Journal of Computational Mathematics
Vol.xx, No.x, 2025, 1–35.
http://www.global-sci.org/jcm
doi:10.4208/jcm.2505-m2024-0057

UNIFORM SUFFICIENT CONDITION FOR THE RECOVERY OF NON-STRICTLY BLOCK k-SPARSE SIGNALS BY WEIGHTED $\ell_{2,p} - \alpha\ell_{2,q}$ NONCONVEX MINIMIZATION METHOD*

Wenhui Liu and Anhua Wan[1)]
School of Mathematics, Sun Yat-sen University, Guangzhou 510275, China
Emails: liuwh37@mail2.sysu.edu.cn, wananhua@mail.sysu.edu.cn

Abstract

Recovery of block sparse signals with partially-known block support information is of particular importance in compressed sensing. A uniform sufficient condition guaranteeing stable recovery of non-strictly block k-sparse signals is established via the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ nonconvex minimization method, and the reconstruction error is precisely bounded in terms of the residual of block-sparsity and the measurement error. Furthermore, a series of contrastive numerical experiments reveal that exploiting the approximate block-sparsity characteristic and the nonuniform prior block support estimate substantially promotes the performance of reconstruction for block-structural signals.

Mathematics subject classification: 94A12, 90C26.
Key words: Compressed sensing, Approximate block-sparsity, Weighted $\ell_{2,p} - \alpha\ell_{2,q}$ minimization, Prior block support information, Block restricted isometry property.

1. Introduction

As is well-known, the aim of compressed sensing is to recover high-dimensional sparse signal $x \in \mathbb{R}^n$ from

$$y = Ax + z, \tag{1.1}$$

where A is the measurement matrix of size $N \times n$ ($N \ll n$), $z \in \mathbb{R}^N$ is a vector of measurement errors, and $y \in \mathbb{R}^N$ is the observation vector [5, 6, 9].

This paper focuses on the practical scenario where the unknown signal to be recovered exhibits a specific structural property known as block-sparsity, that is, the characteristic that the nonzero entries appear in clustered regions. The block-sparsity structure naturally arises in various real-world applications where signals or data exhibit inherent clustering patterns. Such signals have been widely identified and utilized across various fields, including DNA (deoxyribonucleic acid) microarrays [28], color imaging [26] and motion segmentation [31]. DNA microarray data often exhibits block-sparsity as gene expression levels typically show strong correlations within specific functional groups or pathways, and the nonzero coefficients which represent active genes often cluster together rather than being randomly distributed. In image processing, natural images often have block-sparse representations especially in transform domains such as wavelet or DCT (discrete cosine transform), for instance, edges or textures in an image tend to create clusters of significant coefficients, while smooth regions correspond to blocks of near-zero values. In motion segmentation, the motion vectors of pixels associated

* Received March 13, 2024 / Revised version received December 3, 2024 / Accepted May 12, 2025 / Published online October 8, 2025 /
1) Corresponding author

with the same object tend to form clusters, leading to a block-sparse structure in the representation of motion fields. Suppose that the signal $x \in \mathbb{R}^n$ is split into m blocks with every block sequentially satisfying $x[i] \in \mathbb{R}^{d_i}$ $(i = 1, 2, \dots, m)$ and $\sum_{i=1}^{m} d_i = n$, that is to say,

$$x = (\underbrace{x_1, \cdots, x_{d_1}}_{x[1]^\top}, \underbrace{x_{d_1+1}, \cdots, x_{d_1+d_2}}_{x[2]^\top}, \cdots, \underbrace{x_{n-d_m+1}, \cdots, x_n}_{x[m]^\top})^\top. \tag{1.2}$$

Such a signal is referred to as a block-structural signal over the block index set

$$\mathcal{I} = \{d_1, d_2, \cdots, d_m\}.$$

Precisely speaking, $x \in \mathbb{R}^n$ is called block k-sparse over $\mathcal{I}$ if the cardinality of $\{i : x[i] \neq \mathbf{0}\}$ is no more than k. Equivalently,

$$\|x\|_{2,0} = \sum_{i=1}^{m} \mathrm{I}(\|x[i]\|_2) \leq k,$$

where $\mathrm{I}(\cdot)$ represents an indicator function that equals to one for positive variable and equals to zero otherwise.

Denote

$$\|x\|_{2,\infty} := \max_{1\leq i\leq m} \|x[i]\|_2, \quad \|x\|_{2,r} := \left(\sum_{i=1}^{m} \|x[i]\|_2^r\right)^{\frac{1}{r}}, \quad \forall\, r > 0.$$

It is noteworthy that

$$\|x\|_{2,2} = \sqrt{\sum_{i=1}^{m} \|x[i]\|_2^2} = \sqrt{\sum_{i=1}^{n} x_i^2} = \|x\|_2.$$

The subsequent concept of block restricted isometry property (B-RIP) provides a crucial framework for the reconstruction of block sparse signals.

Definition 1.1 ([12]). *Assume that $A \in \mathbb{R}^{N\times n}$ is a measurement matrix, and $k \in \mathbb{N} \cap [1, n]$. The k-ordered block restricted isometry constant (B-RIC) for A is taken as the smallest $\delta_{k|\mathcal{I}}$ which satisfies*

$$\left(1 - \delta_{k|\mathcal{I}}\right) \|x\|_2^2 \leq \|Ax\|_2^2 \leq \left(1 + \delta_{k|\mathcal{I}}\right) \|x\|_2^2, \tag{1.3}$$

where $x \in \mathbb{R}^n$ runs through all block k-sparse vectors over $\mathcal{I}$.

When particularly $d_i = 1$ for $i = 1, 2, \dots, m$, the notion of B-RIC $\delta_{k|\mathcal{I}}$ degenerates to the conventional restricted isometry constant (RIC) δ_k.

In order to make full use of block structure, the $\ell_{2,1}$ minimization method mathematically formulated as

$$\hat{x} = \arg\min_{x\in\mathbb{R}^n} \{\|x\|_{2,1} : \|y - Ax\|_2 \leq \epsilon\} \tag{1.4}$$

was considered in [12, 23, 24], where it is referred to as the ℓ_2/ℓ_1 minimization. For the $\ell_{2,1}$ minimization method (1.4), Eldar and Mishali [12] proved that if the measurement matrix A is subject to $\delta_{2k|\mathcal{I}} < \sqrt{2}-1$, the $\ell_{2,1}$ minimization guarantees the robust recovery of block k-sparse signals. Lin and Li [24] proposed the B-RIP condition $\delta_{2k|\mathcal{I}} < 0.4931$ and also provided another condition $\delta_{k|\mathcal{I}} < 0.307$. Li and Chen [23] proved the sufficient condition $\delta_{tk|\mathcal{I}} < \sqrt{(t-1)/t}$ for $t \geq 4/3$.

In recent years, the $\ell_{2,1} - \ell_2$ minimization method which provides an alternative way to recover block sparse signals by

$$\hat{x} = \arg\min_{x \in \mathbb{R}^n} \{\|x\|_{2,1} - \|x\|_2 : \|y - Ax\|_2 \leq \epsilon\}, \tag{1.5}$$

has attracted much attention. Based on block RIP, Wang *et al.* [34] established a sufficient condition

$$\delta_{2k|\mathcal{I}} + \frac{\sqrt{k}+1}{\sqrt{k}-1}\delta_{3k|\mathcal{I}} < 1,$$

and also numerically showed that $\ell_{2,1} - \ell_2$ minimization presented better recovery performance for recovering block sparse signal from highly coherent measurement matrix in comparison with $\ell_{2,1}$ minimization.

In reality, prior support information can be partially extracted in a variety of applications such as genetics and image processing [13, 17, 21, 22, 30, 36]. Denote $\tilde{T} \subseteq \{1, 2, \ldots, m\}$ as the prior block support estimate, and

$$\mathrm{w}_i = \begin{cases} w, & i \in \tilde{T}, \\ 1, & i \notin \tilde{T}, \end{cases} \quad w \in [0, 1].$$

Recently, Zhang and Zhang [37] extended their previous work for sparse signal recovery in [38] to block sparse signal recovery, proposed the weighted $\ell_{2,1} - \ell_2$ minimization in ℓ_2-bounded noise setting

$$\hat{x} = \arg\min_{x \in \mathbb{R}^n} \left\{ \sum_{i=1}^{m} \mathrm{w}_i \|x[i]\|_2 - \sqrt{\sum_{i=1}^{m} \mathrm{w}_i^2 \|x[i]\|_2^2} : \|y - Ax\|_2 \leq \epsilon \right\}, \tag{1.6}$$

and established a high-order block RIP condition for the robust recovery of block sparse signals with prior block support information.

In the literature on the standard compressed sensing [14, 29, 32, 33], it has been shown that sparse signals can be recovered from fewer linear measurements via ℓ_p $(p \in (0, 1])$ nonconvex minimization compared to ℓ_1 convex minimization.

On the other hand, in some important applications, there always exists non-uniformity in the set of signal blocks, for example, the behavior of neurons exhibits non-uniform clustered responses in computational neuroscience problems [13]. Non-uniform structures inherently carry domain-specific prior knowledge. Motivated by this, we attempt to consider the impact of non-uniform support information on blocks for the purpose of promoting the performance of sparse signal recovery.

Inspired by [16], we assign an appropriate weight vector w such that critical blocks of x with higher expected signal energy $\|x[i]\|_2$ receive reduced penalties in the weighted objective function.

To streamline the presentation of the subsequent theorems, we introduce the following hypothesis, denoted as "**(H)**":

(H) Suppose the prior block support estimate is $\tilde{T} = \bigcup_{j=1}^{L} \tilde{T}_j$, where $\tilde{T}_j, j = 1, 2, \ldots, L$, are disjoint subsets of $\{1, 2, \ldots, m\}$ with $|\tilde{T}_j| = \rho_j k$, $|T_0 \cap \tilde{T}_j| = \nu_j |\tilde{T}_j|$. Denote

$$\mathrm{w} = (\mathrm{w}_1, \mathrm{w}_2, \ldots, \mathrm{w}_m)^{\mathsf{T}},$$

where the non-uniform weights are characterized by

$$\mathrm{w}_i := \begin{cases} w_j, & i \in \tilde{T}_j, \\ 1, & i \notin \tilde{T}, \end{cases} \quad w_j \in [0,1]. \tag{1.7}$$

Denote $w_0 = 1$.

Denote

$$\|x\|_{2,r,\mathrm{w}} := \left(\sum_{i=1}^{m} \mathrm{w}_i^r \|x[i]\|_2^r \right)^{\frac{1}{r}}, \quad \forall\, r > 0.$$

On the basis of the incorporation of the prior block support estimate $\tilde{T} = \bigcup_{j=1}^{L} \tilde{T}_j$, now we introduce the following weighted $\ell_{2,p} - \alpha\ell_{2,q}$ nonconvex minimization:

$$\hat{x} = \arg\min_{x \in \mathbb{R}^n} \left\{ \|x\|_{2,p,\mathrm{w}}^p - \alpha \|x\|_{2,q,\mathrm{w}}^p : y - Ax \in \mathcal{B} \right\}, \tag{1.8}$$

i.e.

$$\hat{x} = \arg\min_{x \in \mathbb{R}^n} \left\{ \sum_{i=1}^{m} \mathrm{w}_i^p \|x[i]\|_2^p - \alpha \left(\sum_{i=1}^{m} \mathrm{w}_i^q \|x[i]\|_2^q \right)^{\frac{p}{q}} : y - Ax \in \mathcal{B} \right\}, \tag{1.9}$$

where $(\alpha, p, q) \in [0,1] \times (0,1] \times [1,2] \setminus (1,1,1)$, w_i is defined in (1.7), and $\mathcal{B} \subseteq \mathbb{R}^N$ denotes some noise structure. In the objective function, smaller weights are assigned to those blocks which are more likely to have non-zero entries.

In particular, when all $w_i = 1$, the $\ell_{2,p} - \alpha\ell_{2,q}$ minimization (1.8) is thus formulated as

$$\hat{x} = \arg\min_{x \in \mathbb{R}^n} \left\{ \sum_{i=1}^{m} \|x[i]\|_2^p - \alpha \left(\sum_{i=1}^{m} \|x[i]\|_2^q \right)^{\frac{p}{q}} : y - Ax \in \mathcal{B} \right\}. \tag{1.10}$$

It is obvious that block-sparse signals are naturally sparse in the standard sense. When all $d_i = 1$ in (1.10), the $\ell_{2,p} - \alpha\ell_{2,q}$ minimization degenerates into the $\ell_p - \alpha\ell_q$ minimization in our previous work [19] which is defined as

$$\hat{x} = \arg\min_{x \in \mathbb{R}^n} \left\{ \sum_{i=1}^{n} |x_i|^p - \alpha \left(\sum_{i=1}^{n} |x_i|^q \right)^{\frac{p}{q}} : y - Ax \in \mathcal{B} \right\}. \tag{1.11}$$

The following example shows the advantage of $\ell_p - \alpha\ell_q$ nonconvex minimization by suitable choices of the adjustable parameters p, q and α.

Example 1.1. Suppose the measurement matrix is $A = (1,1,4,2,1,1,1,3,1,1,3,1) \in \mathbb{R}^{1\times 12}$ and consider the following two vectors in $\mathbb{R}^{12}$:

$$x^* = (0.75, 0.75, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^\mathsf{T},$$
$$x^{**} = (0.01, 0.05, 0.07, 0.05, 0.02, 0.03, 0.04, 0.05, 0.7, 0.02, 0.03, 0.01)^\mathsf{T}.$$

It can be readily verified that $Ax^* = Ax^{**} = 1.5$. The former vector x^* is clearly much sparser than the latter one x^{**}. However, the minimizer of the following objective functions (1)-(3) such as $\ell_1 - \ell_2$ metric in [38] does not correspond to the sparser vector x^*:

(1) For ℓ_1 norm,

$$\|x^*\|_1 = 1.5 > \|x^{**}\|_1 = 1.08. \tag{1.12}$$

(2) For $\ell_1 - \alpha\ell_2$ metric with arbitrary $\alpha \in [0,1]$,

$$\|x^*\|_1 - \alpha\|x^*\|_2 \approx 1.5 - 1.0607\alpha > \|x^{**}\|_1 - \alpha\|x^{**}\|_2 \approx 1.08 - 0.7118\alpha. \tag{1.13}$$

(3) For $\ell_{0.8}$ quasi-norm,

$$\|x^*\|_{0.8}^{0.8} \approx 1.5888 > \|x^{**}\|_{0.8}^{0.8} \approx 1.4788. \tag{1.14}$$

By proper choice of the adjustable parameters $p \in (0,1]$, $q \in [1,2]$ and $\alpha \in [0,1]$, the minimizer of the $\ell_p - \alpha\ell_q$ minimization method presents the truly sparser vector x^*. For instance,

(4) For $\ell_{0.8} - \alpha\ell_2$ metric with arbitrary $\alpha \in [0.39,1]$,

$$\|x^*\|_{0.8}^{0.8} - \alpha\|x^*\|_2^{0.8} \approx 1.5888 - 1.0482\alpha < \|x^{**}\|_{0.8}^{0.8} - \alpha\|x^{**}\|_2^{0.8} \approx 1.4788 - 0.7620\alpha. \tag{1.15}$$

(5) For $\ell_{0.8} - \alpha\ell_{1.2}$ metric with arbitrary $\alpha \in [0.31,1]$,

$$\|x^*\|_{0.8}^{0.8} - \alpha\|x^*\|_{1.2}^{0.8} \approx 1.5888 - 1.2611\alpha < \|x^{**}\|_{0.8}^{0.8} - \alpha\|x^{**}\|_{1.2}^{0.8} \approx 1.4788 - 0.8990\alpha. \tag{1.16}$$

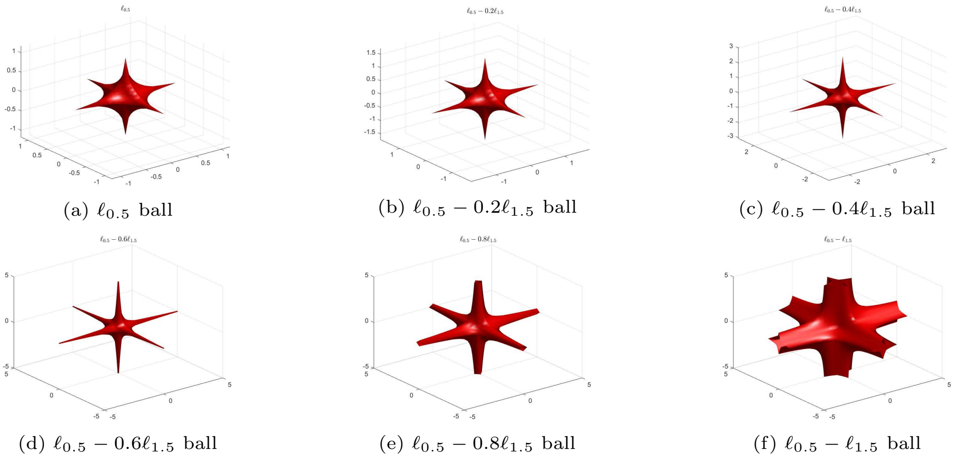
As the aforementioned example implies, when x^* and x^{**} are two feasible solutions of the optimization problem, the $\ell_1 - \alpha\ell_2$ minimization including the conventional ℓ_1 minimization and $\ell_1 - \ell_2$ minimization tends to favor x^{**} as the optimal solution over the sparser one x^*. In contrast, the $\ell_p - \alpha\ell_q$ minimization with appropriate choices of p, q and α does work. As demonstrated in Figs. 1.1-1.3, for any $p \in (0,1]$, the vector sets with smaller $\ell_p - \alpha\ell_q$ values are significantly closer to the set of sparse vectors on the unit sphere of $\mathbb{R}^3$ compared with those of ℓ_p-(quasi-)norm.

Since many studies have already proved that ℓ_1 minimization, ℓ_p minimization and ℓ_{1-2} minimization can stably and robustly recover sparse signals under certain conditions, more generally we consider $\ell_p - \alpha\ell_q$ minimization. The above series of examples shows that this method has notable advantages in sparse signal recovery.

Research on block sparse signals has been relatively mature, and some works on block structured signals have proved that considering the block sparsity can help achieve better performance of recovery. The proposed weighted $\ell_{2,p} - \alpha\ell_{2,q}$ optimization model employs a dual-norm formulation where the $\ell_{2,p}$-norm term promotes block sparsity by aggregating intra-block magnitudes with ℓ_2-norm and enforcing inter-block sparsity through ℓ_p minimization $(0 < p \leq 1)$. The subtractive $\alpha\ell_{2,q}$-norm term $(1 \leq q \leq 2)$ with adaptive parameter $\alpha \in [0,1]$ creates a preferential relaxation effect. This structural treatment ensures effective recovery of block-sparse signals by exploiting their inherent clustering characteristics.

When partial prior information of block support is available, the assignment of non-uniform weights to distinct signal blocks formally embeds this structural prior into the optimization framework through mathematical formalization, while quantitatively regulating the incorporation of such domain knowledge through explicit weighting mechanisms. The known block support experience less penalization in the composite $\ell_{2,p} - \alpha\ell_{2,q}$ objective function, allowing them to retain signal energy more effectively during the optimization process.

