Appl. Comput. Harmon. Anal. 49 (2020) 381–414

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

The $\ell_{2,q}$ regularized group sparse optimization: Lower bound theory, recovery bound and algorithms

Xue Feng, Shi Yan, Chunlin Wu *

School of Mathematical Sciences, Nankai University, Tianjin, 300071, China

ARTICLE INFO

Article history:
Received 15 August 2018
Received in revised form 30 December 2019
Accepted 19 April 2020
Available online 30 April 2020
Communicated by Radu Balan

Keywords:
Group sparsity
Global convergence
Recovery bound
Lower bound
IRLS
Signal recovery

ABSTRACT

In this paper, we consider an unconstrained $\ell_{2,q}$ minimization for group sparse signal recovery. For this nonconvex and non-Lipschitz problem, we mainly focus on its local minimizers. Firstly, a uniform lower bound for nonzero groups of the local minimizers is presented. Secondly, under group restricted isometry property (GRIP) assumption, we provide a global recovery bound for points in a sublevel set of the objective function, as well as a local recovery bound for local minimizers. Thirdly, a sufficient condition for a stationary point to be a local minimizer is shown. Fourthly, inspired by the lower bound theory which indicates the sparsity of solutions, we propose a new efficient iteratively reweighted least square (IRLS) with thresholding algorithm, with nonexpansiveness of the group support set. Compared with the classical IRLS with smoothing algorithm, our algorithm performs better in both theoretical global convergence guarantee and numerical computation.
© 2020 Elsevier Inc. All rights reserved.

1. Introduction

Different from element sparse data in conventional sense, a wide class of data, like segments and features, usually have natural grouping structures. Recently, group sparsity reconstruction has received a great deal of attention. The target is to restore $x^o \in \mathbf{R}^N$, which has few nonzero groups rather than elements, from a noisy observation $d \in \mathbf{R}^M$:

$$d = Ax^o + \xi, \tag{1}$$

where $A \in \mathbf{R}^{M\times N}(M < N)$ is the measurement matrix, and $\xi \in \mathbf{R}^M$ is the measurement error like the most common Gaussian noise.

Using convex ℓ_1 minimization to recover sparse vectors has lasted for a long history [1–3]. Therefore, one natural idea is to consider the group version of ℓ_1 minimization to obtain group sparse vectors, i.e., Group

* Corresponding author.
E-mail address: wucl@nankai.edu.cn (C. Wu).

https://doi.org/10.1016/j.acha.2020.04.002
1063-5203/© 2020 Elsevier Inc. All rights reserved.

Lasso ($\ell_{2,1}$) minimization. Group Lasso was firstly introduced for grouped variable selection in statistics in 2006 [4], and then was applied to other areas like DNA microarrays [5,6], dynamic MRI [7], source localization [8], color imaging [9] and so on. However, these ℓ_1 based minimization methods suffer from some frustrations in practical applications, and the solutions obtained are much less sparse than expected [10–14].

Instead, as shown in [15,16], the group sparsity of x can be better measured by nonconvex $\ell_{p,q}$ "norm", defined as

$$\|x\|_{p,q} := \left(\sum_{i=1}^{n} \|x_{\mathcal{G}_i}\|_p^q\right)^{\frac{1}{q}}, \qquad p \geq 1, 0 < q < 1,$$

where $x_{\mathcal{G}_i}$ is the ith group of x (see Section 2 for more details about group structure). Clearly, when all group sizes equal 1, $\|x\|_{p,q}$ degenerates to be $\|x\|_q^q = \sum_{j=1}^N |x_j|^q$. Meanwhile, since $\|x\|_{2,q}$ regularization performs better than a general $\|x\|_{p,q}$ in most cases [15,16], we focus on the following $\ell_{2,q}$ regularized minimization in this paper:

$$\min_{x \in \mathbf{R}^N} F(x) := \|Ax - d\|^2 + \alpha \|x\|_{2,q}^q, \quad 0 < q < 1, \tag{2}$$

where $\alpha > 0$ is the regularization parameter.