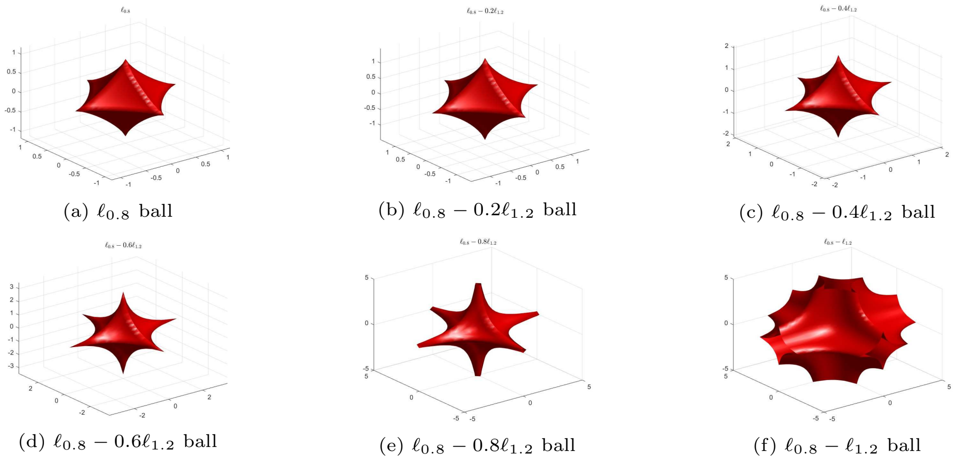
Therefore, in this paper, we propose the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ minimization for block sparse signal reconstruction when partial block support information is available. We are devoted to exploring the recovery of non-strictly block k-sparse signals with prior support estimate in the presence of noise.



(a) $\ell_{0.5}$ ball (b) $\ell_{0.5} - 0.2\ell_{1.5}$ ball (c) $\ell_{0.5} - 0.4\ell_{1.5}$ ball

(d) $\ell_{0.5} - 0.6\ell_{1.5}$ ball (e) $\ell_{0.5} - 0.8\ell_{1.5}$ ball (f) $\ell_{0.5} - \ell_{1.5}$ ball

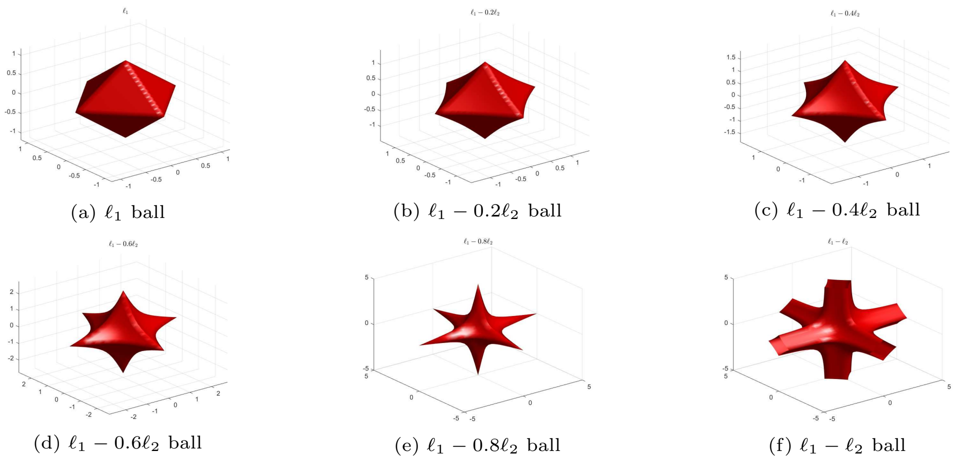
Fig. 1.1. The $\ell_{0.5} - \alpha\ell_{1.5}$ unit balls in $\mathbb{R}^3$ with $\alpha = 0, \alpha = 0.2, \alpha = 0.4, \alpha = 0.6, \alpha = 0.8, \alpha = 1$, respectively.



(a) $\ell_{0.8}$ ball (b) $\ell_{0.8} - 0.2\ell_{1.2}$ ball (c) $\ell_{0.8} - 0.4\ell_{1.2}$ ball

(d) $\ell_{0.8} - 0.6\ell_{1.2}$ ball (e) $\ell_{0.8} - 0.8\ell_{1.2}$ ball (f) $\ell_{0.8} - \ell_{1.2}$ ball

Fig. 1.2. The $\ell_{0.8} - \alpha\ell_{1.2}$ unit balls in $\mathbb{R}^3$ with $\alpha = 0, \alpha = 0.2, \alpha = 0.4, \alpha = 0.6, \alpha = 0.8, \alpha = 1$, respectively.



(a) ℓ_1 ball (b) $\ell_1 - 0.2\ell_2$ ball (c) $\ell_1 - 0.4\ell_2$ ball

(d) $\ell_1 - 0.6\ell_2$ ball (e) $\ell_1 - 0.8\ell_2$ ball (f) $\ell_1 - \ell_2$ ball

Fig. 1.3. The $\ell_1 - \alpha\ell_2$ unit balls in $\mathbb{R}^3$ with $\alpha = 0, \alpha = 0.2, \alpha = 0.4, \alpha = 0.6, \alpha = 0.8, \alpha = 1$, respectively.

The remainder of this paper proceeds as follows. In Section 2, we introduce the necessary terminologies and a series of lemmas. The main theorems are provided in Section 3, with their proofs detailed in the Appendices A-H. We implement a series of numerical experiments and discuss the experimental results in Section 4. Finally, Section 5 concludes the whole work in this paper.

2. Terminologies and Lemmas

In this section, we begin by introducing some preliminary terminologies.

For any $x \in \mathbb{R}^n$ with the block structure of (1.2), the block index set $\text{supp}[x] = \{i : x[i] \neq \mathbf{0}\}$ is defined as block support. For any block index set $\Gamma \subseteq \{1, 2, \dots, m\}$, $x_\Gamma \in \mathbb{R}^n$ denotes the block-structured vector which preserves the blocks of x on the block indices in Γ with all the other blocks adjusted as $\mathbf{0}$. Denote $x_{\max(k)} \in \mathbb{R}^n$ as the block-structured vector which maintains the largest k blocks of x (in the sense of ℓ_2 norm $\|x[i]\|_2$) with all the other blocks set to $\mathbf{0}$, and $x_{-\max(k)} := x - x_{\max(k)}$.

Lemma 2.1 ([39, Lemma 2.1]). *Suppose that $x \in \mathbb{R}^n$ obeys $\|x\|_{2,0} = l, \|x\|_{2,\infty} \leq \tau$ and $\|x\|_{2,p}^p \leq k\tau^p$ with $k \leq l$ being a positive integer, $\tau > 0$ and $0 < p \leq 1$. Then x can be represented as the convex combination of block k-sparse vectors $u_i \in \mathbb{R}^n$, i.e. $x = \sum_i \lambda_i u_i$, where $\lambda_i > 0, \sum_i \lambda_i = 1, \|u_i\|_{2,0} \leq k$. Moreover,*

$$\sum_{i=1}^{M} \lambda_i \|u_i\|_2^2 \leq \min\left\{\frac{l}{k}\|x\|_2^2,\ \tau^p \|x\|_{2,2-p}^{2-p}\right\}. \tag{2.1}$$

Lemma 2.2 ([3, Lemma 5.3]). *Suppose that $s > r \geq 1, a_1 \geq a_2 \geq \cdots \geq a_s \geq 0, b \geq 0$ and $\sum_{i=r+1}^{s} a_i \leq \sum_{i=1}^{r} a_i + b$. Then*

$$\sum_{i=r+1}^{s} a_i^w \leq r\left[\left(\frac{1}{r}\sum_{i=1}^{r} a_i^w\right)^{\frac{1}{w}} + \frac{b}{r}\right]^w \tag{2.2}$$

for any $w \geq 1$.

Lemma 2.3 ([4, Lemma 5.1]). *Suppose $A \in \mathbb{R}^{N\times n}, k \geq 2$ is an integer, $s > 1$, and sk is an integer. Then $\delta_{sk} \leq (2s-1)\delta_k$.*

Lemma 2.4 ([15, Lemma 9.2]). *Suppose that $A \in \mathbb{R}^{N\times n}$ is a sub-Gaussian random matrix. Then there exists a constant $C > 0$ (depending only on the sub-Gaussian parameters) such that the restricted isometry constant of $A/\sqrt{m}$ satisfies $\delta_s < \delta$ with probability at least $1-\varepsilon$ provided that*

$$m \geq C\delta^{-2}\left(s\ln\frac{\mathrm{e}n}{s} + \ln\frac{2}{\varepsilon}\right).$$

Setting $\varepsilon = 2\mathrm{e}^{-\delta^2 m/(2C)}$ yields the condition $m \geq 2C\delta^{-2}s\ln(\mathrm{e}n/s)$ which guarantees the probability $P(\delta_s < \delta) \geq 1 - 2\mathrm{e}^{-\delta^2 m/(2C)}$.

Lemma 2.5. *Suppose $A \in \mathbb{R}^{N\times n}$, $k \geq 2$ is an integer, $s > 1$, and sk is an integer. Then $\delta_{sk|\mathcal{I}} \leq (2s-1)\delta_{k|\mathcal{I}}$.*

In what follows in this section, we establish a series of lemmas which will play important roles in the proof of the main theorems.

Lemma 2.6. *Suppose that $\hat{x} \in \mathbb{R}^n$ is the minimizer of the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ nonconvex minimization (1.8), where $(\alpha, p, q) \in [0,1] \times (0,1] \times [1,2] \setminus (1,1,1)$ and $x \in \mathbb{R}^n$ is block structural over the block index set $\mathcal{I} = \{d_1, d_2, \cdots, d_m\}$. Denote $h = \hat{x} - x$ and $T_0 = \mathrm{supp}[x_{\max(k)}]$. Suppose that $\tilde{T}_j, j = 1, 2, \ldots, L$, are disjoint subsets of $\{1, 2, \ldots, m\}$,*

$$|\tilde{T}_j| = \rho_j k, \quad |T_0 \cap \tilde{T}_j| = \nu_j |\tilde{T}_j|, \quad \tilde{T} = \bigcup_{j=1}^{L} \tilde{T}_j$$

is the block support estimate. Then

$$\begin{aligned} \|h_{T_0^\complement}\|_{2,p}^p \leq w_L^p \|h_{T_0}\|_{2,p}^p &+ \sum_{j=2}^{L} \left(w_{j-1}^p - w_j^p\right) \Big\| h_{(T_0 \cup \bigcup_{i=j}^{L} \tilde{T}_i) \setminus \bigcup_{i=j}^{L} (\tilde{T}_i \cap T_0)} \Big\|_{2,p}^p \\ &+ \left(1 - w_1^p\right) \Big\| h_{(T_0 \cup \tilde{T}) \setminus \bigcup_{i=1}^{L} (\tilde{T}_i \cap T_0)} \Big\|_{2,p}^p + 2\|x_{T_0^\complement}\|_{2,p,\mathrm{w}}^p + \alpha \|h\|_{2,q}^p. \end{aligned} \tag{2.3}$$

Lemma 2.7. *Suppose $A \in \mathbb{R}^{N \times n}, k > 0, t > \hat{d} \geq 1, \hat{d}k$ is an integer, and $tk \geq 2$ is an integer. Then*

$$\delta_{2(t-\hat{d})k|\mathcal{I}} \leq \mu \delta_{tk|\mathcal{I}}, \tag{2.4}$$

where

$$\mu = \begin{cases} 1, & \hat{d} < t \leq 2\hat{d}, \\ \dfrac{3t - 4\hat{d}}{t}, & t > 2\hat{d}. \end{cases} \tag{2.5}$$

Lemma 2.8. *Suppose that $\hat{x} \in \mathbb{R}^n$ is the minimizer of the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ nonconvex minimization (1.8), where $(\alpha, p, q) \in [0,1] \times (0,1] \times [1,2] \setminus (1,1,1)$ and $x \in \mathbb{R}^n$ is block structural over the block index set $\mathcal{I} = \{d_1, d_2, \cdots, d_m\}$. For any $d \geq 1$, suppose that the matrix $A \in \mathbb{R}^{N \times n}$ satisfies block RIP of order tk with $\delta_{tk|\mathcal{I}} \in [0,1)$ and block RIP of order $2(t-d)k$ with $\delta_{2(t-d)k|\mathcal{I}} \in [0,1)$ for some $t > d$, and $K \subseteq \{1, 2, \ldots, m\}$ is of the size $|K| \leq dk$. Denote*

$$c = \frac{1}{2} - \frac{1}{4}\left(\sqrt{\mu^2 p^2 + 4(1-p)} - \mu p\right), \tag{2.6}$$

where μ is defined in (2.5). Then

$$\|h_K\|_2 \leq \eta \|Ah\|_2 + \beta \left\|h_{K^\complement}\right\|_{2,p}, \tag{2.7}$$

where

$$\eta = \sqrt{1 + \delta_{tk|\mathcal{I}}} \max\left\{\frac{1}{1 + (2c-1)\delta_{tk|\mathcal{I}}}, \frac{1 + (4c-1)\delta_{tk|\mathcal{I}}}{1 - \delta_{tk|\mathcal{I}}^2}\right\}, \tag{2.8}$$

$$\beta = \delta_{tk|\mathcal{I}}^{\frac{1}{p}} \left(\frac{p}{1 - \delta_{tk|\mathcal{I}}^2}\right)^{\frac{1}{2}} \left(\frac{\sqrt{\mu^2 p^2/4 + 1 - p} + \mu p/2}{1 + \mu \delta_{tk|\mathcal{I}}}\right)^{\frac{1}{p}-1} \left[\frac{\mu(1-2c) + 1 - 2c + 2c^2}{(1-p/2)(t-d)k}\right]^{\frac{1}{p}-\frac{1}{2}}. \tag{2.9}$$

Lemma 2.9. *Suppose that $\hat{x} \in \mathbb{R}^n$ is the minimizer of the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ nonconvex minimization (1.8) where $(\alpha, p, q) \in [0,1] \times (0,1] \times [1,2] \setminus (1,1,1)$, and $x \in \mathbb{R}^n$ is block structural over the block index set $\mathcal{I} = \{d_1, d_2, \cdots, d_m\}$. Denote $h = x - \hat{x}$. For any $d \geq 1$, if*

$$\|h_{-\max(dk)}\|_{2,p}^p \leq \|h_{\max(dk)}\|_{2,p}^p + 2\|x_{-\max(k)}\|_{2,p,w}^p + \alpha \|h\|_{2,q}^p, \tag{2.10}$$

and there exist $\phi_1 \geq 0$ *and* $\phi_2 \geq 0$ *such that*

$$\|h\|_{2,q} \leq \phi_1 \|Ah\|_2 + \phi_2 \|x_{-\max(k)}\|_{2,p,w}, \tag{2.11}$$

then

$$\|h\|_{2,q} \leq \phi_3 \|A^{\mathsf{T}} Ah\|_{2,\infty} + \phi_4 \|x_{-\max(k)}\|_{2,p,w}, \tag{2.12}$$

where

$$\begin{aligned} \phi_3 &= \left(\frac{\phi_1}{\theta}\right)^2 \left[\left(1 + 3^{\frac{1}{p}-1}\right)(dk)^{1-\frac{1}{q}} + \alpha^{\frac{1}{p}} \left(\frac{3}{dk}\right)^{\frac{1}{p}-1}\right], \\ \phi_4 &= \max\left\{\frac{\phi_2}{1-\theta}, \frac{2^{1/p}}{(3^{1-1/p}+1)(dk)^{1/p-1/q} + \alpha^{1/p}}\right\} \end{aligned} \tag{2.13}$$

for any $\theta \in (0,1)$.

Some symbols and notations used in this paper are defined in Table 2.1.

Table 2.1: Symbols and notations.

Notation	Definition
$\mathcal{B}^{\ell_2}(\varepsilon)$	$\{z \in \mathbb{R}^N : \|z\|_2 \leq \varepsilon\}$
$\mathcal{B}^{DS}(\varepsilon)$	$\{z \in \mathbb{R}^N : \|A^{\mathsf{T}} z\|_\infty \leq \varepsilon\}$
$\mathcal{B}^{BDS}(\varepsilon)$	$\{z \in \mathbb{R}^N : \|A^{\mathsf{T}} z\|_{2,\infty} \leq \varepsilon\}$
T_0	$\mathrm{supp}[x_{\max(k)}]$
ζ_i	$\max\left\{\sum_{j=i}^{L} \nu_j\rho_j, \sum_{j=i}^{L} (1-\nu_j)\rho_j\right\}$
d	$\begin{cases} 1, & \prod_{j=1}^{L} w_i = 1 \\ \max_{1\leq i\leq L} \left\{1 - \sum_{j=i}^{L} \nu_j\rho_j + \zeta_i\right\}, & \text{otherwise.} \end{cases}$
μ	$\begin{cases} 1, & d < t \leq 2d \\ \frac{3t-4d}{t}, & t > 2d \end{cases}$
c	$\frac{1}{2} - \frac{1}{4}\left(\sqrt{\mu^2 p^2 + 4(1-p)} - \mu p\right)$
σ	$\left[w_L^p + \sum_{j=1}^{L} \left(w_{j-1}^p - w_j^p\right)\left(1 + \sum_{i=j}^{L} \rho_i - 2\sum_{i=j}^{L} \nu_i\rho_i\right)^{\frac{2-p}{2}}\right]^{\frac{2}{2-p}}$

3. Main Results

In realistic situations, it is inconceivable to measure a signal with infinite precision, which indicates that the measurement vector y is only an approximation of the vector Ax in (1.1) [15]. In block sparse signal recovery, the following three categories of measurement noise settings are of particular significance. The ℓ_2-bounded noise setting is

$$y - Ax \in \mathcal{B}^{\ell_2}(\varepsilon), \tag{3.1}$$

which was motivated by [10]. The second noise setting is

$$y - Ax \in \mathcal{B}^{DS}(\varepsilon). \tag{3.2}$$

Candès and Tao [7] considered the above noise setting in the Dantzig selector (DS) procedure. For the sake of convenience, we call (3.2) DS-bounded noise setting.

In the study of recovery for block sparse signals, Zhou and Huang [39] considered

$$y - Ax \in \mathcal{B}^{BDS}(\varepsilon). \tag{3.3}$$

In order to distinguish from (3.2), we denote (3.3) as BDS-bounded noise setting.

Firstly, in noisy setting under ℓ_2-hounded perturbation, the next theorem establishes a uniform sufficient condition which guarantees stable recovery of general non-strictly block k-sparse signals.

Theorem 3.1. *Consider the signal recovery model* (1.1) *with* $\|z\|_2 \le \epsilon$, *where* $x \in \mathbb{R}^n$ *is block structural over the block index set* $\mathcal{I} = \{d_1, d_2, \cdots, d_m\}$. *Suppose that* $(\alpha, p, q) \in [0,1] \times (0,1] \times [1,2] \setminus (1,1,1)$, $\varepsilon \ge \epsilon$, $k \in \{1, 2, \dots, m\}$, $k \ne \alpha$.

Under the hypothesis **(H)**, *if the measurement matrix A satisfies the B-RIP condition*

$$\delta_{tk|\mathcal{I}} < \delta(p, q, \alpha, t, \sigma, d) \tag{3.4}$$

for some $t > d$, *where* $\delta := \delta(p, q, \alpha, t, \sigma, d)$ *is the unique positive solution of*

$$\begin{aligned} &\left(\frac{1}{\delta^2} - 1\right)\left(\frac{1}{\delta} + \mu\right)^{\frac{2(1-p)}{p}} \\ &= p\left(\frac{\mu p}{2} + \sqrt{\frac{\mu^2 p^2}{4} + 1 - p}\right)^{\frac{2(1-p)}{p}} \\ &\quad \times \left[\frac{\mu(1-2c) + 1 - 2c + 2c^2}{(1-p/2)(t-d)}\right]^{\frac{2-p}{p}} \left[\sigma^{\frac{2-p}{2}} + \frac{2^{p/q} d^{(2-p)/2}\alpha}{(dk)^{(q-p)/q} - \alpha}\right]^{\frac{2}{p}}, \end{aligned} \tag{3.5}$$

then the minimizer $\hat{x}^{\ell_2}$ *of*

$$\hat{x}^{\ell_2} = \arg\min_{x \in \mathbb{R}^n} \left\{\|x\|_{2,p,\mathrm{w}}^p - \alpha \|x\|_{2,q,\mathrm{w}}^p : \ \|y - Ax\|_2 \le \varepsilon\right\}$$

obeys

$$\|\hat{x}^{\ell_2} - x\|_{2,q} \le C_1(\varepsilon + \epsilon) + C_2 \|x_{-\max(k)}\|_{2,p,w}, \tag{3.6}$$

where

$$\begin{aligned} C_1 &= \frac{2^{1/p+1/q-1}(dk)^{(2-p)/(2p)}\eta}{\{[1 - (\sigma k)^{(2-p)/2}\beta^p][(dk)^{(q-p)/q} - \alpha] - 2^{p/q}\alpha(dk)^{(2-p)/2}\beta^p\}^{1/p}}, \\ C_2 &= 2^{\frac{2}{p}-1}\left\{\frac{2^{p/q}(dk)^{(2-p)/2}\beta^p + 1 - (\sigma k)^{(2-p)/2}\beta^p}{[1 - (\sigma k)^{(2-p)/2}\beta^p][(dk)^{(q-p)/q} - \alpha] - 2^{p/q}\alpha(dk)^{(2-p)/2}\beta^p}\right\}^{\frac{1}{p}} \end{aligned} \tag{3.7}$$

with η *and* β *respectively defined in* (2.8) *and* (2.9).

Theorem 3.1 shows that if the measurement matrix satisfies certain RIP condition, all block k-sparse signals can be stably and robustly recovered by the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ minimization (1.8) based on $\|z\|_2 \le \epsilon$, where z is a type of noise. According to Theorem 3.1, we infer that the new results improve the existing ones, and the details are presented in the following remark.

Remark 3.1. In particular, if $p = 1, q = 2, \alpha = 1$ and all w_j's take the same value w, then the above results for the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ minimization (1.8) degenerate to the results for the weighted $\ell_{2,1} - \ell_2$ minimization (1.6) in [37]. Meanwhile, the new results improve the reconstruction error estimation in [37].

Actually, when $p = 1, q = 2, \alpha = 1$ and $\varepsilon = \epsilon$, we arrive at

$$\eta = \frac{\sqrt{1+\delta_{tk|\mathcal{I}}}}{1-\delta_{tk|\mathcal{I}}}, \quad \beta = \frac{\delta_{tk|\mathcal{I}}}{\sqrt{(1-\delta_{tk|\mathcal{I}}^2)(t-d)k}},$$

and (3.6) turns to be

$$\|\hat{x}^{\ell_2} - x\|_2 \leq D_1\epsilon + D_2\|x_{-\max(k)}\|_{2,1,w},$$

where

$$D_1 = \frac{2\sqrt{2dk}\eta}{(1-\sqrt{\sigma k}\beta)(\sqrt{dk}-1) - \sqrt{2dk}\beta},$$
$$D_2 = \frac{2(\sqrt{2dk}\beta - \sqrt{\sigma k}\beta + 1)}{(1-\sqrt{\sigma k}\beta)(\sqrt{dk}-1) - \sqrt{2dk}\beta}.$$

The reconstruction error estimation in [37] is

$$\|\hat{x}^{\ell_2} - x\|_2 \leq \hat{D}_1\epsilon + D_2\|x_{-\max(k)}\|_{2,1,w}, \tag{3.8}$$

where

$$\hat{D}_1 = \frac{2\sqrt{2dk}\hat{\eta}}{(1-\sqrt{\sigma k}\beta)(\sqrt{dk}-1) - \sqrt{2dk}\beta}, \quad \hat{\eta} = \frac{2}{(1-\delta_{tk|\mathcal{I}})\sqrt{1+\delta_{tk|\mathcal{I}}}}.$$

It follows from $\hat{\eta} > \eta$ that $\hat{D}_1 > D_1$, and thus the new result provides tighter reconstruction error estimation.

Furthermore, when particularly $d_i = 1$ in the block index set $\mathcal{I} = \{d_1, d_2, \cdots, d_m\}$, the results in [37] degenerates into the results in [38] via the weighted ℓ_{1-2} minimization, hence the new results also include the results in [38] as special cases.

Remark 3.2. In particular, when $d_i = 1, L = 1, \alpha = 0$ and $\nu_1 \geq 1/2$, we obtain $d = 1$, and the new RIC upper bound (3.5) for $t \in (1, 2]$ turns to be

$$\left(\frac{1}{\delta(p,q,0,t,1,1)} - 1\right)\left(\frac{1}{\delta(p,q,0,t,1,1)} + 1\right)^{\frac{2-p}{p}} = p\left(\frac{2-p}{t-1}\sigma\right)^{\frac{2-p}{p}}. \tag{3.9}$$

It can be readily certified that our results degenerate to the standard weighted ℓ_p minimization and contain the RIP conditions for $t \in [1 + \sigma(2-p)/(2+p), 2]$ in our previous work [11].

Remark 3.3. In particular, when $p = 1, d_i = 1, w_i = 1$ and $t = 2$, the new B-RIP condition (3.4) becomes

$$\delta_{2k} \leq \delta(1,q,\alpha,2,1,1) = \sqrt{\frac{k^{1-1/q} - \alpha}{k^{1-1/q} - \alpha + \alpha^2 2^{1/q}}}. \tag{3.10}$$

The RIP condition in [20] is

$$\delta_{2k} \leq \frac{\tau}{\sqrt{\tau^2 + \hat{a}k^{-1}}}, \tag{3.11}$$

where

$$\tau = \frac{s^{1-1/q-\alpha}}{s^{1-1/q+\alpha}},$$

and

$$\hat{a} = \begin{cases} \left[\frac{1}{1-\alpha(\sqrt[q]{2}-1)}\right]^2 \left[(k-\sqrt[q]{k}\alpha)+\alpha(2-\sqrt[q]{2})\right], & \sqrt[q]{k}\alpha < 1, \\ \left[\frac{1}{1-\alpha(\sqrt[q]{2}-1)}\right]^2 (k-\sqrt[q]{k}\alpha)+1, & \sqrt[q]{k}\alpha \geq 1. \end{cases}$$

By setting $k = 20$, we can find that the RIC upper bound in [20] is larger when q is close to 1, but the RIC upper bound in our result will be larger when $1.2 \leq q \leq 2$.

According to Theorem 3.1, we demonstrate that when the accuracy of estimated block supports is sufficiently high, the minimizer of the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ minimization (1.8) can recover block sparse signals under less stringent condition compared with those required when no block support information is utilized.

Remark 3.4. When in particular $w_j = 1$ for any $j \in \{1, 2, \ldots, L\}$, then $\sigma = 1$ and $d = 1$. In this case, the recovery condition (3.5) on the RIC upper bound degenerates into

$$\begin{aligned} &\left(\frac{1}{\delta^2(p,q,\alpha,t,1,1)} - 1\right)\left(\frac{1}{\delta(p,q,\alpha,t,1,1)} + \mu\right)^{\frac{2(1-p)}{p}} \\ &= p\left(\sqrt{\frac{\mu^2p^2}{4}+1-p}+\frac{\mu p}{2}\right)^{\frac{2(1-p)}{p}} \left[\frac{\mu(1-2c)+1-2c+2c^2}{(1-p/2)(t-1)}\right]^{\frac{2-p}{p}} \left(1+\frac{2^{p/q}\alpha}{k^{(q-p)/q}-\alpha}\right)^{\frac{2}{p}}. \quad (3.12) \end{aligned}$$

In contrast, for general cases $w_j \in [0, 1]$, without loss of generality, we may arrange $\{w_j\}$ in descending order

$$1 = w_0 \geq w_1 \geq w_2 \geq \cdots \geq w_L \geq 0.$$

If $\nu_i \geq 1/2$ for all $i \in \{1, 2, \ldots, L\}$, then $\sigma < 1$ and $d = 1$. As a result, the condition on the RIC upper bound turns to be

$$\begin{aligned} &\left(\frac{1}{\delta^2(p,q,\alpha,t,\sigma,1)} - 1\right)\left(\frac{1}{\delta(p,q,\alpha,t,\sigma,1)} + \mu\right)^{\frac{2(1-p)}{p}} \\ &= p\left(\sqrt{\frac{\mu^2p^2}{4}+1-p}+\frac{\mu p}{2}\right)^{\frac{2(1-p)}{p}} \left[\frac{\mu(1-2c)+1-2c+2c^2}{(1-p/2)(t-1)}\right]^{\frac{2-p}{p}} \left(\sigma^{\frac{2-p}{2}}+\frac{2^{p/q}\alpha}{k^{(q-p)/q}-\alpha}\right)^{\frac{2}{p}}. \quad (3.13) \end{aligned}$$

Obviously, $(1/\delta^2 - 1)(1/\delta + \mu)^{2(1-p)/p}$ is monotonically decreasing with respect to x and it is easy to get

$$\left(1+\frac{2^{p/q}\alpha}{k^{(q-p)/q}-\alpha}\right)^{\frac{2}{p}} > \left(\sigma^{\frac{2-p}{2}}+\frac{2^{p/q}\alpha}{k^{(q-p)/q}-\alpha}\right)^{\frac{2}{p}}.$$

Thus, the RIC upper bound obeying the condition (3.13) is bigger than the one satisfying (3.12). Fig. 3.1 shows the comparison of the RIC upper bound $\delta(1, q, \alpha, 2, 1, 1)$ with that in [20].

Therefore, as long as the accuracy is more than half for all estimated block supports, the minimizer of the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ minimization method recovers block sparse signals under a weaker condition compared with the case when block support information is not employed.

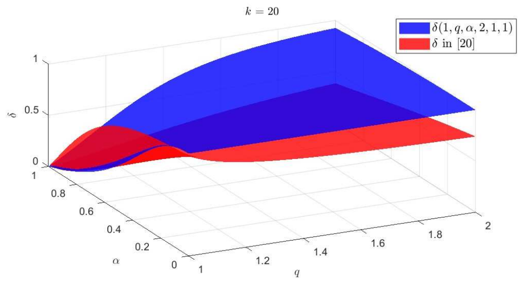


Fig. 3.1. The comparison of the RIC upper bound $\delta(1,q,\alpha,2,1,1)$ with that in [20].

Now we discuss the situations where the observations are in two other noisy settings.

Theorem 3.2. *Consider the signal recovery model* (1.1) *with* $\|A^{\top}z\|_{2,\infty}\le\epsilon$, *where* $x\in\mathbb{R}^n$ *is block structural over the block index set* $\mathcal{I}=\{d_1,d_2,\cdots,d_m\}$. *Suppose that* $(\alpha,p,q)\in[0,1]\times(0,1]\times[1,2]\setminus(1,1,1)$, $\varepsilon\ge\epsilon$, $k\in\{1,2,\ldots,m\},k\neq\alpha$.

Under the hypothesis **(H)**, *if the matrix* A *satisfies the B-RIP condition* (3.4) *for any fixed* $t\in(d,+\infty)$, *then the minimizer of*

$$\hat{x}^{BDS}=\arg\min_{x\in\mathbb{R}^n}\left\{\|x\|_{2,p,\mathrm{w}}^p-\alpha\|x\|_{2,q,\mathrm{w}}^p:\ \|A^{\top}(y-Ax)\|_{2,\infty}\le\varepsilon\right\}$$

obeys

$$\|\hat{x}^{BDS}-x\|_{2,q}\le C_3(\varepsilon+\epsilon)+C_4\|x_{-\max(k)}\|_{2,p,w},\tag{3.14}$$

where

$$\begin{aligned}C_3&=\left(\frac{C_1}{\theta}\right)^2\left[(1+3^{\frac{1}{p}-1})(dk)^{1-\frac{1}{q}}+\left(\frac{3}{dk}\right)^{\frac{1}{p}-1}\alpha^{\frac{1}{p}}\right],\\ C_4&=\max\left\{\frac{C_2}{1-\theta},\frac{2^{1/p}}{(3^{1-1/p}+1)(dk)^{1/p-1/q}+\alpha^{1/p}}\right\}\end{aligned}\tag{3.15}$$

for any $\theta\in(0,1)$ *with* C_1 *and* C_2 *defined in* (3.7).

The following result in the third type of noise setting can be analogously obtained.

Theorem 3.3. *Consider the signal recovery model* (1.1) *with* $\|A^{\top}z\|_{\infty}\le\epsilon$, *where* $x\in\mathbb{R}^n$ *is block structural over the block index set* $\mathcal{I}=\{d_1,d_2,\cdots,d_m\}$. *Suppose that* $(\alpha,p,q)\in[0,1]\times(0,1]\times[1,2]\setminus(1,1,1),\varepsilon\ge\epsilon,k\in\{1,2,\ldots,m\},k\neq\alpha$.

Under the hypothesis **(H)**, *if the matrix* A *satisfies* (3.4) *for any fixed* $t\in(d,+\infty)$, *the minimizer of*

$$\hat{x}^{DS}=\arg\min_{x\in\mathbb{R}^n}\left\{\|x\|_{2,p,\mathrm{w}}^p-\alpha\|x\|_{2,q,\mathrm{w}}^p:\ \|A^{\top}(y-Ax)\|_{\infty}\le\varepsilon\right\}$$

obeys

$$\|\hat{x}^{DS}-x\|_{2,q}\le\sqrt{\max_i d_i}\,C_3(\varepsilon+\epsilon)+C_4\|x_{-\max(k)}\|_{2,p,w},\tag{3.16}$$

where C_3 *and* C_4 *are defined in* (3.15).

4. Numerical Experiments

In this section, we numerically verify the newly-derived results for successful recovery of block structural signals.

4.1. Experiments in ℓ_2-bounded noise setting

In previous related studies [11,15,18,19,37], a well-established computational methodology has been documented wherein constrained optimization problems are reformulated as unconstrained penalty problems through the strategic introduction of a regularization parameter. When the regularization parameter is set to a very large value, it effectively enforces the constraint as a strict requirement. This paradigm shift achieves computational facilitation by circumventing explicit constraint handling mechanisms, and achieves theoretical consistency through asymptotic equivalence when the penalty parameter exceeds critical thresholds.

We firstly transform the constrained problem (1.8) into the unconstrained penalty problem

$$\min_{x\in\mathbb{R}^N} \big(\|x\|_{2,p,w}^p - \alpha\|x\|_{2,q,w}^p\big) + \hat{\lambda}\|Ax-y\|_2^2. \tag{4.1}$$

Herein, $\hat{\lambda} > 0$ denotes the regularizer parameter. Clearly, this problem is equivalent to

$$\min_{x\in\mathbb{R}^N} \frac{1}{2}\|Ax-y\|_2^2 + \lambda\big(\|x\|_{2,p,w}^p - \alpha\|x\|_{2,q,w}^p\big), \tag{4.2}$$

where $\lambda > 0$. Inspired by [19], we use the iteratively reweighted least squares (IRLS) algorithm proposed in [8] to approximate $\|x\|_{2,p,w}^p$. We have

$$x^{(n+1)} = \arg\min_{x\in\mathbb{R}^N} \left(\frac{1}{2}\|Ax-y\|_2^2 + \lambda\|W^{(n)}x\|_2^2\right) - \lambda\alpha\|x^{(n)}\|_{2,q,w}^{p-1}\|x\|_{2,q,w}, \tag{4.3}$$

where

$$W^{(n)} = \operatorname{diag}\Big(w_i^{\frac{p}{2}}\big(\tau_n^2 + \|x^{(n)}[i]\|_2^2\big)^{\frac{p}{4}-\frac{1}{2}}\Big).$$

Now we are in a position to solve the problem via the difference of convex functions algorithm (DCA). Suppose that $[r(x)]_k$ is the k-th largest value in ℓ_2 norm of the blocks of x and $\gamma \in (0,1)$. We obtain

$$x^{(n+1)} = \big(A^\mathsf{T}A + 2\lambda(W^{(n)})^\mathsf{T}W^{(n)}\big)^{-1}\big(A^\mathsf{T}y + \lambda\alpha\|x^{(n)}\|_{2,q,\mathrm{w}}^{p-q}S^{(n)}x^{(n)}\big),$$

where

$$S^{(n)} = \operatorname{diag}\Big(w_i^q\big(\tau_n^2 + \|x^{(n)}[i]\|_2^2\big)^{\frac{q}{2}-1}\Big),$$
$$\tau_{n+1} = \min\big\{\tau_n,\ \gamma[r(x^{(n+1)})]_{\hat{k}+1}\big\}.$$

Set $\gamma = 0.9$ and $\lambda = 10^{-6}$. The performance is measured by the signal to noise ratio (SNR) defined as

$$\mathrm{SNR} = 20\log_{10}\frac{\|x\|_2}{\|x - x^{(n+1)}\|_2}. \tag{4.4}$$

By Lemma 2.4, we conclude that when m is sufficiently large, the sub-Gaussian random matrix satisfies the RIP condition with high probability. Consequently, according to Theorem 3.1, it is guaranteed that the block sparse signal can be successfully recovered. Therefore, in this part, we employ the sub-Gaussian random matrix A.

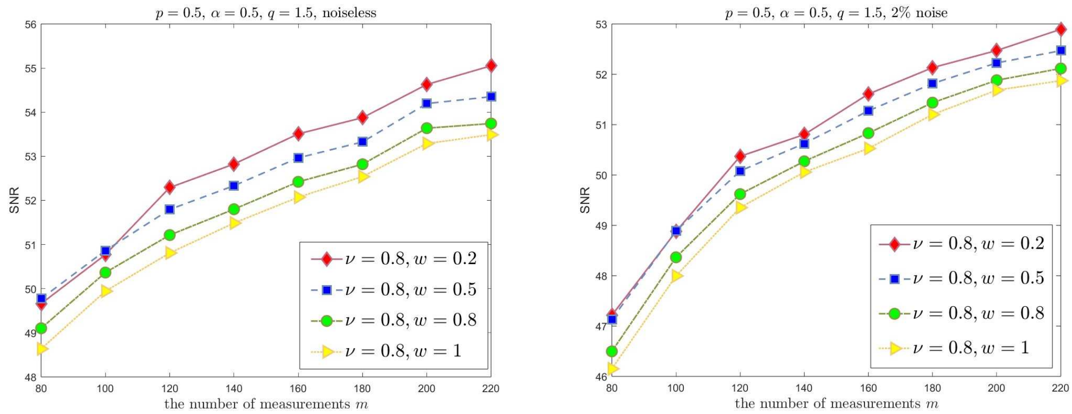
- By observing Fig. 4.1, we find out that whether in the noisy case or in the noiseless case, the performance of recovery is better by employing lower weight w when the block support accuracy is high. Moreover, the recovery performance of the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ minimization is superior to that of the $\ell_p - \alpha\ell_q$ minimization, and this result agrees with the theoretical analyses in Section 3.