Generally speaking, the minimizers of $F(x)$ are not x^o in (1), thus it is important to estimate the distance between them. There is quite a lot of literature on the recovery of conventional (non-group) ℓ_q minimization. Their results were established under some mild conditions on A, such as the restricted isometry property (RIP, [17–22,14,23]) and restricted eigenvalue condition (REC, [24]). For group sparse minimization, see [25–27,15]. By generalizing RIP to group sparse case, namely GRIP, [25,26] showed that the convex Group Lasso is guaranteed to exactly recover group sparse signals. [27] also established the robust recovery for constrained $\ell_{2,q}$ model under GRIP conditions. For the recovery property of unconstrained $\ell_{2,q}$ minimization, [15] is the only work, as far as we know. Their local and global recovery bounds were established based on group restricted eigenvalue condition (GREC). However, the local recovery result in [15] is given under some assumptions like the activeness and group support conditions. To the best of our knowledge, a general form of local recovery bound for unconstrained $\ell_{2,q}$ problem is still undiscovered.

It is well known that the nonsmooth and non-Lipschitz $\ell_{2,q}$ minimization is a great challenge in algorithm design and convergence analysis. Some existing solvers include majorization minimization approach in [16], proximal gradient method with explicit subsolvers for $q = 1/2, 2/3$ in [15], a support-shrinking iterative reweighted ℓ_1 (IRL1) algorithm with two loops of iterations in [28] extended from the references in it. Especially, motivated by the great success of the iteratively reweighted least square (IRLS) method for conventional ℓ_q minimization problem (which smoothes the objective function at first and then solves a series of linear systems) [11,29–31], [27] generalized this IRLS with smoothing algorithm to the unconstrained $\ell_{2,q}$ minimization. Although this IRLS algorithm is simple to implement and performs well in numerical experiments, the sequence generated is shown to have only convergent subsequences.

In this work, starting from a lower bound theory of local minimizers of $F(x)$, we show both local and global recovery bounds of $F(x)$ and propose an effective and globally convergent IRLS with thresholding algorithm. We also provide a sufficient condition for a stationary point of $F(x)$ to be a local minimizer. Our main contributions are summarized below.

- We show a uniform lower bound for nonzero groups of local minimizers of $F(x)$; see Theorem 3.1. Compared with the smoothed $\ell_{2,q}$ regularized model, the solutions of $F(x)$ are sharper; see Proposition 6.1.
- Under the group restricted isometry property (GRIP) [26] assumption, we provide a recovery bound for each point x satisfying $F(x) \leq F(x^o)$. Meanwhile, if x is also a local minimizer with the columns of A

corresponding to its support linearly independent, a local recovery bound is derived by the means of the lower bound theory in Theorem 3.1; see Theorem 4.3. The assumptions we need are weaker than those for $\ell_{2,1}$ model, which implies the better group sparse recovery ability of the nonconvex $\ell_{2,q}$ model.
- A sufficient condition for a stationary point of $F(x)$ to be a local minimizer is proven; see Theorem 5.3. Since most algorithms for this problem can only converge to stationary points, this helps to identify whether an algorithm solution is a local minimizer.
- Motivated by our lower bound theory, we propose a new IRLS with thresholding algorithm for $F(x)$; see Algorithm 2. In each iteration, the groups whose norms are less than a threshold are truncated, and then fixed to be 0 in the following steps. Compared with the classical IRLS with smoothing algorithm, our algorithm is superior in both global convergence and computational efficiency. An error bound analysis of our algorithm is also given.

The rest of this paper is organized as follows. In Section 2, there are some basic notations and preliminaries. In Section 3, we show a uniform lower bound for nonzero groups of local minimizers. In Section 4, both local and global recovery bounds for $F(x)$ are established. In Section 5, we aim at the relationship between stationary points and local minimizers. In Section 6, the focus is on two algorithms. Numerical experiments are discussed in Section 7. Conclusions are presented in Section 8.

2. Notations and preliminaries

In this paper, we use $x = (x_{\mathcal{G}_1}^T, \ldots, x_{\mathcal{G}_n}^T)^T \in \mathbf{R}^N$, i.e.,

$$x = (\underbrace{x_1, \cdots, x_{N_1}}_{x_{\mathcal{G}_1}}, \underbrace{x_{N_1+1}, \cdots, x_{N_1+N_2}}_{x_{\mathcal{G}_2}}, \cdots, \underbrace{x_{N-N_n+1}, \cdots, x_N}_{x_{\mathcal{G}_n}})^T, \tag{3}$$

to represent a predefined group structure of x. Here, $x_{\mathcal{G}_i}$ denotes the ith group of x where $\mathcal{G}_i$ is an index subset of $\{1, 2, \ldots, n\}$, N_i is the group size of the ith group, and n is the group number. For a group $x_{\mathcal{G}_i}$, $x_{\mathcal{G}_i} = 0$ means $x_j = 0$ for all $j \in \mathcal{G}_i$. There is no overlapping between any two groups, i.e., $\mathcal{G}_i \cap \mathcal{G}_k = \emptyset$ for any $i \neq k$. Obviously, if $N_i = 1$ for all i, the group structured signal degenerates to be a conventional signal.