- Fig. 4.2 visually suggests that when there are some prior block support estimates, the incorporation of non-uniform weighting can make more efficient use of the block support information of the original signal and result in more successful recovery.

- Fig. 4.3 reveals that utilizing the characteristic of block sparsity is necessarily beneficial for better performance of recovery for block structural signals.

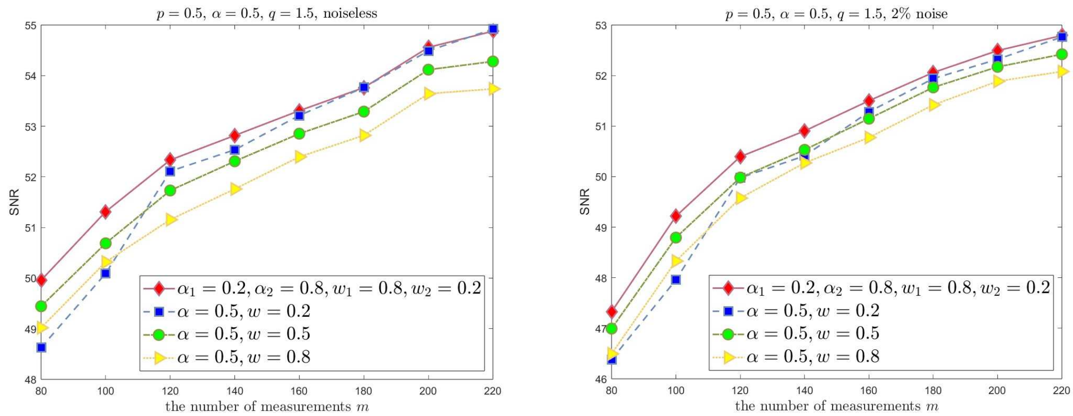
- As shown in Fig. 4.4, whether in the noisy case or in the noiseless case, the recovery performance is always more excellent for smaller $p \in (0, 1]$.

- From the observation of Fig. 4.5, we can find the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ minimization performs better in the comparisons of the weighted $\ell_{2,1} - \ell_2$, $\ell_{2,p}$, and $\ell_{2,p} - \alpha\ell_{2,q}$ minimization. Moreover, if the appropriate q is selected, the recovery performance can be better.



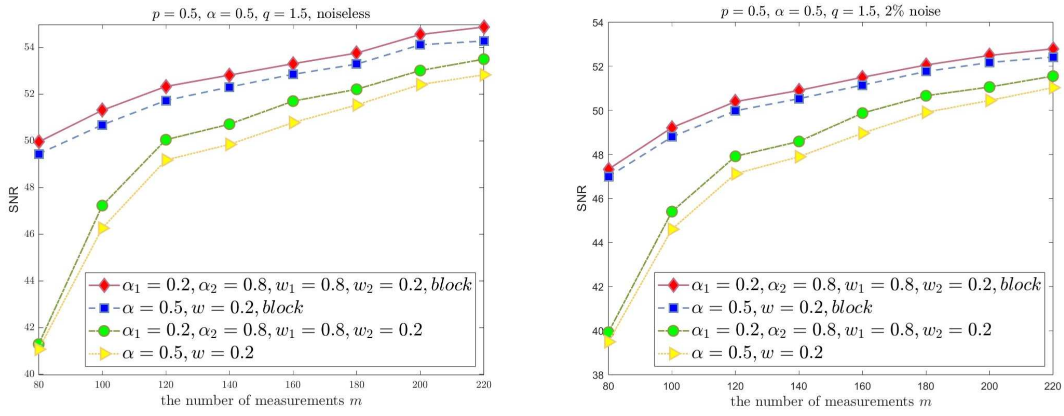
(a) (b)

Fig. 4.1. For $p = 0.5, \alpha = 0.5, q = 1.5$, the recovery performance of the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ minimization for block sparse signals in (a) the noiseless case, (b) the noisy case.



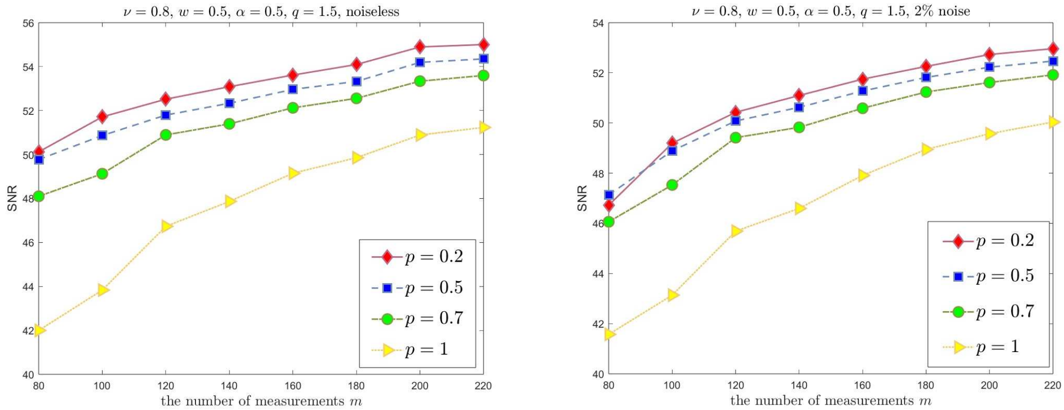
(a) (b)

Fig. 4.2. For $p = 0.5, \alpha = 0.5, q = 1.5$, the recovery performance of the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ minimization for block sparse signals in (a) the noiseless case, (b) the noisy case.



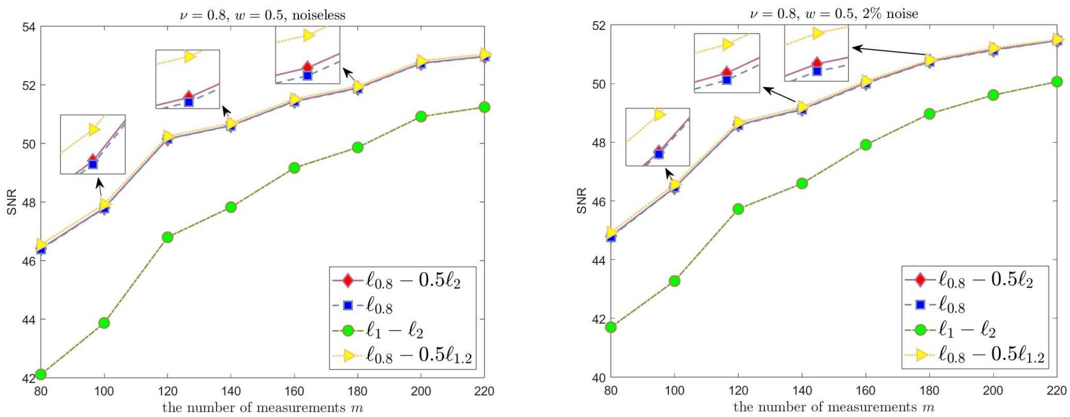
(a) (b)

Fig. 4.3. For $p = 0.5, \alpha = 0.5, q = 1.5$, the recovery performance of the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ minimization for block sparse signals in (a) the noiseless case, (b) the noisy case.



(a) (b)

Fig. 4.4. For $\nu = 0.8, w = 0.5, \alpha = 0.5, q = 1.5$, the recovery performance of the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ minimization for block sparse signals in (a) the noiseless case, (b) the noisy case.



(a) (b)

Fig. 4.5. For $\nu = 0.8, w = 0.5$, the comparisons of the weighted $\ell_{2,1} - \ell_2, \ell_{2,p}$, and $\ell_{2,p} - \alpha\ell_{2,q}$ minimization for block sparse signals in (a) the noiseless case, (b) the noisy case.

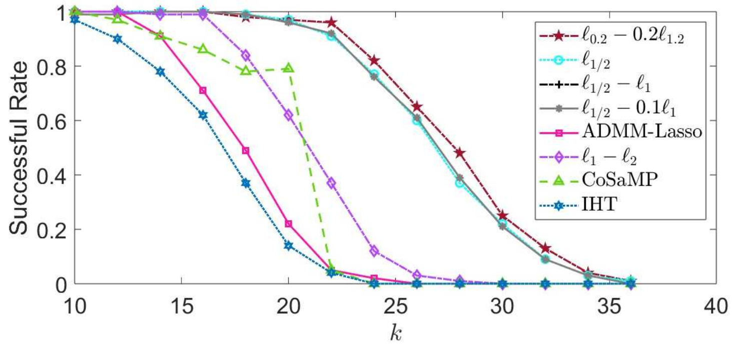


Fig. 4.6. Comparison of the performance of sparse signal recovery via different algorithms for Gaussian random matrix.

- To ensure the comparability of the results, we consider general signals. From the observation of Fig. 4.6, we observe that the weighted $\ell_{0.2} - 0.2\ell_{1.2}$ minimization method is better than other well-known algorithms, including $\ell_p - \alpha\ell_1$ minimization [41] (denoted as IRLS $\ell_{0.5}$, IRLS $\ell_{0.5} - \ell_1$), DCA $\ell_{0.5} - 0.1\ell_1$ [40], ADMM-Lasso [2], $\ell_1 - \ell_2$ minimization [25], CoSaMP [27], iterative hard thresholding (IHT) [1].

4.2. Experiments in DS-bounded noise setting and BDS-bounded noise setting

Inspired by [18], based on the fact that Dantzig selector and Lasso estimator exhibit similar behavior, we propose an unconstraint penalty problem as follows:

$$\min_{x,y\in\mathcal{B}^\infty(\hat{\eta})} \frac{1}{2}\|A^\mathsf{T}Ax - y - A^\mathsf{T}b\|_2^2 + \lambda\big(\|x\|_{2,p,w}^p - \alpha\|x\|_{2,q,w}^p\big), \tag{4.5}$$

where $\lambda > 0$ and

$$\mathcal{B}^\infty(\hat{\eta}) = \big\{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_\infty \leqslant \hat{\eta}\big\}.$$

Let $B = A^\mathsf{T}A$ and $\hat{c} = A^\mathsf{T}b$. Then (4.5) is equivalent to

$$\begin{aligned}&\min_{\hat{w},x,y\in\mathcal{B}^\infty(\hat{\eta})} \lambda\left(\|\hat{w}\|_{2,p,w}^p - \alpha\|\hat{w}\|_{2,q,w}^p\right) + \frac{1}{2}\|Bx - y - \hat{c}\|_2^2\\ &\text{s.t.} \quad x - \hat{w} = 0.\end{aligned} \tag{4.6}$$

We solve it directly using the alternating direction method of multipliers (ADMM) algorithm [35]. The augmented Lagrangian function of (4.6) is as follows:

$$\mathcal{L}(x,y,\hat{w};z) = \lambda\big(\|\hat{w}\|_{2,p,w}^p - \alpha\|\hat{w}\|_{2,q,w}^p\big) + \frac{1}{2}\|Bx - y - \hat{c}\|_2^2 + \frac{\hat{\beta}}{2}\|x - \hat{w}\|^2 + \langle z, x - w\rangle, \tag{4.7}$$

where z is the Lagrangian multiplier and $\hat{\beta} > 0$ is the regularized parameter. Therefore, we have

$$\begin{aligned}\hat{w}^{k+1} &= \mathrm{Prox}_{\frac{\lambda}{\hat{\beta}}\ell_{2,p} - \frac{\alpha}{\hat{\beta}}\ell_{2,q}}\left(x^k + \frac{z^k}{\hat{\beta}}\right)\\ &= \arg\min_{\hat{w}\in\mathbb{R}^N} \frac{1}{2}\left\|A\hat{w} - \left(x^k + \frac{z^k}{\hat{\beta}}\right)\right\|_2^2 + \frac{\lambda}{\hat{\beta}}\left(\|\hat{w}\|_{2,p,w}^p - \frac{\alpha}{\hat{\beta}}\|\hat{w}\|_{2,q,w}^p\right),\end{aligned} \tag{4.8}$$

$$x^{k+1} = (B^{\mathsf{T}}B + \hat{\beta}I)^{-1}\left(B^{\mathsf{T}}(y^k + \hat{c}) + \hat{\beta}\left(\hat{w}^{k+1} - \frac{z^k}{\hat{\beta}}\right)\right). \tag{4.9}$$

The iteration of y and z are

$$y_j^{k+1} = \min\big\{\max\{x_j, -\hat{\eta}\}, \hat{\eta}\big\}, \tag{4.10}$$

$$z^{k+1} = z^k + \hat{\beta}(x^{k+1} - \hat{w}^{k+1}). \tag{4.11}$$

Similar to DS-bounded noise setting, we need to solve the following unconstrained problem:

$$\min_{x,y\in\mathcal{B}^{2,\infty}(\hat{\eta})} \frac{1}{2}\|A^{\mathsf{T}}Ax - y - A^{\mathsf{T}}b\|_2^2 + \lambda\big(\|x\|_{2,p,w}^p - \alpha\|x\|_{2,q,w}^p\big), \tag{4.12}$$

where $\lambda > 0$ and

$$\mathcal{B}^{2,\infty}(\hat{\eta}) = \big\{\boldsymbol{x}\in\mathbb{R}^n : \|\boldsymbol{x}\|_{2,\infty} \leqslant \hat{\eta}\big\}.$$

Except

$$y^{k+1}[i] = \begin{cases} x[i], & \|x[i]\|_2 \le \hat{\eta}, \\ \dfrac{\hat{\eta}}{\|x[i]\|_2}x[i], & \|x[i]\|_2 > \hat{\eta}, \end{cases} \tag{4.13}$$

the other iterations are the same as in the DS-bounded noise setting.

Observing Fig. 4.7, we can find that appropriate p, q, α can make the recovery performance of the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ minimization better than that of the weighted $\ell_{2,p}$ minimization and the weighted $\ell_{2,1} - \ell_2$ minimization in DS-bounded noise setting and BDS-bounded noise setting.

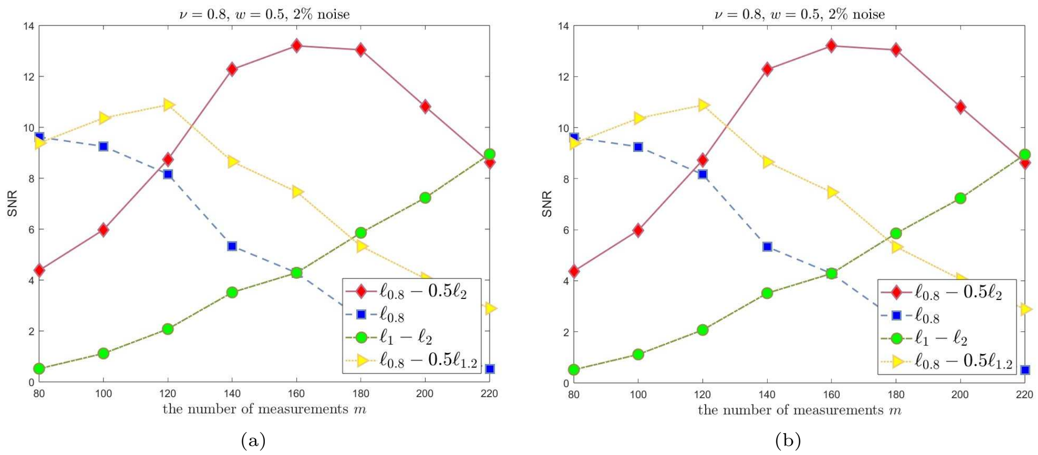


Fig. 4.7. For $\nu = 0.8, w = 0.5$, the comparisons of the weighted $\ell_{2,1} - \ell_2, \ell_{2,p}$, and $\ell_{2,p} - \alpha\ell_{2,q}$ minimization for block sparse signals in (a) DS-bounded noise setting, (b) BDS-bounded noise setting.

5. Conclusions

In block sparse signal recovery, we take into account non-uniform prior block support information, which always occurs in some important practical applications. We derive stable and robust recovery of non-strictly block k-sparse signals in different noise settings via the weighted

$\ell_{2,p} - \alpha\ell_{2,q}$ nonconvex minimization under arbitrary prior support information in noisy settings. The obtained results substantially generalize and improve the state-of-the-art results on block sparse signal recovery. By virtue of theoretical deduction and numerical experiments, we demonstrate that by exploiting the block sparsity and the known support information of original signals, better performance of recovery for block-structural signals can be achieved under more general conditions.

Appendix A

Proof of Lemma 2.5. Similar to the proof of Lemma 2.3. Denote

$$x_i = (0, \cdots, 0, x[i]^{\mathsf{T}}, 0, \cdots, 0)^{\mathsf{T}}.$$

Then $x = \sum_{i=1}^{m} x_i$. Suppose x is block sk-sparse signal, $\hat{x}_i \in \{x_j : \|x_j\|_2 \neq 0\}$ and $\hat{x}_i \hat{x}_j^{\mathsf{T}} = 0$ $(i \neq j)$. Without loss of generality, we can set $\hat{x}_i = 0$ if $|\{x_j : \|x_j\|_2 \neq 0\}| < sk$. Thus, $x = \sum_{i=1}^{sk} \hat{x}_i$. Then we obtain

$$\begin{aligned}
\|Ax\|_2^2 &= \left\| \sum_{i=1}^{sk} A\hat{x}_i \right\|_2^2 = \left\langle \sum_{i=1}^{sk} A\hat{x}_i, \sum_{j=1}^{sk} A\hat{x}_j \right\rangle = \sum_{i=1}^{sk}\sum_{j=1}^{sk} \langle A\hat{x}_i, A\hat{x}_j \rangle \\
&= \sum_{i=1}^{sk} \|A\hat{x}_i\|_2^2 + 2 \sum_{1\leqslant i<j\leqslant sk} \langle A\hat{x}_i, A\hat{x}_j \rangle \\
&= s\sum_{i=1}^{sk} \|A\hat{x}_i\|_2^2 - (s-1) \sum_{1\leqslant i\leqslant sk} \|A\hat{x}_i\|_2^2 + 2\left(1 - \frac{s-1}{sk-1} + \frac{s-1}{sk-1}\right) \sum_{1\leqslant i<j\leqslant sk} \langle A\hat{x}_i, A\hat{x}_j \rangle \\
&= s\sum_{i=1}^{sk} \|A\hat{x}_i\|_2^2 - \frac{s-1}{sk-1} \sum_{1\leqslant i\leqslant sk} (sk-1)\|A\hat{x}_i\|_2^2 + 2\left(1 - \frac{s-1}{sk-1} + \frac{s-1}{sk-1}\right) \sum_{1\leqslant i<j\leqslant sk} \langle A\hat{x}_i, A\hat{x}_j \rangle \\
&= s\sum_{i=1}^{sk} \|A\hat{x}_i\|_2^2 - \frac{s-1}{sk-1} \sum_{1\leqslant i\leqslant sk} [(sk-i)+(i-1)]\|A\hat{x}_i\|_2^2 \\
&\quad + 2\left(1 - \frac{s-1}{sk-1}\right) \sum_{1\leqslant i<j\leqslant sk} \langle A\hat{x}_i, A\hat{x}_j \rangle + \frac{s-1}{sk-1} \sum_{1\leqslant i<j\leqslant sk} 2\langle A\hat{x}_i, A\hat{x}_j \rangle \\
&= s\sum_{i=1}^{sk} \|A\hat{x}_i\|_2^2 - \frac{s-1}{sk-1} \sum_{1\leqslant i\leqslant sk} \sum_{i<j\leqslant sk} \|A\hat{x}_i\|_2^2 - \frac{s-1}{sk-1} \sum_{1\leqslant i\leqslant sk} \sum_{1\leq j<i} \|A\hat{x}_i\|_2^2 \\
&\quad + 2\left(1 - \frac{s-1}{sk-1}\right) \sum_{1\leqslant i<j\leqslant sk} \langle A\hat{x}_i, A\hat{x}_j \rangle + \frac{s-1}{sk-1} \sum_{1\leqslant i<j\leqslant sk} 2\langle A\hat{x}_i, A\hat{x}_j \rangle \\
&= s\sum_{i=1}^{sk} \|A\hat{x}_i\|_2^2 - \frac{s-1}{sk-1} \sum_{1\leqslant i\leqslant sk} \sum_{i<j\leqslant sk} \|A\hat{x}_i\|_2^2 - \frac{s-1}{sk-1} \sum_{1\leqslant j\leqslant sk} \sum_{1\leq i<j} \|A\hat{x}_j\|_2^2 \\
&\quad + 2\left(1 - \frac{s-1}{sk-1}\right) \sum_{1\leqslant i<j\leqslant sk} \langle A\hat{x}_i, A\hat{x}_j \rangle + \frac{s-1}{sk-1} \sum_{1\leqslant i<j\leqslant sk} 2\langle A\hat{x}_i, A\hat{x}_j \rangle \\
&= s\sum_{i=1}^{sk} \|A\hat{x}_i\|_2^2 - \frac{s-1}{sk-1} \sum_{1\leqslant i\leqslant sk} \sum_{i<j\leqslant sk} \|A\hat{x}_i\|_2^2 - \frac{s-1}{sk-1} \sum_{1\leqslant i\leqslant sk} \sum_{i<j\leqslant sk} \|A\hat{x}_j\|_2^2
\end{aligned}$$

$$
\begin{aligned}
&+2\Big(1-\frac{s-1}{sk-1}\Big)\sum_{1\leqslant i<j\leqslant sk}\langle A\hat{x}_i, A\hat{x}_j\rangle+\frac{s-1}{sk-1}\sum_{1\leqslant i<j\leqslant sk}2\langle A\hat{x}_i, A\hat{x}_j\rangle\\
&=s\sum_{i=1}^{sk}\|A\hat{x}_i\|_2^2+2\Big(1-\frac{s-1}{sk-1}\Big)\sum_{1\leqslant i<j\leqslant sk}\langle A\hat{x}_i, A\hat{x}_j\rangle\\
&\quad-\frac{s-1}{sk-1}\sum_{1\leqslant i<j\leqslant sk}\big(\|A\hat{x}_i\|_2^2+\|A\hat{x}_j\|_2^2-2\langle A\hat{x}_i, A\hat{x}_j\rangle\big)\\
&=s\sum_{i=1}^{sk}\|A\hat{x}_i\|_2^2+2\Big(1-\frac{s-1}{sk-1}\Big)\sum_{1\leqslant i<j\leqslant sk}\langle A\hat{x}_i, A\hat{x}_j\rangle-\frac{s-1}{sk-1}\sum_{1\leqslant i<j\leqslant sk}\|A\hat{x}_i-A\hat{x}_j\|_2^2\\
&=\frac{s^2}{\binom{sk}{k}}\sum_{i=1}^{sk}\frac{\binom{sk}{k}}{s}\|A\hat{x}_i\|_2^2+2\frac{s^2}{\binom{sk}{k}}\sum_{1\leqslant i<j\leqslant sk}\frac{\binom{sk}{k}(k-1)}{s(sk-1)}\langle A\hat{x}_i, A\hat{x}_j\rangle\\
&\quad-\frac{s-1}{sk-1}\sum_{1\leqslant i<j\leqslant sk}\|A\hat{x}_i-A\hat{x}_j\|_2^2\\
&=\frac{s^2}{\binom{sk}{k}}\Bigg[\sum_{i=1}^{sk}\binom{sk-1}{k-1}\|A\hat{x}_i\|_2^2+\sum_{1\leqslant i<j\leqslant sk}\binom{sk-2}{k-2}2\langle A\hat{x}_i, A\hat{x}_j\rangle\Bigg]\\
&\quad-\frac{s-1}{sk-1}\sum_{1\leqslant i<j\leqslant sk}\|A\hat{x}_i-A\hat{x}_j\|_2^2\\
&=\frac{s^2}{\binom{sk}{k}}\sum_{1\leqslant i_1<\cdots<i_k\leqslant sk}\|A\hat{x}_{i_1}+A\hat{x}_{i_2}+\cdots+A\hat{x}_{i_k}\|_2^2-\frac{s-1}{sk-1}\sum_{1\leqslant i<j\leqslant sk}\|A\hat{x}_i-A\hat{x}_j\|_2^2. \quad \text{(A.1)}
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
\|Ax\|_2^2 &\le \frac{s^2(1+\delta_{k|\mathcal{I}})}{\binom{sk}{k}}\sum_{1\leqslant i_1<\cdots<i_k\leqslant sk}\big(\|\hat{x}_{i_1}\|_2^2+\cdots+\|\hat{x}_{i_k}\|_2^2\big)\\
&\quad-\frac{(s-1)(1-\delta_{k|\mathcal{I}})}{sk-1}\sum_{1\leqslant i<j\leqslant sk}\big(\|\hat{x}_i\|_2^2+\|\hat{x}_j\|_2^2\big)\\
&=\frac{s^2(1+\delta_{k|\mathcal{I}})}{\binom{sk}{k}}\binom{sk-1}{k-1}\sum_{i=1}^{sk}\|\hat{x}_i\|_2^2-\frac{(s-1)(1-\delta_{k|\mathcal{I}})}{sk-1}\sum_{i=1}^{sk}[(sk-i)+(i-1)]\|\hat{x}_i\|_2^2\\
&=\big(s(1+\delta_{k|\mathcal{I}})-(s-1)(1-\delta_{k|\mathcal{I}})\big)\sum_{i=1}^{sk}\|\hat{x}_i\|_2^2\\
&=\big(1+(2s-1)\delta_{k|\mathcal{I}}\big)\|x\|_2^2, \quad \text{(A.2)}\\
\|Ax\|_2^2 &\ge \frac{s^2(1-\delta_{k|\mathcal{I}})}{\binom{sk}{k}}\sum_{1\leqslant i_1<\cdots<i_k\leqslant sk}\big(\|\hat{x}_{i_1}\|_2^2+\cdots+\|\hat{x}_{i_k}\|_2^2\big)\\
&\quad-\frac{(s-1)(1+\delta_{k|\mathcal{I}})}{sk-1}\sum_{1\leqslant i<j\leqslant sk}\big(\|\hat{x}_i\|_2^2+\|\hat{x}_j\|_2^2\big)\\
&=\big(s(1-\delta_{k|\mathcal{I}})-(s-1)(1+\delta_{k|\mathcal{I}})\big)\sum_{i=1}^{sk}\|\hat{x}_i\|_2^2\\
&=\big(1-(2s-1)\delta_{k|\mathcal{I}}\big)\|x\|_2^2. \quad \text{(A.3)}
\end{aligned}
$$

Therefore, $\delta_{sk|\mathcal{I}} \le (2s-1)\delta_{k|\mathcal{I}}$. □

Appendix B

Proof of Lemma 2.6. Since $\hat{x} = x + h$ is the minimizer of the weighted $\ell_{2,p} - \alpha\ell_{2,q}$ nonconvex minimization (1.8), we have

$$\|x+h\|_{2,p,\mathrm{w}}^p - \alpha\|x+h\|_{2,q,\mathrm{w}}^p \le \|x\|_{2,p,\mathrm{w}}^p - \alpha\|x\|_{2,q,\mathrm{w}}^p. \tag{B.1}$$

By the triangle inequalities for $q \in [1,2]$ and $p \in (0,1]$, we obtain

$$\|x+h\|_{2,q,\mathrm{w}}^p \le (\|x\|_{2,q,\mathrm{w}} + \|h\|_{2,q,\mathrm{w}})^p \le \|x\|_{2,q,\mathrm{w}}^p + \|h\|_{2,q,\mathrm{w}}^p \le \|x\|_{2,q,\mathrm{w}}^p + \|h\|_{2,q}^p. \tag{B.2}$$

Therefore,

$$\begin{aligned} &\|x+h\|_{2,p,\mathrm{w}}^p - \alpha\big(\|x\|_{2,q,\mathrm{w}}^p + \|h\|_{2,q}^p\big) \\ &\le \|x+h\|_{2,p,\mathrm{w}}^p - \alpha\|x+h\|_{2,q,\mathrm{w}}^p \le \|x\|_{2,p,\mathrm{w}}^p - \alpha\|x\|_{2,q,\mathrm{w}}^p, \end{aligned} \tag{B.3}$$

and thus,

$$\|x+h\|_{2,p,\mathrm{w}}^p \le \|x\|_{2,p,\mathrm{w}}^p + \alpha\|h\|_{2,q}^p. \tag{B.4}$$

Applying the reverse triangle inequalities yields

$$\|x_{T_0} + h_{T_0}\|_{2,p,\mathrm{w}}^p \ge \|x_{T_0}\|_{2,p,\mathrm{w}}^p - \|h_{T_0}\|_{2,p,\mathrm{w}}^p, \tag{B.5}$$

$$\big\|x_{T_0^\complement} + h_{T_0^\complement}\big\|_{2,p,\mathrm{w}}^p \ge \big\|h_{T_0^\complement}\big\|_{2,p,\mathrm{w}}^p - \big\|x_{T_0^\complement}\big\|_{2,p,\mathrm{w}}^p. \tag{B.6}$$

By virtue of the combination of (B.4)-(B.6), we deduce

$$\begin{aligned} &\|x\|_{2,p,\mathrm{w}}^p + \alpha\|h\|_{2,q}^p \\ &\ge \|x+h\|_{2,p,\mathrm{w}}^p = \|x_{T_0} + h_{T_0}\|_{2,p,\mathrm{w}}^p + \big\|x_{T_0^\complement} + h_{T_0^\complement}\big\|_{2,p,\mathrm{w}}^p \\ &\ge \|x_{T_0}\|_{2,p,\mathrm{w}}^p - \|h_{T_0}\|_{2,p,\mathrm{w}}^p + \big\|h_{T_0^\complement}\big\|_{2,p,\mathrm{w}}^p - \big\|x_{T_0^\complement}\big\|_{2,p,\mathrm{w}}^p \\ &= \|x\|_{2,p,\mathrm{w}}^p - 2\big\|x_{T_0^\complement}\big\|_{2,p,\mathrm{w}}^p - \|h_{T_0}\|_{2,p,\mathrm{w}}^p + \big\|h_{T_0^\complement}\big\|_{2,p,\mathrm{w}}^p, \end{aligned} \tag{B.7}$$

and thus,

$$\big\|h_{T_0^\complement}\big\|_{2,p,\mathrm{w}}^p \le \|h_{T_0}\|_{2,p,\mathrm{w}}^p + 2\big\|x_{T_0^\complement}\big\|_{2,p,\mathrm{w}}^p + \alpha\|h\|_{2,q}^p. \tag{B.8}$$

We derive

$$\begin{aligned} \|h_{T_0}\|_{2,p,\mathrm{w}}^p &= \sum_{i=1}^{L} w_i^p \|h_{T_0\cap\tilde{T}_i}\|_{2,p}^p + \|h_{T_0\cap\tilde{T}^\complement}\|_{2,p}^p \\ &= w_L^p\|h_{T_0\cap\tilde{T}}\|_{2,p}^p + \sum_{i=1}^{L-1}\big(w_i^p - w_L^p\big)\|h_{T_0\cap\tilde{T}_i}\|_{2,p}^p + \|h_{T_0\cap\tilde{T}^\complement}\|_{2,p}^p \\ &= w_L^p\|h_{T_0}\|_{2,p}^p + \sum_{i=1}^{L-1}\big(w_i^p - w_L^p\big)\|h_{T_0\cap\tilde{T}_i}\|_{2,p}^p + \big(1 - w_L^p\big)\big\|h_{T_0\cap\tilde{T}^\complement}\big\|_{2,p}^p, \end{aligned} \tag{B.9}$$

$$\big\|h_{T_0^\complement}\big\|_{2,p,\mathrm{w}}^p = \big\|h_{T_0^\complement}\big\|_{2,p}^p - \sum_{i=1}^{L}\big(1 - w_i^p\big)\big\|h_{T_0^\complement\cap\tilde{T}_i}\big\|_{2,p}^p. \tag{B.10}$$

Concatenating (B.8)-(B.10) leads to

$$\begin{aligned}
\left\|h_{T_0^{\complement}}\right\|_{2,p}^p &\le w_L^p\|h_{T_0}\|_{2,p}^p+\sum_{i=1}^{L}\left(1-w_i^p\right)\left\|h_{T_0^{\complement}\cap\widetilde{T}_i}\right\|_{2,p}^p+\sum_{i=1}^{L-1}\left(w_i^p-w_L^p\right)\|h_{T_0\cap\widetilde{T}_i}\|_{2,p}^p\\
&\quad+\left(1-w_L^p\right)\left\|h_{T_0\cap\widetilde{T}^{\complement}}\right\|_{2,p}^p+2\left\|x_{T_0^{\complement}}\right\|_{2,p,\mathrm{w}}^p+\alpha\|h\|_{2,q}^p\\
&=w_L^p\|h_{T_0}\|_{2,p}^p+\left(1-w_1^p\right)\left\|h_{(T_0\cup\widetilde{T})\setminus\bigcup_{i=1}^{L}(\widetilde{T}_i\cap T_0)}\right\|_{2,p}^p\\
&\quad+\sum_{i=2}^{L}\left[\sum_{j=2}^{i}\left(w_{j-1}^p-w_j^p\right)\right]\|h_{\widetilde{T}_i\cap T_0^{\complement}}\|_{2,p}^p+\sum_{i=2}^{L}\left[\sum_{j=i}^{L}\left(w_{j-1}^p-w_j^p\right)\right]\|h_{T_0\cap\widetilde{T}_{i-1}}\|_{2,p}^p\\
&\quad+\left[\sum_{j=2}^{L}\left(w_{j-1}^p-w_j^p\right)\right]\|h_{T_0\cap\widetilde{T}^{\complement}}\|_{2,p}^p+2\left\|x_{T_0^{\complement}}\right\|_{2,p,\mathrm{w}}^p+\alpha\|h\|_{2,q}^p\\
&=w_L^p\|h_{T_0}\|_{2,p}^p+\left(1-w_1^p\right)\left\|h_{(T_0\cup\widetilde{T})\setminus\bigcup_{i=1}^{L}(\widetilde{T}_i\cap T_0)}\right\|_{2,p}^p\\
&\quad+\sum_{j=2}^{L}\sum_{i=j}^{L}\left(w_{j-1}^p-w_j^p\right)\|h_{\widetilde{T}_i\cap T_0^{\complement}}\|_{2,p}^p+\sum_{j=2}^{L}\sum_{i=2}^{j}\left(w_{j-1}^p-w_j^p\right)\|h_{T_0\cap\widetilde{T}_{i-1}}\|_{2,p}^p\\
&\quad+\sum_{j=2}^{L}\left(w_{j-1}^p-w_j^p\right)\|h_{T_0\cap\widetilde{T}^{\complement}}\|_{2,p}^p+2\left\|x_{T_0^{\complement}}\right\|_{2,p,\mathrm{w}}^p+\alpha\|h\|_{2,q}^p\\
&=w_L^p\|h_{T_0}\|_{2,p}^p+\left(1-w_1^p\right)\left\|h_{(T_0\cup\widetilde{T})\setminus\bigcup_{i=1}^{L}(\widetilde{T}_i\cap T_0)}\right\|_{2,p}^p\\
&\quad+\sum_{j=2}^{L}\sum_{i=j}^{L}\left(w_{j-1}^p-w_j^p\right)\|h_{\widetilde{T}_i\cap T_0^{\complement}}\|_{2,p}^p+\sum_{j=2}^{L}\sum_{i=1}^{j-1}\left(w_{j-1}^p-w_j^p\right)\|h_{T_0\cap\widetilde{T}_i}\|_{2,p}^p\\
&\quad+\sum_{j=2}^{L}\left(w_{j-1}^p-w_j^p\right)\|h_{T_0\cap\widetilde{T}^{\complement}}\|_{2,p}^p+2\left\|x_{T_0^{\complement}}\right\|_{2,p,\mathrm{w}}^p+\alpha\|h\|_{2,q}^p\\
&=w_L^p\|h_{T_0}\|_{2,p}^p+\left(1-w_1^p\right)\left\|h_{(T_0\cup\widetilde{T})\setminus\bigcup_{i=1}^{L}(\widetilde{T}_i\cap T_0)}\right\|_{2,p}^p+\alpha\|h\|_{2,q}^p\\
&\quad+\sum_{j=2}^{L}\left(w_{j-1}^p-w_j^p\right)\left(\sum_{i=j}^{L}\|h_{\widetilde{T}_i\cap T_0^{\complement}}\|_{2,p}^p+\sum_{i=1}^{j-1}\|h_{T_0\cap\widetilde{T}_i}\|_{2,p}^p+\left\|h_{T_0\cap\widetilde{T}^{\complement}}\right\|_{2,p}^p\right)+2\left\|x_{T_0^{\complement}}\right\|_{2,p,\mathrm{w}}^p\\
&=w_L^p\|h_{T_0}\|_{2,p}^p+\left(1-w_1^p\right)\left\|h_{(T_0\cup\widetilde{T})\setminus\bigcup_{i=1}^{L}(\widetilde{T}_i\cap T_0)}\right\|_{2,p}^p+\alpha\|h\|_{2,q}^p\\
&\quad+\sum_{j=2}^{L}\left(w_{j-1}^p-w_j^p\right)\left[\left\|h_{\left(\bigcup_{i=j}^{L}\widetilde{T}_i\right)\cap T_0^{\complement}}\right\|_{2,p}^p+\left\|h_{T_0\cap\left(\bigcup_{i=j}^{L}\widetilde{T}_i\right)^{\complement}}\right\|_{2,p}^p\right]+2\left\|x_{T_0^{\complement}}\right\|_{2,p,\mathrm{w}}^p\\
&=w_L^p\|h_{T_0}\|_{2,p}^p+\left(1-w_1^p\right)\left\|h_{(T_0\cup\widetilde{T})\setminus\bigcup_{i=1}^{L}(\widetilde{T}_i\cap T_0)}\right\|_{2,p}^p+\alpha\|h\|_{2,q}^p\\
&\quad+\sum_{j=2}^{L}\left(w_{j-1}^p-w_j^p\right)\left\|h_{\left(T_0\cup\bigcup_{i=j}^{L}\widetilde{T}_i\right)\setminus\bigcup_{i=j}^{L}(\widetilde{T}_i\cap T_0)}\right\|_{2,p}^p+2\left\|x_{T_0^{\complement}}\right\|_{2,p,\mathrm{w}}^p. \qquad \text{(B.11)}
\end{aligned}$$

The proof is complete. □

Appendix C

Proof of Lemma 2.7. For $t \in (\hat{d}, 2\hat{d}]$, we have $2(t-\hat{d})k \leq tk$, and thus $\delta_{2(t-\hat{d})k|\mathcal{I}} \leq \delta_{tk|\mathcal{I}}$.