Denote $\mathbb{I} = \{1, 2, \ldots, n\}, \mathbb{J} = \{1, 2, \ldots, N\}$. For a vector $x \in \mathbf{R}^N$, we denote the (group) support set of x by

$$\text{supp}(x) := \{j \in \mathbb{J} : x_j \neq 0\}, \qquad \text{gsupp}(x) := \{i \in \mathbb{I} : x_{\mathcal{G}_i} \neq 0\}. \tag{4}$$

For any $S \subseteq \mathbb{J}$, let $S^c \subseteq \mathbb{J}$ be the complementary set of S. For any $\mathbb{G} \subseteq \mathbb{I}$, let $\mathbb{G}^c \subseteq \mathbb{I}$ be the complementary set of $\mathbb{G}$. We denote x_S as the subvector of x indexed by S, and denote A_S as the submatrix of A consisting of columns indexed by S. For $S \subseteq \mathbb{J}$, we construct $\overline{x_S} \in \mathbf{R}^N$ as an extension of x_S:

$$(\overline{x_S})_S = x_S, \quad (\overline{x_S})_{S^c} = 0. \tag{5}$$

Without loss of generality, we use $\|\cdot\|$ to represent ℓ_2 norm, i.e., $\|\cdot\| = \|\cdot\|_2$. Note that $\|x\|_0$ denotes the number of the nonzero entries in x. In particular, for $x \in \mathbf{R}^N$, $\|x\|_{2,2} = \|x\|$, $\|x\|_{2,0} = \#\{i \in \mathbb{I} \ : \ x_{\mathcal{G}_i} \neq 0\}$, and $\|x\|_{2,\infty} = \max_{i\in\mathbb{I}} \|x_{\mathcal{G}_i}\|$.

We denote

$$\sigma_{t,q'}(x) = \min_{\|y\|_{2,0}\leq t} \|x - y\|_{2,q'}^{q'}, \quad \text{where} \quad q' > 0. \tag{6}$$

If $\sigma_{t,q'}(x)$ is small, we say x has a very small tail. Then, if $S_0 \subset \mathbb{J}$ where S_0 is the indices over the first t largest groups of x, we have $\sigma_{t,q'}(x) = \|x - \overline{x_{S_0}}\|_{2,q'}^{q'}$. Here, the first t largest groups mean the first t groups after a rearrangement of the vector with decreasing group ℓ_2 norms.

Lemma 2.1. *([32]) Let $0 < s < s' \leq \infty$. Then, for all $x \in \mathbf{R}^N$,*

$$\|x\|_{s'} \leq \|x\|_s \leq N^{\frac{1}{s}-\frac{1}{s'}} \|x\|_{s'}. \tag{7}$$

Moreover, for $x \in \mathbf{R}^N$ with group structure in (3), we have

$$\|x\|_{2,s'} \leq \|x\|_{2,s} \leq n^{\frac{1}{s}-\frac{1}{s'}} \|x\|_{2,s'}. \tag{8}$$

We then introduce the concept of RIP.

Definition 2.2. (RIP, or, restricted isometry property, [17]) For $t = 1, 2, \ldots, N$, the restricted isometry constant $\delta_t \in (0, 1)$ of matrix A is the smallest number such that

$$(1 - \delta_t)\|x\|^2 \leq \|Ax\|^2 \leq (1 + \delta_t)\|x\|^2, \quad \forall\, x \in \mathbf{R}^N \text{satisfying } \|x\|_0 \leq t. \tag{9}$$

For simplification, we say matrix A satisfies RIP of order t with constant δ_t.

The matrix 2-norm is denoted by $\|A\|_2$, which equals the largest singular value of A. We denote

$$\varrho(A) := \min\left\{\frac{\|Ax\|^2}{\|x\|^2} : x \in \mathbf{R}^N, \det\left(A^T_{\text{supp}(x)} A_{\text{supp}(x)}\right) \neq 0\right\} > 0. \tag{10}$$

Next, we present the generalization of RIP in group sparse setting.