For $t > 2\hat{d}$, we have $2(t-\hat{d})/t > 1$. Then we have $tk \geq 2$ is an integer and

$$\frac{2(t-\hat{d})}{t} \cdot tk = 2(t-\hat{d})k$$

is an integer. Therefore, by virtue of Lemma 2.5, we get

$$\delta_{2(t-\hat{d})k|\mathcal{I}} \leq \left(\frac{4(t-\hat{d})}{t} - 1\right)\delta_{tk|\mathcal{I}} = \frac{3t-4\hat{d}}{t}\delta_{tk|\mathcal{I}}. \tag{C.1}$$

The proof is complete. □

Appendix D

Proof of Lemma 2.8. For $K \subseteq \{1, 2, \ldots, m\}$, we denote

$$K_1 = \left\{ i \in K^{\complement} : \|h[i]\|_2 > \frac{\|h_{K^{\complement}}\|_{2,p}}{[(t-d)k]^{1/p}} \right\}, \tag{D.1}$$

$$K_2 = \left\{ i \in K^{\complement} : \|h[i]\|_2 \leq \frac{\|h_{K^{\complement}}\|_{2,p}}{[(t-d)k]^{1/p}} \right\}. \tag{D.2}$$

We obtain $h_{K^{\complement}} = h_{K_1} + h_{K_2}$ and

$$\|h_{K^{\complement}}\|_{2,p}^p = \|h_{K_1} + h_{K_2}\|_{2,p}^p \geq \|h_{K_1}\|_{2,p}^p > |K_1|\frac{\|h_{K^{\complement}}\|_{2,p}^p}{(t-d)k},$$

and thus $|K_1| < (t-d)k$.

We have

$$\|h_{K_2}\|_{2,p}^p = \|h_{K^{\complement}}\|_{2,p}^p - \|h_{K_1}\|_{2,p}^p \leq [(t-d)k - |K_1|]\frac{\|h_{K^{\complement}}\|_{2,p}^p}{(t-d)k}. \tag{D.3}$$

By the definition of K_2, that is,

$$K_2 = \left\{ i \in K^{\complement} : \|h[i]\|_2 \leq \frac{\|h_{K^{\complement}}\|_{2,p}}{[(t-d)k]^{1/p}} \right\},$$

we obtain

$$\|h_{K_2}\|_{2,\infty} = \max\{\|h[i]\|_2 : i \in K_2\} \leq \frac{\|h_{K^{\complement}}\|_{2,p}}{[(t-d)k]^{1/p}}. \tag{D.4}$$

By virtue of Lemma 2.1, we deduce that $h_{K_2} = \sum_i \lambda_i u_i$, where $\lambda_i > 0, \sum_i \lambda_i = 1$,

$$\|u_i\|_{2,0} \leq (t-d)k - |K_1|, \quad \sum_{i=1}^{M} \lambda_i \|u_i\|_2^2 \leq \frac{\|h_{K^{\complement}}\|_{2,p}^p}{(t-d)k}\|h_{K_2}\|_{2,2-p}^{2-p}.$$

Therefore, by Hölder inequality,

$$\frac{1}{(2-p)/\big(2(1-p)\big)} + \frac{1}{(2-p)/p} = 1,$$

and $K_2 \subseteq K^{\complement}$, we obtain

$$\begin{aligned}\sum_{i=1}^{M}\lambda_i\|u_i\|_2^2 &\le \frac{\|h_{K^{\complement}}\|_{2,p}^p}{(t-d)k}\|h_{K_2}\|_{2,2-p}^{2-p} = \frac{\|h_{K^{\complement}}\|_{2,p}^p}{(t-d)k}\sum_{i=1}^{m}\|h_{K_2[i]}\|_2^{2-p}\\ &= \frac{\|h_{K^{\complement}}\|_{2,p}^p}{(t-d)k}\sum_{i=1}^{m}\|h_{K_2[i]}\|_2^{\frac{4(1-p)}{2-p}}\|h_{K_2[i]}\|_2^{\frac{p^2}{2-p}}\\ &\le \frac{\|h_{K^{\complement}}\|_{2,p}^p}{(t-d)k}\left(\sum_{i=1}^{m}\|h_{K_2[i]}\|_2^2\right)^{\frac{2(1-p)}{2-p}}\left(\sum_{i=1}^{m}\|h_{K_2[i]}\|_2^p\right)^{\frac{p}{2-p}}\\ &= \frac{\|h_{K^{\complement}}\|_{2,p}^p}{(t-d)k}\left(\|h_{K_2}\|_2^2\right)^{\frac{2(1-p)}{2-p}}\left(\|h_{K_2}\|_{2,p}^p\right)^{\frac{p}{2-p}}\\ &\le \frac{\|h_{K^{\complement}}\|_{2,p}^p}{(t-d)k}\left(\|h_{K_2}\|_2^2\right)^{\frac{2(1-p)}{2-p}}\left(\|h_{K^{\complement}}\|_{2,p}^p\right)^{\frac{p}{2-p}}\\ &= \frac{1}{(t-d)k}\left(\|h_{K_2}\|_2^2\right)^{\frac{2(1-p)}{2-p}}\left(\|h_{K^{\complement}}\|_{2,p}^p\right)^{\frac{2}{2-p}}. \end{aligned} \tag{D.5}$$

Denote $\beta_i = h_K + h_{K_1} + \kappa u_i, i = 1, 2, \ldots, M$, for any $\kappa \in [0, 1]$. Then

$$\sum_{j=1}^{M}\lambda_j\beta_j - c\beta_i = (1-c-\kappa)(h_K + h_{K_1}) - c\kappa u_i + \kappa h. \tag{D.6}$$

Combining $\sum_i \lambda_i = 1$, it is easy to get

$$\begin{aligned}&\sum_{i=1}^{M}\lambda_i\left\|A\left(\sum_{j=1}^{M}\lambda_j\beta_j - c\beta_i\right)\right\|_2^2\\ &= \sum_{i=1}^{M}\lambda_i\left(\left\|\sum_{j=1}^{M}\lambda_jA\beta_j\right\|_2^2 + c^2\|A\beta_i\|_2^2 - 2c\left\langle\sum_{j=1}^{M}\lambda_jA\beta_j, A\beta_i\right\rangle\right)\\ &= \left\|\sum_{j=1}^{M}\lambda_jA\beta_j\right\|_2^2 + c^2\sum_{i=1}^{M}\lambda_i\|A\beta_i\|_2^2 - 2c\left\langle\sum_{j=1}^{M}\lambda_jA\beta_j, \sum_{i=1}^{M}\lambda_iA\beta_i\right\rangle\\ &= (1-2c)\left\|\sum_{j=1}^{M}\lambda_jA\beta_j\right\|_2^2 + c^2\sum_{i=1}^{M}\lambda_i\|A\beta_i\|_2^2, \end{aligned} \tag{D.7}$$

$$\begin{aligned}&\sum_{i=1}^{M}\sum_{j=1}^{M}\lambda_i\lambda_j\|A(\beta_i-\beta_j)\|_2^2\\ &= \sum_{i=1}^{M}\sum_{j=1}^{M}\lambda_i\lambda_j\left(\|A\beta_i\|_2^2 + \|A\beta_j\|_2^2 - 2\langle A\beta_j, A\beta_i\rangle\right)\\ &= 2\sum_{i=1}^{M}\lambda_i\|A\beta_i\|_2^2 - 2\left\langle\sum_{j=1}^{M}\lambda_jA\beta_j, \sum_{i=1}^{M}\lambda_iA\beta_i\right\rangle\\ &= 2\sum_{i=1}^{M}\lambda_i\|A\beta_i\|_2^2 - 2\left\|\sum_{j=1}^{M}\lambda_jA\beta_j\right\|_2^2. \end{aligned} \tag{D.8}$$

Combining (D.7) and (D.8), we obtain

$$\begin{aligned}&\sum_{i=1}^{M}\lambda_i\left\|A\left(\sum_{j=1}^{M}\lambda_j\beta_j - c\beta_i\right)\right\|_2^2 + \frac{1-2c}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\lambda_i\lambda_j\|A(\beta_i-\beta_j)\|_2^2\\ &= (1-c)^2\sum_{i=1}^{M}\lambda_i\|A\beta_i\|_2^2. \end{aligned}\tag{D.9}$$

Combining (D.6), $\sum_i \lambda_i = 1$ and $h - (h_K + h_{K_1}) = h_{K_2} = \sum_i \lambda_i u_i$, we obtain

$$\begin{aligned}&\sum_{i=1}^{M}\lambda_i\left\|A\left(\sum_{j=1}^{M}\lambda_j\beta_j - c\beta_i\right)\right\|_2^2\\
&= \sum_{i=1}^{M}\lambda_i\|A[(1-c-\kappa)(h_K+h_{K_1}) - c\kappa u_i + \kappa h]\|_2^2\\
&= \sum_{i=1}^{M}\lambda_i\|A[(1-c-\kappa)(h_K+h_{K_1}) - c\kappa u_i]\|_2^2 + \kappa^2\|Ah\|_2^2\\
&\quad + 2\sum_{i=1}^{M}\lambda_i\langle A(1-c-\kappa)(h_K+h_{K_1}) - c\kappa A u_i, \kappa Ah\rangle\\
&= \sum_{i=1}^{M}\lambda_i\|A[(1-c-\kappa)(h_K+h_{K_1}) - c\kappa u_i]\|_2^2 + \kappa^2\|Ah\|_2^2\\
&\quad + 2(1-c-\kappa)\kappa\langle A(h_K+h_{K_1}), Ah\rangle - 2c\kappa^2\langle A\sum_{i=1}^{M}\lambda_i u_i, Ah\rangle\\
&= \sum_{i=1}^{M}\lambda_i\|A[(1-c-\kappa)(h_K+h_{K_1}) - c\kappa u_i]\|_2^2 + \kappa^2\|Ah\|_2^2\\
&\quad + 2(1-c-\kappa)\kappa\langle A(h_K+h_{K_1}), Ah\rangle - 2c\kappa^2\langle Ah - A(h_K+h_{K_1}), Ah\rangle\\
&= \sum_{i=1}^{M}\lambda_i\|A[(1-c-\kappa)(h_K+h_{K_1}) - c\kappa u_i]\|_2^2 + \kappa^2\|Ah\|_2^2\\
&\quad + 2(1-c-\kappa)\kappa\langle A(h_K+h_{K_1}), Ah\rangle + 2c\kappa^2\langle A(h_K+h_{K_1}), Ah\rangle - c\kappa^2\|Ah\|_2^2\\
&= \sum_{i=1}^{M}\lambda_i\|A[(1-c-\kappa)(h_K+h_{K_1}) - c\kappa u_i]\|_2^2 + \kappa^2(1-2c)\|Ah\|_2^2\\
&\quad + 2(1-c)\kappa(1-\kappa)\langle A(h_K+h_{K_1}), Ah\rangle.\end{aligned}\tag{D.10}$$

Thus combining (D.9) and (D.10), we have

$$\begin{aligned}0 &= \sum_{i=1}^{M}\lambda_i\|A[(1-c-\kappa)(h_K+h_{K_1}) - c\kappa u_i]\|_2^2\\
&\quad + \frac{1-2c}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\lambda_i\lambda_j\|A(\beta_i-\beta_j)\|_2^2 - (1-c)^2\sum_{i=1}^{M}\lambda_i\|A\beta_i\|_2^2\\
&\quad + 2(1-c)\kappa(1-\kappa)\langle A(h_K+h_{K_1}), Ah\rangle + (1-2c)\kappa^2\|Ah\|_2^2.\end{aligned}\tag{D.11}$$

Since $|K| \leq dk, |K_1| < (t-d)k$ and $\|u_i\|_{2,0} \leq (t-d)k - |K_1|$, we know that $h_K + h_{K_1}, u_i$ and $(1-c-\kappa)(h_K + h_{K_1}) - c\kappa u_i$ are all block tk-sparse and $\beta_i - \beta_j = \kappa u_i - \kappa u_j$ are all block $2(t-d)k$-sparse. By virtue of the definition of the B-RIC and Lemma 2.7, we derive

$$\begin{aligned}
&\sum_{i=1}^{M} \lambda_i \|A[(1-c-\kappa)(h_K + h_{K_1}) - c\kappa u_i]\|_2^2 \\
&\leq (1+\delta_{tk|\mathcal{I}}) \left[(1-c-\kappa)^2 \|h_K + h_{K_1}\|_2^2 + c^2\kappa^2 \sum_{i=1}^{M} \lambda_i \|u_i\|_2^2 \right], \quad \text{(D.12)} \\
&\sum_{i=1}^{M} \sum_{j=1}^{M} \lambda_i \lambda_j \|A(\beta_i - \beta_j)\|_2^2 \\
&\leq \kappa^2 (1+\delta_{2(t-d)k|\mathcal{I}}) \sum_{i=1}^{M} \sum_{j=1}^{M} \lambda_i \lambda_j \|u_i - u_j\|_2^2 \\
&\leq 2\kappa^2 (1+\mu\delta_{tk|\mathcal{I}}) \left(\sum_{i=1}^{M} \lambda_i \|u_i\|_2^2 - \|h_{K_2}\|_2^2 \right), \quad \text{(D.13)}
\end{aligned}$$

where μ is defined in (2.5). In addition,

$$\sum_{i=1}^{M} \lambda_i \|A\beta_i\|_2^2 \geq (1-\delta_{tk|\mathcal{I}}) \left(\|h_K + h_{K_1}\|_2^2 + \kappa^2 \sum_{i=1}^{M} \lambda_i \|u_i\|_2^2 \right). \quad \text{(D.14)}$$

Denote $\vartheta = \|h_{K_2}\|_2^2$. By the inequalities (D.5)-(D.14), we obtain

$$\begin{aligned}
0 \leq\ & \big[(1+\delta_{tk|\mathcal{I}})(1-c-\kappa)^2 - (1-\delta_{tk|\mathcal{I}})(1-c)^2\big] \|h_K + h_{K_1}\|_2^2 \\
&+ \big[\mu\delta_{tk|\mathcal{I}}(1-2c) + \delta_{tk|\mathcal{I}}(1-2c+2c^2)\big] \kappa^2 \frac{\vartheta^{2(1-p)/(2-p)}}{(t-d)k} \big(\|h_{K^\complement}\|_{2,p}^p\big)^{\frac{2}{2-p}} \\
&- \kappa^2(1-2c)(1+\mu\delta_{tk|\mathcal{I}})\vartheta \\
&+ 2(1-c)\kappa(1-\kappa)\langle A(h_K + h_{K_1}), Ah\rangle + (1-2c)\kappa^2 \|Ah\|_2^2. \quad \text{(D.15)}
\end{aligned}$$

It follows from

$$c = \frac{1}{2} - \frac{1}{4}\big(\sqrt{\mu^2 p^2 + 4(1-p)} - \mu p\big)$$

that $c \in (0, 1/2)$ for $p \in (0,1)$ and $c = 1/2$ for $p = 1$.

We derive that the following function:

$$-(1-2c)(1+\mu\delta_{tk|\mathcal{I}})\vartheta + \delta_{tk|\mathcal{I}}\big[\mu(1-2c) + (1-2c+2c^2)\big] \frac{\big(\|h_{K^\complement}\|_{2,p}^p\big)^{2/(2-p)}}{(t-d)k} \vartheta^{\frac{2(1-p)}{2-p}} \quad \text{(D.16)}$$

attains its maximum at

$$\vartheta = \left\{ \frac{2\big(\sqrt{\mu^2p^2/4 + 1 - p} + \mu p/2\big)\big[\mu(1-2c) + 1 - 2c + 2c^2\big]\delta_{tk|\mathcal{I}}}{(2-p)(t-d)k(1+\mu\delta_{tk|\mathcal{I}})} \right\}^{\frac{2-p}{p}} \|h_{K^\complement}\|_{2,p}^2.$$

Therefore, we deduce

$$\big[(1+\delta_{tk|\mathcal{I}})[(1-c)s - c]^2 - (1-\delta_{tk|\mathcal{I}})[(1-c)s + (1-c)]^2\big] \|h_K + h_{K_1}\|_2^2$$

$$
\begin{aligned}
&+ 2(1-c)s\sqrt{1+\delta_{tk|\mathcal{I}}}\|h_K + h_{K_1}\|_2\|Ah\|_2 + (1-2c)\|Ah\|_2^2 \\
&+ \frac{p}{2}\left(\frac{\sqrt{\mu^2p^2/4+1-p}+\mu p/2}{1+\mu\delta_{tk|\mathcal{I}}}\right)^{\frac{2(1-p)}{p}} \\
&\times \left\{\frac{2\delta_{tk|\mathcal{I}}\big[\mu(1-2c)+1-2c+2c^2\big]}{(2-p)(t-d)k}\right\}^{\frac{2-p}{p}} \|h_{K^\complement}\|_{2,p}^2 \geq 0, \qquad \text{(D.17)}
\end{aligned}
$$

where $s := 1/\kappa - 1$. Due to $\kappa \in [0,1]$, we have $s \in [0, +\infty)$. Thus,

$$
\begin{aligned}
&2\delta_{tk|\mathcal{I}}(1-c)^2\|h_K + h_{K_1}\|_2^2 s^2 \\
&+ 2(1-c)\left\{\sqrt{1+\delta_{tk|\mathcal{I}}}\|Ah\|_2 - [1+(2c-1)\delta_{tk|\mathcal{I}}]\|h_K + h_{K_1}\|_2\right\}\|h_K + h_{K_1}\|_2 s \\
&+ \big[(2c^2+1-2c)\delta_{tk|\mathcal{I}} - (1-2c)\big]\|h_K + h_{K_1}\|_2^2 + (1-2c)\|Ah\|_2^2 \\
&+ \frac{p}{2}\left(\frac{\sqrt{\mu^2p^2/4+1-p}+\mu p/2}{1+\mu\delta_{tk|\mathcal{I}}}\right)^{\frac{2(1-p)}{p}} \\
&\times \left\{\frac{2\delta_{tk|\mathcal{I}}\big[\mu(1-2c)+1-2c+2c^2\big]}{(2-p)(t-d)k}\right\}^{\frac{2-p}{p}} \|h_{K^\complement}\|_{2,p}^2 \geq 0. \qquad \text{(D.18)}
\end{aligned}
$$

(i) For $\delta_{tk|\mathcal{I}} = 0$, we get $\|h_K + h_{K_1}\|_2 \leq \|Ah\|_2$.