Definition 2.3. (GRIP, or, group restricted isometry property, [26]) For integer $t = 1, 2, \ldots, n$, the group restricted isometry constant $\delta'_t \in (0, 1)$ of matrix A over $G = \{\mathcal{G}_1, \ldots, \mathcal{G}_n\}$ is the smallest number such that

$$(1 - \delta'_t)\|x\|^2 \leq \|Ax\|^2 \leq (1 + \delta'_t)\|x\|^2, \quad \forall\, x \in \mathbf{R}^N \text{satisfying } \|x\|_{2,0} \leq t. \tag{11}$$

For simplification, we say matrix A satisfies GRIP of order t over G with constant δ'_t.

As shown in [26], the GRIP constant is typically smaller than the standard RIP constant. Furthermore, if A is a random matrix, it satisfies GRIP almost surely.

Lemma 2.4. *Given $A \in \mathbf{R}^{M\times N}$, we have*

$$|\langle Ax, Ay\rangle| \leq \delta'_{2t}(A)\|x\|\|y\|, \quad \forall\, x, y \in \mathbf{R}^N \textit{ satisfying } \|x\|_{2,0} \leq t, \|y\|_{2,0} \leq t. \tag{12}$$

Proof. It follows from

$$\|A(x + y)\|^2 \leq (1 + \delta'_{2t}(A))\|x + y\|^2, \quad -\|A(x - y)\|^2 \leq -(1 - \delta'_{2t}(A))\|x - y\|^2. \quad \square$$

A point whose subdifferential contains 0 is called a **stationary point**. Moreover, the well known Fermat's rule remains barely unchanged, i.e., if x is a local minimizer of a function ϕ, then $0 \in \partial\phi(x)$. The subdifferential of $F(x)$ at x is given by $\forall\, j \in \mathbb{J}$ with $j \in \mathcal{G}_i$,

$$(\partial F(x))_j = \begin{cases} \{2A_j^T(Ax-d) + \alpha q\|x_{\mathcal{G}_i}\|^{q-2}x_j\}, \text{if } x_{\mathcal{G}_i} \neq 0; \\ (-\infty, +\infty), \qquad\qquad\qquad\qquad \text{otherwise.} \end{cases}$$

See Appendix 9.1 or Section 3.1 of [28] for more details. Thus, if x is a local minimizer of $F(x)$, we have

$$2A_S^T(Ax-d) + \alpha \left[\, q\|x_{\mathcal{G}_i}\|^{q-2}x_{\mathcal{G}_i} \,\right]_{i\in\mathbb{G}} = 0, \tag{13}$$

where $\mathbb{G} = \text{gsupp}(x), S = \cup_{i\in\mathbb{G}}\mathcal{G}_i$.

3. Lower bound theory

In this section, we aim at the lower bound theory about the local minimizers of $F(x)$, i.e., there is a uniform lower bound for nonzero groups of any local minimizer of $F(x)$.

Our lower bound theory applies to a general $\ell_{p,q}(p \geq 1)$ regularized minimization:

$$\min_{x\in\mathbf{R}^N} F^p(x) := \|Ax-d\|^2 + \alpha\|x\|_{p,q}^q, \qquad \text{where} \quad \alpha > 0, \quad 0 < q < 1. \tag{14}$$

Note that $F(x)$ defined in our paper equals $F^2(x)$ here.

Theorem 3.1. *(Lower bound theory) For any $d \in \mathbf{R}^M$, if $F^p(x)$ has a local minimum at x^*, then for all $i \in \mathbb{I}$,*

$$x^*_{\mathcal{G}_i} \neq 0, \quad \Longrightarrow \quad \begin{cases} \|x^*_{\mathcal{G}_i}\|_p \geq \left(\dfrac{\alpha q(1-q)}{2\|A\|_2^2}\right)^{\frac{1}{2-q}}, \qquad \textit{when} \quad 1 \leq p \leq 2, \\ \|x^*_{\mathcal{G}_i}\|_p \geq \left(\dfrac{\alpha q(1-q)}{2N_i^{1-2/p}\|A\|_2^2}\right)^{\frac{1}{2-q}}, \textit{when} \quad p > 2. \end{cases} \tag{15}$$