(ii) For $\delta_{tk|\mathcal{I}} \in (0,1)$, there exist two cases:

1) $\|h_K + h_{K_1}\|_2 \leq \dfrac{\sqrt{1+\delta_{tk|\mathcal{I}}}}{1+(2c-1)\delta_{tk|\mathcal{I}}}\|Ah\|_2$.

2) $\|h_K + h_{K_1}\|_2 > \dfrac{\sqrt{1+\delta_{tk|\mathcal{I}}}}{1+(2c-1)\delta_{tk|\mathcal{I}}}\|Ah\|_2$. Then it follows from (D.18) that

$$
\begin{aligned}
\|h_K + h_{K_1}\|_2 \leq{}& \frac{[1+(4c-1)\delta_{tk|\mathcal{I}}]\sqrt{1+\delta_{tk}}}{1-\delta_{tk|\mathcal{I}}^2}\|Ah\|_2 \\
&+ \Bigg(\frac{p\delta_{tk|\mathcal{I}}\big((\sqrt{\mu^2p^2/4+1-p}+\mu p/2)/(1+\mu\delta_{tk|\mathcal{I}})\big)^{2(1-p)/p}}{1-\delta_{tk|\mathcal{I}}^2} \\
&\times \left\{\frac{\delta_{tk|\mathcal{I}}\big[\mu(1-2c)+(1-2c+2c^2)\big]}{(1-p/2)(t-d)k}\right\}^{\frac{2-p}{p}}\Bigg)^{\frac{1}{2}} \|h_{K^\complement}\|_{2,p}. \qquad \text{(D.19)}
\end{aligned}
$$

Denote

$$
\eta = \max\left\{\frac{\sqrt{1+\delta_{tk|\mathcal{I}}}}{1+(2c-1)\delta_{tk|\mathcal{I}}}, \frac{\sqrt{1+\delta_{tk}}[1+(4c-1)\delta_{tk|\mathcal{I}}]}{1-\delta_{tk|\mathcal{I}}^2}\right\}.
$$

It is clear that $\eta \geq 1$ since

$$
\frac{\sqrt{1+\delta_{tk|\mathcal{I}}}}{1+(2c-1)\delta_{tk|\mathcal{I}}} \geq \frac{1}{1+(2c-1)\delta_{tk|\mathcal{I}}} \geq 1.
$$

To sum up, for $\delta_{tk|\mathcal{I}} \in [0,1)$, we have

$$
\|h_K\|_2 \leq \|h_K + h_{K_1}\|_2 \leq \eta\|Ah\|_2 + \beta\|h_{K^\complement}\|_{2,p}, \qquad \text{(D.20)}
$$

where

$$\beta=\delta_{tk|\mathcal{I}}^{\frac{1}{p}}\left(\frac{\sqrt{\mu^2p^2/4+1-p}+\mu p/2}{1+\mu\delta_{tk|\mathcal{I}}}\right)^{\frac{1}{p}-1}\sqrt{\frac{p}{1-\delta_{tk|\mathcal{I}}^2}\left[\frac{\mu(1-2c)+1-2c+2c^2}{(1-p/2)(t-d)k}\right]^{\frac{2}{p}-1}}.$$

The proof is complete. □

Appendix E

Proof of Lemma 2.9. Recall (2.10), i.e.

$$\begin{aligned}&\sum_{i=1}^m\|h_{-\max(dk)}[i]\|_2^p\\&=\|h_{-\max(dk)}\|_{2,p}^p\\&\leq\|h_{\max(dk)}\|_{2,p}^p+2\|x_{-\max(k)}\|_{2,p,w}^p+\alpha\|h\|_{2,q}^p\\&=\sum_{i=1}^m\|h_{\max(dk)}[i]\|_2^p+\big(2\|x_{-\max(k)}\|_{2,p,w}^p+\alpha\|h\|_{2,q}^p\big).\end{aligned}\tag{E.1}$$

By Lemma 2.2 and $1/p\geq 1$, we derive

$$\begin{aligned}&\|h_{-\max(dk)}\|_{2,1}\\&=\sum_{i=1}^m\|h_{-\max(dk)}[i]\|_2\leq\sum_{i=1}^m\big(\|h_{-\max(dk)}[i]\|_2^p\big)^{\frac{1}{p}}\\&\leq dk\left\{\left[\frac{1}{dk}\sum_{i=1}^m\big(\|h_{\max(dk)}[i]\|_2^p\big)^{\frac{1}{p}}\right]^p+\frac{2\|x_{-\max(k)}\|_{2,p,w}^p+\alpha\|h\|_{2,q}^p}{dk}\right\}^{\frac{1}{p}}\\&=\left\{\left[\sum_{i=1}^m\big(\|h_{\max(dk)}[i]\|_2^p\big)^{\frac{1}{p}}\right]^p+\frac{2\|x_{-\max(k)}\|_{2,p,w}^p+\alpha\|h\|_{2,q}^p}{(dk)^{1-p}}\right\}^{\frac{1}{p}}\\&=\left[\|h_{\max(dk)}\|_{2,1}^p+\frac{2}{(dk)^{1-p}}\|x_{-\max(k)}\|_{2,p,w}^p+\frac{\alpha}{(dk)^{1-p}}\|h\|_{2,q}^p\right]^{\frac{1}{p}}.\end{aligned}\tag{E.2}$$

Then by virtue of Jensen inequality, we derive

$$\begin{aligned}&\|h_{-\max(dk)}\|_{2,1}\\&\leq\left\{3^{1-p}\left[\|h_{\max(dk)}\|_{2,1}+\frac{2^{1/p}}{(dk)^{1/p-1}}\|x_{-\max(k)}\|_{2,p,w}+\frac{\alpha^{1/p}}{(dk)^{1/p-1}}\|h\|_{2,q}\right]^p\right\}^{\frac{1}{p}}\\&\leq 3^{\frac{1}{p}-1}\left[\|h_{\max(dk)}\|_{2,1}+\frac{2^{1/p}}{(dk)^{1/p-1}}\|x_{-\max(k)}\|_{2,p,w}+\frac{\alpha^{1/p}}{(dk)^{1/p-1}}\|h\|_{2,q}\right].\end{aligned}\tag{E.3}$$

By Hölder inequality, we derive

$$\begin{aligned}&\|h_{\max(dk)}\|_{2,1}\\&=\sum_{i=1}^m\|h_{\max(dk)}[i]\|_2\leq(dk)^{1-\frac{1}{q}}\left(\sum_{i=1}^m\|h_{\max(dk)}[i]\|_2^q\right)^{\frac{1}{q}}\\&=(dk)^{\frac{q-1}{q}}\|h_{\max(dk)}\|_{2,q}.\end{aligned}\tag{E.4}$$

Then we obtain

$$\begin{aligned}
\|h\|_{2,1} &= \|h_{\max(dk)}\|_{2,1} + \|h_{-\max(dk)}\|_{2,1} \\
&\leq \big(1 + 3^{\frac{1}{p}-1}\big)\|h_{\max(dk)}\|_{2,1} \\
&\quad + 3^{\frac{1}{p}-1}\left[\frac{2^{1/p}}{(dk)^{1/p-1}}\|x_{-\max(k)}\|_{2,p,w} + \frac{\alpha^{1/p}}{(dk)^{1/p-1}}\|h\|_{2,q}\right] \\
&\overset{(E.4)}{\leq} \big(1 + 3^{\frac{1}{p}-1}\big)(dk)^{\frac{q-1}{q}}\|h_{\max(dk)}\|_{2,q} \\
&\quad + 3^{\frac{1}{p}-1}\left[\frac{2^{1/p}}{(dk)^{1/p-1}}\|x_{-\max(k)}\|_{2,p,w} + \frac{\alpha^{1/p}}{(dk)^{1/p-1}}\|h\|_{2,q}\right] \\
&\leq \left[\big(1 + 3^{\frac{1}{p}-1}\big)(dk)^{\frac{q-1}{q}} + \frac{3^{1/p-1}\alpha^{1/p}}{(dk)^{1/p-1}}\right]\|h\|_{2,q} + \frac{2^{1/p}3^{1/p-1}}{(dk)^{1/p-1}}\|x_{-\max(k)}\|_{2,p,w}.
\end{aligned} \tag{E.5}$$

Thus, we have

$$\begin{aligned}
\|Ah\|_2 \leq \Bigg(&\left[\big(1 + 3^{\frac{1}{p}-1}\big)(dk)^{\frac{q-1}{q}} + \frac{3^{1/p-1}\alpha^{1/p}}{(dk)^{1/p-1}}\right]\|h\|_{2,q}\|A^\top Ah\|_{2,\infty} \\
&+ \frac{2^{1/p}3^{1/p-1}}{(dk)^{1/p-1}}\|x_{-\max(k)}\|_{2,p,w}\|A^\top Ah\|_{2,\infty}\Bigg)^{\frac{1}{2}}.
\end{aligned} \tag{E.6}$$

By (2.11), for any $\theta \in (0,1)$, we deduce

$$\begin{aligned}
&[\theta + (1-\theta)]\|h\|_{2,q} \\
&\leq \phi_1\Bigg(\left[\big(1 + 3^{\frac{1}{p}-1}\big)(dk)^{\frac{q-1}{q}} + \frac{3^{1/p-1}\alpha^{1/p}}{(dk)^{1/p-1}}\right]\|h\|_{2,q}\|A^\top Ah\|_{2,\infty} \\
&\qquad + \frac{2^{1/p}3^{1/p-1}}{(dk)^{1/p-1}}\|x_{-\max(k)}\|_{2,p,w}\|A^\top Ah\|_{2,\infty}\Bigg)^{\frac{1}{2}} \\
&\quad + \phi_2\|x_{-\max(k)}\|_{2,p,w}.
\end{aligned} \tag{E.7}$$

Therefore,

$$\begin{aligned}
\theta\|h\|_{2,q} \leq \phi_1\Bigg(&\left[\big(1 + 3^{\frac{1}{p}-1}\big)(dk)^{\frac{q-1}{q}} + \frac{3^{1/p-1}\alpha^{1/p}}{(dk)^{1/p-1}}\right]\|h\|_{2,q}\|A^\top Ah\|_{2,\infty} \\
&+ \frac{2^{1/p}3^{1/p-1}}{(dk)^{1/p-1}}\|x_{-\max(k)}\|_{2,p,w}\|A^\top Ah\|_{2,\infty}\Bigg)^{\frac{1}{2}}
\end{aligned} \tag{E.8}$$

or

$$\|h\|_{2,q} \leq \frac{\phi_2}{1-\theta}\|x_{-\max(k)}\|_{2,p,w}. \tag{E.9}$$

Otherwise, that is, neither (E.8) nor (E.9) holds, then we have

$$\begin{aligned}
&[\theta + (1-\theta)]\|h\|_{2,q} \\
&> \phi_1\Bigg(\left[\big(1 + 3^{\frac{1}{p}-1}\big)(dk)^{\frac{q-1}{q}} + \frac{3^{1/p-1}\alpha^{1/p}}{(dk)^{1/p-1}}\right]\|h\|_{2,q}\|A^\top Ah\|_{2,\infty} \\
&\qquad + \frac{2^{1/p}3^{1/p-1}}{(dk)^{1/p-1}}\|x_{-\max(k)}\|_{2,p,w}\|A^\top Ah\|_{2,\infty}\Bigg)^{\frac{1}{2}} \\
&\quad + \phi_2\|x_{-\max(k)}\|_{2,p,w},
\end{aligned} \tag{E.10}$$

and (E.7) contradicts.

The inequality (E.8) can be readily converted to

$$\theta^2\|h\|_{2,q}^2-\phi_1^2\left[\left(1+3^{\frac{1}{p}-1}\right)(dk)^{1-\frac{1}{q}}+\frac{3^{1/p-1}\alpha^{1/p}}{(dk)^{1/p-1}}\right]\|A^{\mathsf{T}}Ah\|_{2,\infty}\|h\|_{2,q} - \frac{2^{1/p}3^{1/p-1}}{(dk)^{1/p-1}}\phi_1^2\|x_{-\max(k)}\|_{2,p,w}\|A^{\mathsf{T}}Ah\|_{2,\infty}\le 0, \tag{E.11}$$

where the left-hand side is a quadratic polynomial in $\|h\|_{2,q}$, and thus,

$$\begin{aligned}\|h\|_{2,q}&\le\frac{\phi_1^2}{2\theta^2}\left[\left(1+3^{\frac{1}{p}-1}\right)(dk)^{1-\frac{1}{q}}+\frac{3^{1/p-1}\alpha^{1/p}}{(dk)^{1/p-1}}\right]\|A^{\mathsf{T}}Ah\|_{2,\infty}\\&\quad+\frac{1}{2\theta^2}\left(\phi_1^4\left[\left(1+3^{\frac{1}{p}-1}\right)(dk)^{1-\frac{1}{q}}+\frac{3^{1/p-1}\alpha^{1/p}}{(dk)^{1/p-1}}\right]^2\|A^{\mathsf{T}}Ah\|_{2,\infty}^2\right.\\&\qquad\left.+4\theta^2\frac{2^{1/p}3^{1/p-1}\phi_1^2}{(dk)^{1/p-1}}\|x_{-\max(k)}\|_{2,p,w}\|A^{\mathsf{T}}Ah\|_{2,\infty}\right)^{\frac{1}{2}}\\&\le\frac{\phi_1^2}{2\theta^2}\left[\left(1+3^{\frac{1}{p}-1}\right)(dk)^{1-\frac{1}{q}}+\frac{3^{1/p-1}\alpha^{1/p}}{(dk)^{1/p-1}}\right]\|A^{\mathsf{T}}Ah\|_{2,\infty}\\&\quad+\frac{\phi_1^2}{2\theta^2}\left[\left(1+3^{\frac{1}{p}-1}\right)(dk)^{1-\frac{1}{q}}+\frac{3^{1/p-1}\alpha^{1/p}}{(dk)^{1/p-1}}\right]\|A^{\mathsf{T}}Ah\|_{2,\infty}\\&\quad+\frac{2^{1/p}3^{1/p-1}\phi_1^2\|x_{-\max(k)}\|_{2,p,w}}{\phi_1^2\left[(1+3^{1/p-1})(dk)^{1-1/q}+3^{1/p-1}\alpha^{1/p}/(dk)^{1/p-1}\right](dk)^{1/p-1}}\\&=\frac{\phi_1^2}{\theta^2}\left[\left(1+3^{\frac{1}{p}-1}\right)(dk)^{1-\frac{1}{q}}+\left(\frac{3}{dk}\right)^{\frac{1}{p}-1}\alpha^{\frac{1}{p}}\right]\|A^{\mathsf{T}}Ah\|_{2,\infty}\\&\quad+\frac{2^{1/p}}{(3^{1-1/p}+1)(dk)^{(q-p)/(qp)}+\alpha^{1/p}}\|x_{-\max(k)}\|_{2,p,w}.\end{aligned}\tag{E.12}$$

In all,

$$\|h\|_{2,q}\le\phi_3\|A^{\mathsf{T}}Ah\|_{2,\infty}+\phi_4\|x_{-\max(k)}\|_{2,p,w},$$

where ϕ_3 and ϕ_4 are defined in (2.13). □

Appendix F

Proof of Theorem 3.1. We have $d\ge 1$ and

$$\left|\left(T_0\cup\bigcup_{i=j}^{L}\widetilde{T}_i\right)\setminus\bigcup_{i=j}^{L}\left(\widetilde{T}_i\cap T_0\right)\right|=\left(1+\sum_{i=j}^{L}\rho_i-2\sum_{i=j}^{L}\nu_i\rho_i\right)k\le dk.$$

Denote $h=\hat{x}^{\ell_2}-x$. By Lemma 2.6, we obtain

$$\begin{aligned}\|h_{-\max(dk)}\|_{2,p}^p&\le\left\|h_{T_0^{\complement}}\right\|_{2,p}^p\le w_L^p\|h_{T_0}\|_{2,p}^p+\sum_{j=2}^{L}\left(w_{j-1}^p-w_j^p\right)\left\|h_{(T_0\cup\bigcup_{i=j}^{L}\widetilde{T}_i)\setminus\bigcup_{i=j}^{L}(\widetilde{T}_i\cap T_0)}\right\|_{2,p}^p\\&\quad+\left(1-w_1^p\right)\left\|h_{(T_0\cup\widetilde{T})\setminus\bigcup_{i=1}^{L}(\widetilde{T}_i\cap T_0)}\right\|_{2,p}^p+2\left\|x_{T_0^{\complement}}\right\|_{2,p,\mathrm{w}}^p+\alpha\|h\|_{2,q}^p\\&\le w_L^p\|h_{\max(dk)}\|_{2,p}^p+\sum_{j=2}^{L}\left(w_{j-1}^p-w_j^p\right)\|h_{\max(dk)}\|_{2,p}^p\end{aligned}$$

$$
\begin{aligned}
&+\big(1-w_1^p\big)\|h_{\max(dk)}\|_{2,p}^p+2\big\|x_{T_0^{\complement}}\big\|_{2,p,\mathrm{w}}^p+\alpha\|h\|_{2,q}^p\\
&=\|h_{\max(dk)}\|_{2,p}^p+2\big\|x_{T_0^{\complement}}\big\|_{2,p,\mathrm{w}}^p+\alpha\|h\|_{2,q}^p, \qquad \text{(F.1)}
\end{aligned}
$$

$$
\begin{aligned}
\|h_{-\max(dk)}\|_{2,p}^p &\le w_L^p k^{\frac{2-p}{2}}\|h_{\max(dk)}\|_2^p+\alpha\|h\|_{2,q}^p\\
&+\sum_{j=2}^{L}\big(w_{j-1}^p-w_j^p\big)\left(k+\sum_{i=j}^{L}\rho_i k-2\sum_{i=j}^{L}\nu_i\rho_i k\right)^{\frac{2-p}{2}}\|h_{\max(dk)}\|_2^p\\
&+\big(1-w_1^p\big)\left(k+\sum_{i=1}^{L}\rho_i k-2\sum_{i=1}^{L}\nu_i\rho_i k\right)^{\frac{2-p}{2}}\|h_{\max(dk)}\|_2^p+2\big\|x_{T_0^{\complement}}\big\|_{2,p,w}^p\\
&=(\sigma k)^{\frac{2-p}{2}}\|h_{\max(dk)}\|_2^p+2\big\|x_{T_0^{\complement}}\big\|_{2,p,w}^p+\alpha\|h\|_{2,q}^p. \qquad \text{(F.2)}
\end{aligned}
$$

By Lemma 2.8, we have

$$
\|h_{\max(dk)}\|_2\le\eta\|Ah\|_2+\beta\|h_{-\max(dk)}\|_{2,p}, \qquad \text{(F.3)}
$$

we derive

$$
\big[1-(\sigma k)^{\frac{2-p}{2}}\beta^p\big]\|h_{\max(dk)}\|_2^p\le\eta^p\|Ah\|_2^p+\beta^p\big(2\big\|x_{T_0^{\complement}}\big\|_{2,p,w}^p+\alpha\|h\|_{2,q}^p\big). \qquad \text{(F.4)}
$$

By (2.9), we have $\beta|_{\delta_{tk|\mathcal{I}}=0}=0$, and

$$
\begin{aligned}
\beta|_{\delta_{tk|\mathcal{I}}>0}&=\delta_{tk|\mathcal{I}}^{\frac{1}{p}}\sqrt{\frac{p}{1-\delta_{tk|\mathcal{I}}^2}}\left(\frac{\sqrt{\mu^2p^2/4+1-p}+\mu p/2}{1+\mu\delta_{tk|\mathcal{I}}}\right)^{\frac{1}{p}-1}\left[\frac{\mu(1-2c)+1-2c+2c^2}{(1-p/2)(t-d)k}\right]^{\frac{1}{p}-\frac{1}{2}}\\
&=\sqrt{\frac{p}{(1/\delta_{tk|\mathcal{I}})^2-1}}\left(\frac{\sqrt{\mu^2p^2/4+1-p}+\mu p/2}{1/\delta_{tk|\mathcal{I}}+\mu}\right)^{\frac{1}{p}-1}\left[\frac{\mu(1-2c)+1-2c+2c^2}{(1-p/2)(t-d)k}\right]^{\frac{1}{p}-\frac{1}{2}}. \qquad \text{(F.5)}
\end{aligned}
$$

Clearly, β is monotonically increasing with respect to $\delta_{tk|\mathcal{I}}$. Hence, $1-(\sigma k)^{(2-p)/2}\beta^p$ and

$$
1-\alpha\left[\frac{2^{p/q}(dk)^{(2-q)p/(2q)}\beta^p}{1-(\sigma k)^{(2-p)/2}\beta^p}+\frac{1}{(dk)^{(q-p)/q}}\right],
$$

which are both monotonically decreasing with respect to β, are also both monotonically decreasing with respect to $\delta_{tk|\mathcal{I}}$.