Proof. Denote $S = \text{supp}(x^*)$. Since x^* is a local minimizer of $F^p(x)$, x^* is also a local minimizer of the following problem:

$$\min_{x\in\mathbf{R}^N,\, x_{S^c}=0} \qquad F^p(x).$$

It follows that $y^* := x^*_S$ with $\left((y^*_{\bar{\mathcal{G}}_1})^T, \ldots, (y^*_{\bar{\mathcal{G}}_{\bar{n}}})^T\right)^T$ as a group formula of y^*, i.e., $\forall\, i = 1, 2, \ldots, \bar{n}$,

$$\exists\, l \in \mathbb{I}, \quad \text{such that} \quad y^*_{\bar{\mathcal{G}}_i} = x^*_{\mathcal{G}_l\cap S},$$

is a local minimizer of the following problem:

$$\min_{y\in\mathbf{R}^{\bar{N}}} \quad H(y) := \|A_S y - d\|^2 + \alpha \sum_{i=1}^{\bar{n}} \|y_{\bar{\mathcal{G}}_i}\|_p^q,$$

where $\bar{N} = \#S$, $\bar{N}_i = \#\bar{\mathcal{G}}_i$. Then, it is sufficient to find the lower bound for each $y^*_{\bar{\mathcal{G}}_i}$.

We can see that $H(y)$ is smooth at y^*, and its first order derivative at y^* is

$$2A_S^T(A_S y^* - d) + \alpha \begin{pmatrix} q\|y^*_{\bar{\mathcal{G}}_1}\|_p^{q-p} \left[\, |y^*_j|^{p-1}\text{sgn}(y^*_j) \,\right]_{j\in\bar{\mathcal{G}}_1} \\ \vdots \\ q\|y^*_{\bar{\mathcal{G}}_{\bar{n}}}\|_p^{q-p} \left[\, |y^*_j|^{p-1}\text{sgn}(y^*_j) \,\right]_{j\in\bar{\mathcal{G}}_{\bar{n}}} \end{pmatrix} = 0.$$

To obtain its second order derivative at y^*, for all $i = 1, 2, \ldots, \bar{n}$, we define

$$\phi_j(y) = q\|y_{\bar{\mathcal{G}}_i}\|_p^{q-p}|y_j|^{p-1}\text{sgn}(y_j), \quad \text{if} \quad j \in \bar{\mathcal{G}}_i.$$

Take $k \in \bar{\mathcal{G}}_i$. If $k \neq j$, we have

$$\frac{\partial \phi_j(y)}{\partial y_k} = q(q-p)\|y_{\bar{\mathcal{G}}_i}\|_p^{q-2p} \cdot |y_k|^{p-1}\text{sgn}(y_k) \cdot |y_j|^{p-1}\text{sgn}(y_j).$$

If $k = j$, we have when $p > 1$,

$$\frac{\partial \phi_j(y)}{\partial y_j} = q(q-p)\|y_{\bar{\mathcal{G}}_i}\|_p^{q-2p} \cdot |y_j|^{p-1}\text{sgn}(y_j) \cdot |y_j|^{p-1}\text{sgn}(y_j) \ + \ q(p-1)\|y_{\bar{\mathcal{G}}_i}\|_p^{q-p}|y_j|^{p-2};$$

when $p = 1$,

$$\frac{\partial \phi_j(y)}{\partial y_j} = q(q-p)\|y_{\bar{\mathcal{G}}_i}\|_p^{q-2p} \cdot |y_j|^{p-1}\text{sgn}(y_j) \cdot |y_j|^{p-1}\text{sgn}(y_j).$$

Thus, the second derivative of H at y^* is

$$\nabla^2 H(y^*) = 2A_S^T A_S + \alpha \begin{pmatrix} M^1 + D^1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & M^{\bar{n}} + D^{\bar{n}} \end{pmatrix} \succeq 0, \tag{16}$$

where $M^i, D^i \in \mathbf{R}^{\bar{N}_i \times \bar{N}_i}$,

$$M^i = q(q-p)\|y^*_{\bar{\mathcal{G}}_i}\|_p^{q-2p} \ \mu_i\mu_i^T, \qquad D^i = q(p-1)\|y^*_{\bar{\mathcal{G}}_i}\|_p^{q-p} \ \text{diag}(d_i),$$

and $\mu_i, d_i \in \mathbf{R}^{\bar{N}_i}$,

$$\mu_i = \begin{pmatrix} |(y^*_{\bar{\mathcal{G}}_i})_1|^{p-1}\text{sgn}(\ (y^*_{\bar{\mathcal{G}}_i})_1 \ ) \\ \vdots \\ |(y^*_{\bar{\mathcal{G}}_i})_{\bar{N}_i}|^{p-1}\text{sgn}((y^*_{\bar{\mathcal{G}}_i})_{\ \bar{N}_i} \ ) \end{pmatrix}, \quad d_i = \begin{pmatrix} |(y^*_{\bar{\mathcal{G}}_i})_1|^{p-2} \\ \vdots \\ |(y^*_{\bar{\mathcal{G}}_i})_{\bar{N}_i}|^{p-2} \end{pmatrix}.$$

Note that $D^i = 0$ when $p = 1$.

Let $i \in \{1, 2, \ldots \bar{n}\}$. We define $z \in \mathbf{R}^{\bar{N}}$ as

$$z_{\bar{\mathcal{G}}_l} = \begin{cases} y^*_{\bar{\mathcal{G}}_i}, \text{if } l = i, \\ 0, \quad \text{otherwise.} \end{cases}$$

Therefore,

$$\begin{aligned} 0 &\leq z^T \nabla^2 H(y^*) z \\ &= 2\|A_S z\|^2 + \alpha q(q-p)\|y^*_{\bar{\mathcal{G}}_i}\|_p^{q-2p}((y^*_{\bar{\mathcal{G}}_i})^T\mu_i)^2 + \alpha q(p-1)\|y^*_{\bar{\mathcal{G}}_i}\|_p^{q-p}(y^*_{\bar{\mathcal{G}}_i})^T\text{diag}(d_i)y^*_{\bar{\mathcal{G}}_i} \\ &= 2\|A_S z\|^2 + \alpha q(q-p)\|y^*_{\bar{\mathcal{G}}_i}\|_p^{q-2p}\|y^*_{\bar{\mathcal{G}}_i}\|_p^{2p} + \alpha q(p-1)\|y^*_{\bar{\mathcal{G}}_i}\|_p^{q-p}\|y^*_{\bar{\mathcal{G}}_i}\|_p^p \\ &\leq 2\|A_S\|_2^2\|y^*_{\bar{\mathcal{G}}_i}\|^2 + \alpha q(q-1)\|y^*_{\bar{\mathcal{G}}_i}\|_p^q, \end{aligned}$$

followed by,

$$\alpha q(1-q)\|y^*_{\bar{\mathcal{G}}_i}\|_p^q \leq 2\|A_S\|_2^2\|y^*_{\bar{\mathcal{G}}_i}\|^2.$$

If $p \leq 2$, then $\|y^*_{\bar{\mathcal{G}}_i}\|_2 \leq \|y^*_{\bar{\mathcal{G}}_i}\|_p$ by (7). Thus, combining it with $\|A_S\|_2 \leq \|A\|_2$, we can deduce

$$\|y^*_{\bar{\mathcal{G}}_i}\|_p \geq \left(\frac{\alpha q(1-q)}{2\|A\|_2^2}\right)^{\frac{1}{2-q}}.$$

If $p > 2$, then $\|y^*_{\bar{\mathcal{G}}_i}\|_2 \leq \bar{N}_i^{\frac{1}{2}-\frac{1}{p}}\|y^*_{\bar{\mathcal{G}}_i}\|_p$ by (7), implying

$$\|y^*_{\bar{\mathcal{G}}_i}\|_p \geq \left(\frac{\alpha q(1-q)}{2\bar{N}_i^{1-\frac{2}{p}}\|A\|_2^2}\right)^{\frac{1}{2-q}}.$$

Finally, (15) is obtained by using $y^* = x^*_{\text{supp}(x^*)}$. □

The corresponding result for the non-group case

$$f(x) := \|Ax - d\|^2 + \alpha\|x\|_q^q, \qquad \text{where} \quad \alpha > 0, \quad 0 < q < 1, \tag{17}$$

was already given in Theorem 2.1 of [33]. However, due to the group structure of x in $F(x)$, it is not trivial to generalize their result here.

4. The recovery bound

This section aims at conditions on the exponent q and matrix A to guarantee the local and global recovery for $F(x)$, e.g., to estimate the distance between the local (global) minimizers of $F(x)$ and the target signal x^o in (1).