Consider $\hat{\delta}(p,q,\alpha,t,\sigma,d)>0$ obeying

$$
\begin{aligned}
&\left(\frac{1}{\hat{\delta}^2(p,q,\alpha,t,\sigma,d)}-1\right)\left(\frac{1}{\hat{\delta}(p,q,\alpha,t,\sigma,d)}+\mu\right)^{\frac{2(1-p)}{p}}\\
&=p\left(\sqrt{\frac{\mu^2p^2}{4}+1-p}+\frac{\mu p}{2}\right)^{\frac{2(1-p)}{p}}\left[\frac{\mu(1-2c)+1-2c+2c^2}{(1-p/2)(t-d)}\right]^{\frac{2-p}{p}}\sigma^{\frac{2-p}{p}}. \qquad \text{(F.6)}
\end{aligned}
$$

We derive $1-(\sigma k)^{(2-p)/2}\beta^p=0$ when $\delta_{tk|\mathcal{I}}=\hat{\delta}(p,q,\alpha,t,\sigma,d)$, and thus $1-(\sigma k)^{(2-p)/2}\beta^p>0$ for $\delta_{tk|\mathcal{I}}<\hat{\delta}(p,q,\alpha,t,\sigma,d)$.

In view of (3.5), i.e.

$$
\begin{aligned}
&\left(\frac{1}{\delta^2(p,q,\alpha,t,\sigma,d)}-1\right)\left(\frac{1}{\delta(p,q,\alpha,t,\sigma,d)}+\mu\right)^{\frac{2(1-p)}{p}}\\
&=p\left(\sqrt{\frac{\mu^2p^2}{4}+1-p}+\frac{\mu p}{2}\right)^{\frac{2(1-p)}{p}}\left[\frac{\mu\,(1-2c)+1-2c+2c^2}{(1-p/2)(t-d)}\right]^{\frac{2-p}{p}}\left[\sigma^{\frac{2-p}{2}}+\frac{2^{p/q}d^{(2-p)/2}\alpha}{(dk)^{(q-p)/q}-\alpha}\right]^{\frac{2}{p}},
\end{aligned}
$$

we have

$$1-\alpha\left[\frac{2^{p/q}(dk)^{(2-q)p/(2q)}\beta^p}{1-(\sigma k)^{(2-p)/2}\beta^p}+\frac{1}{(dk)^{(q-p)/q}}\right]=0,$$

when $\delta_{tk|\mathcal{I}}=\delta(p,q,\alpha,t,\sigma,d)$. Hence, it follows from (3.4) that

$$1-\alpha\left[\frac{2^{p/q}(dk)^{(2-q)p/(2q)}\beta^p}{1-(\sigma k)^{(2-p)/2}\beta^p}+\frac{1}{(dk)^{(q-p)/q}}\right]>0.$$

The function $(1/\delta^2-1)(1/\delta+\mu)^{2(1-p)/p}$ is monotonically decreasing with respect to $\delta\in(0,1]$, and

$$\sigma^{\frac{2-p}{p}}\leq\left[\sigma^{\frac{2-p}{2}}+\frac{2^{p/q}d^{(2-p)/2}\alpha}{(dk)^{(q-p)/q}-\alpha}\right]^{\frac{2}{p}},$$

hence, $\delta(p,q,\alpha,t,\sigma,d)\leq\hat{\delta}(p,q,\alpha,t,\sigma,d)$. As a consequence, for $\delta_{tk|\mathcal{I}}<\delta(p,q,\alpha,t,\sigma,d)$, we have $\delta_{tk|\mathcal{I}}<\hat{\delta}(p,q,\alpha,t,\sigma,d)$, and thus $1-(\sigma k)^{(2-p)/2}\beta^p>0$.

Thus, by (F.4), we obtain

$$\begin{aligned}\|h_{\max(dk)}\|_2^p\leq{}&\frac{\eta^p}{1-(\sigma k)^{(2-p)/2}\beta^p}\|Ah\|_2^p\\&+\frac{\beta^p}{1-(\sigma k)^{(2-p)/2}\beta^p}\big(2\|x_{T_0^\complement}\|_{2,p,w}^p+\alpha\|h\|_{2,q}^p\big).\end{aligned}\tag{F.7}$$

Then by Lemma 2.2, (F.1) and $q/p\geq 1$, we deduce

$$\begin{aligned}\|h_{-\max(dk)}\|_{2,q}^p&=\left[\sum_{i=1}^m\big(\|h_{-\max(dk)}[i]\|_2^p\big)^{\frac{q}{p}}\right]^{\frac{p}{q}}\\&\leq\left\{dk\left\{\left[\frac{1}{dk}\sum_{i=1}^m\big(\|h_{\max(dk)}[i]\|_2^p\big)^{\frac{q}{p}}\right]^{\frac{p}{q}}+\frac{2\|x_{-\max(k)}\|_{2,p,w}^p+\alpha\|h\|_{2,q}^p}{dk}\right\}^{\frac{q}{p}}\right\}^{\frac{p}{q}}\\&=\left(\sum_{i=1}^m\|h_{\max(dk)}[i]\|_2^q\right)^{\frac{p}{q}}+\frac{2\|x_{-\max(k)}\|_{2,p,w}^p+\alpha\|h\|_{2,q}^p}{(dk)^{1-p/q}}\\&=\|h_{\max(dk)}\|_{2,q}^p+\frac{2}{(dk)^{(q-p)/q}}\|x_{T_0^\complement}\|_{2,p,w}^p+\frac{\alpha}{(dk)^{(q-p)/q}}\|h\|_{2,q}^p.\end{aligned}\tag{F.8}$$

By (F.7), (F.8) and $q/p\geq 1$, we deduce

$$\begin{aligned}\|h\|_{2,q}^p&=\big(\|h_{\max(dk)}\|_{2,q}^q+\|h_{-\max(dk)}\|_{2,q}^q\big)^{\frac{p}{q}}\\&\leq\left\{\|h_{\max(dk)}\|_{2,q}^q+\left[\|h_{\max(dk)}\|_{2,q}^p+\frac{1}{(dk)^{(q-p)/q}}\big(2\|x_{T_0^\complement}\|_{2,p,w}^p+\alpha\|h\|_{2,q}^p\big)\right]^{\frac{q}{p}}\right\}^{\frac{p}{q}}\\&\overset{\varpi}{\leq}2^{\frac{p}{q}}\|h_{\max(dk)}\|_{2,q}^p+\frac{1}{(dk)^{(q-p)/q}}\big(2\|x_{T_0^\complement}\|_{2,p,w}^p+\alpha\|h\|_{2,q}^p\big)\\&\leq\frac{2^{p/q}(dk)^{(2-q)p/(2q)}\eta^p}{1-(\sigma k)^{(2-p)/2}\beta^p}\|Ah\|_2^p\\&\quad+\left[\frac{2^{p/q}(dk)^{(2-q)p/(2q)}\beta^p}{1-(\sigma k)^{(2-p)/2}\beta^p}+\frac{1}{(dk)^{(q-p)/q}}\right]\big(2\|x_{T_0^\complement}\|_{2,p,w}^p+\alpha\|h\|_{2,q}^p\big),\end{aligned}\tag{F.9}$$

where the details of Step ϖ are as follows. Denote

$$\digamma_1 := \left(\|h_{\max(dk)}\|_{2,q}^p, \|h_{\max(dk)}\|_{2,q}^p\right) \in \mathbb{R}^2,$$
$$\digamma_2 := \left(0, \frac{2\|x_{T_0^\complement}\|_{2,p,w}^p + \alpha\|h\|_{2,q}^p}{(dk)^{(q-p)/q}}\right) \in \mathbb{R}^2.$$

Since $q/p \geq 1$, we have

$$\begin{aligned}
&\left\{\|h_{\max(dk)}\|_{2,q}^q + \left[\|h_{\max(dk)}\|_{2,q}^p + \frac{1}{(dk)^{(q-p)/q}}\left(2\big\|x_{T_0^\complement}\big\|_{2,p,w}^p + \alpha\|h\|_{2,q}^p\right)\right]^{\frac{q}{p}}\right\}^{\frac{p}{q}} \\
&= \|\digamma_1 + \digamma_2\|_{\frac{q}{p}} \leq \|\digamma_1\|_{\frac{q}{p}} + \|\digamma_2\|_{\frac{q}{p}} \\
&= 2^{\frac{p}{q}}\|h_{\max(dk)}\|_{2,q}^p + \frac{1}{(dk)^{(q-p)/q}}\left(2\big\|x_{T_0^\complement}\big\|_{2,p,w}^p + \alpha\|h\|_{2,q}^p\right). \qquad \text{(F.10)}
\end{aligned}$$

Thus,

$$\begin{aligned}
&\left\{1 - \alpha\left[\frac{2^{p/q}(dk)^{(2-q)p/(2q)}\beta^p}{1-(\sigma k)^{(2-p)/2}\beta^p} + \frac{1}{(dk)^{(q-p)/q}}\right]\right\}\|h\|_{2,q}^p \\
&\leq \frac{2^{p/q}(dk)^{(2-q)p/(2q)}\eta^p}{1-(\sigma k)^{(2-p)/2}\beta^p}\|Ah\|_2^p + 2\left[\frac{2^{p/q}(dk)^{(2-q)p/(2q)}\beta^p}{1-(\sigma k)^{(2-p)/2}\beta^p} + \frac{1}{(dk)^{(q-p)/q}}\right]\|x_{T_0^\complement}\|_{2,p,w}^p. \qquad \text{(F.11)}
\end{aligned}$$

Since

$$\|Ah\|_2 \leq \|y - A\hat{x}^{\ell_2}\|_2 + \|Ax - y\|_2 \leq \varepsilon + \epsilon,$$

we obtain

$$\|h\|_{2,q} \leq C_1(\varepsilon + \epsilon) + C_2\big\|x_{T_0^\complement}\big\|_{2,p,w}, \qquad \text{(F.12)}$$

where C_1 and C_2 are defined in (3.7). □

Appendix G

Proof of Theorem 3.2. Denote $h = \hat{x}^{BDS} - x$. We derive

$$\|A^\mathsf{T}Ah\|_{2,\infty} \leq \|A^\mathsf{T}(A\hat{x}^{BDS} - y)\|_{2,\infty} + \|A^\mathsf{T}(y - Ax)\|_{2,\infty} \leq \varepsilon + \epsilon. \qquad \text{(G.1)}$$

In the proof of Theorem 3.1, the specific form of noise structure is not needed for the derivation of (F.1) and (F.11), hence the difference of noise structure has no effect on them. In view of $\delta_{tk|\mathcal{I}} < \delta(p, q, \alpha, t, \sigma, d)$, applying (F.1), (F.11) and Lemma 2.9 leads to

$$\|h\|_{2,q} \leq C_3(\varepsilon + \epsilon) + C_4\big\|x_{T_0^\complement}\big\|_{2,p,w},$$

where C_3 and C_4 are defined in (3.15). □

Appendix H

Proof of Theorem 3.3. Denote $h = \hat{x}^{DS} - x$. The proof of Theorem 3.2 can be subtly carried over to this theorem by replacing (G.1) with

$$\|A^\mathsf{T}Ah\|_{2,\infty} \leq \sqrt{\max_i d_i}\,(\varepsilon + \epsilon), \qquad \text{(H.1)}$$

and therefore

$$\|h\|_{2,q} \le C_3\sqrt{\max_i d_i}\,(\varepsilon+\epsilon) + C_4\|x_{T_0^\complement}\|_{2,p,w}.$$

The proof is complete. □

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China (Grant No. 12171496), and in part by the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2024A1515012057).

References

[1] T. Blumensath and M.E. Davies, Iterative hard thresholding for compressed sensing, *Appl. Comput. Harmon. Anal.*, **27**:3 (2009), 265–274.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, Distributed optimization and statistics via alternating direction method of multipliers, *Found. Trends Mach. Learn.*, **3**:1 (2010), 1–122.

[3] T. Cai and A. Zhang, Sharp RIC bound for sparse signal and low-rank matrix recovery, *Appl. Comput. Harmon. Anal.*, **35**:1 (2013), 74–93.

[4] T. Cai and A. Zhang, Sparse representation of a polytope and recovery of sparse signals and low-rank matrices, *IEEE Trans. Inf. Theory*, **60**:1 (2014), 122–132.

[5] E. Candès, J. Romberg, and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Comm. Pure Appl. Math.*, **59**:8 (2006), 1207–1223.

[6] E. Candès and T. Tao, Decoding by linear programming, *IEEE Trans. Inf. Theory*, **51**:12 (2005), 4203–4215.

[7] E. Candès and T. Tao, The Dantzig selector: Statistical estimation when p is much larger than n, *Ann. Statist.*, **35**:6 (2007), 2313–2351.

[8] R. Chartrand and W. Yin, Iteratively reweighted algorithms for compressive sensing, in: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, (2008), 3869–3872.

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

[10] D.L. Donoho, M. Elad, and V.N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, *IEEE Trans. Inf. Theory*, **52**:1 (2006), 6–18.

[11] B. Du and A. Wan, Stable and robust recovery of approximately k-sparse signals with partial support information in noise settings via weighted $\ell_p\,(0 < p \le 1)$ minimization, *J. Comput. Math.*, **41**:6 (2023), 1137–1170.

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

[13] T. Euler and T. Baden, Computational neuroscience: Species-specific motion detectors, *Nature*, **535** (2016), 45–46.

[14] S. Foucart and M. 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.

[15] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser, 2013.

[16] H. Ge and W. Chen, Recovery of signals by a weighted ℓ_2/ℓ_1 minimization under arbitrary prior support information, *Signal Process.*, **148** (2018), 288–302.

[17] H. Ge, W. Chen, M. Ng, New RIP bounds for recovery of sparse signals with partial support information via weighted ℓ_p minimization, *IEEE Transactions on Information Theory*, **66**:6 (2020), 3914–3928.

[18] H. Ge and P. Li, The Dantzig selector: Recovery of signal via $\ell_1 - \alpha\ell_2$ minimization, *Inverse Problems*, **38**:1 (2022), 015006.

[19] F. Gu and A. Wan, Nonconvex $\ell_p - \alpha\ell_q$ minimization method and p-RIP condition for stable recovery of approximately k-sparse signals, *Comput. Appl. Math.*, **43** (2024), 29.

[20] L. Huo, W. Chen, H. Ge, and M. Ng, $L_1 - \beta L_q$ minimization for signal and image recovery, *SIAM J. Imaging Sci.*, **16**:4 (2023), 1886–1928.

[21] L. Jacques, A short note on compressed sensing with partially known signal support, *Signal Process.*, **90**:12 (2010), 3308–3312.

[22] L. Kang and C. Lu, Distributed compressive video sensing, in: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, (2009), 1169–1172.

[23] Y. Li and W. Chen, The high order block RIP condition for signal recovery, *J. Comput. Math.*, **37**:1 (2019), 61–75.

[24] J. Lin and S. Li, Block sparse recovery via mixed ℓ_2/ℓ_1 minimization, *Acta Math. Sin. (Engl. Ser.)*, **29**:7 (2013), 1401–1412.

[25] Y. Lou, Y. Qi, and P. Xin, Minimization of $\ell_1 - \ell_2$ for compressed sensing, *SIAM J. Sci. Comput.*, **37**:1 (2015), 536–563.

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

[27] D. Needell and J.A. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, *Appl. Comput. Harmon. Anal.*, **26**:3 (2009), 301–321.

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

[29] Q. Sun, Recovery of sparsest signals via ℓ_q-minimization, *Appl. Comput. Harmon. Anal.*, **32**:3 (2012), 329–341.

[30] N. Vaswani and W. Lu, Modified-CS: Modifying compressive sensing for problems with partially known support, *IEEE Trans. Signal Process.*, **58**:9 (2010), 4595–4607.

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

[32] A. Wan, Stable recovery of approximately k-sparse signals in noisy cases via ℓ_p minimization, *Neurocomputing*, **406** (2020), 159–168.

[33] A. Wan, Uniform RIP conditions for recovery of sparse signals by $\ell_p\,(0 < p \leq 1)$ minimization, *IEEE Trans. Signal Process.*, **68** (2020), 5379–5394.

[34] W. Wang, J. Wang, and Z. Zhang, Block-sparse signal recovery via ℓ_2/ℓ_{1-2} minimisation method, *IET Signal Process.*, **12**:4 (2018), 422–430.

[35] Y. Wang, W. Yin, and J. Zeng, Global convergence of ADMM in nonconvex nonsmooth optimization, *J. Sci. Comput.*, **78**:1 (2019), 29–63.

[36] L. Weizman, Y. Eldar, and D. Bashat, Compressed sensing for longitudinal MRI: An adaptive-weighted approach, *Med. Phys.*, **42**:9 (2015), 5195–5208.

[37] J. Zhang and S. Zhang, Recovery analysis for ℓ_2/ℓ_{1-2} minimization via prior support information, *Digital Signal Process.*, **121** (2022), 103315.

[38] J. Zhang, S. Zhang, and W. Wang, Robust signal recovery for ℓ_{1-2} minimization via prior support information, *Inverse Problems*, **37** (2021), 115001.

[39] J. Zhou and W. Huang, Block sparse representation of a polytope and non-convex compressed sensing, *Sci. Sin. Math.*, **52** (2022), 105–120. (in Chinese)

[40] Z. Zhou, RIP analysis for the weighted $\ell_r - \ell_1$ minimization method, *Signal Process.*, **202** (2023), 108754.

[41] Z. Zhou and J. Yu, A new nonconvex sparse recovery method for compressive sensing, *Front. Appl. Math. Stat.*, **5** (2019), 14.