Instead of the unconstrained $\ell_{2,q}$ group sparse minimization, most existing analysis focuses on recovering element t sparse x^o via constrained ℓ_q minimization. We therefore mainly review some results for ℓ_q minimization. Among the first is to ensure the uniqueness of the solution. It is easy to verify that the solution, at sparsity level t, of the linear system $Ax = d$ is unique provided that

$$\rho_{2t}(A) := \min_{\|x\|_0 \leq 2t} \frac{x^T A^T Ax}{\|x\|^2} > 0; \tag{18}$$

see [15] for more details. However, it is not enough to recover x^o from ℓ_q minimization. Note that $\rho_{2t} > 0$ is satisfied if A has RIP of order $2t$ with constant $\delta_{2t} \in (0,1)$ as defined in Definition 2.2. Most of the recovery analysis has been built under different assumptions on RIP condition. For instance, $\delta_t(A)+\delta_{2t}(A)+\delta_{3t}(A) < 1$ in [19], $\delta_{3t}(A)+3\delta_{4t}(A) < 2$ in [20], and $\delta_{2t}(A) < 0.3333, 0.4142, 0.4531, 0.4731$ in [21,22,14,23] respectively. Many random matrices with i.i.d. entries satisfy those requirements, but when $\delta_{2t} \to 1$, all of these conditions fail. To handle this problem, Sun proved that the exponent q can be chosen to be about $0.6796 \times (1 - \delta_{2t})$ so that x^o can be recovered from constrained ℓ_q minimization in [18]. For $\ell_{2,q}$ group sparse optimization, [27] also established the robust recovery for constrained $\ell_{2,q}$ model by using GRIP defined in Definition 2.3.

Inspired by the work above, we propose a recovery bound theory for $F(x)$ under GRIP assumption. Especially, by the means of the lower bound theory, we derive a recovery bound for the local minimizers.

We adopt a continuous function $b(q,C)$ with $(q,C) \in (0,1) \times (0,1)$ as [18]:

$$b(q,C) := C^{-1} \inf_{0<z<1} \max\left\{ \frac{1+zC}{(1+z^q C^q)^{1/q}}, \quad \sup_{\sqrt{2}(1-z)C/2 \le y \le 1} \frac{2y}{(1+2^{-q/2} y^{2+q})^{1/q}}, \right. \tag{19}$$

$$\left. \sup_{\sqrt{2}(1-z)C/2 \le y \le 1} \frac{3y}{(1+y)^{1/q}}, \quad \sup_{1 \le y} \frac{2y}{(1+y)^{1/q}} \right\}.$$

Lemma 4.1. *([18]) Let $0<q<1, s \ge 1$ be a positive integer, and let $\{z_j\}_{j\ge 1}$ be a finite decreasing sequence of nonnegative numbers with*

$$\sum_{k\ge 1} \left(\sum_{i=1}^{s} z_{ks+i}^2 \right)^{\frac{1}{2}} \ge C \left(\sum_{i=1}^{s} z_i^2 \right)^{\frac{1}{2}},$$

for some $C \in (0,1)$. Then,

$$\sum_{k\ge 1} \left(\sum_{i=1}^{s} z_{ks+i}^2 \right)^{\frac{1}{2}} \le C b(q,C) s^{\frac{1}{2}-\frac{1}{q}} \left(\sum_{j\ge 1} z_j^q \right)^{\frac{1}{q}}.$$

Lemma 4.2. *Let $\mathbb{G} \subset \mathbb{I}, T = \cup_{i\in\mathbb{G}} \mathcal{G}_i \subseteq \mathbb{J}$. Then, for any $x, y \in \mathbf{R}^N$, we have*

$$\|(x-y)_{T^c}\|_{2,q}^q \le \|(x-y)_T\|_{2,q}^q + (\,\|x\|_{2,q}^q - \|y\|_{2,q}^q + 2\|y_{T^c}\|_{2,q}^q\,). \tag{20}$$

Proof. It is due to

$$\begin{aligned} \|(x-y)_{T^c}\|_{2,q}^q &= \sum_{i\notin\mathbb{G}} \|(x-y)_{\mathcal{G}_i}\|^q \\ &\le \sum_{i\notin\mathbb{G}} (\|x_{\mathcal{G}_i}\| + \|y_{\mathcal{G}_i}\|)^q \\ &\le \|x_{T^c}\|_{2,q}^q + \|y_{T^c}\|_{2,q}^q \\ &= \|x\|_{2,q}^q - \|x_T\|_{2,q}^q - \|y\|_{2,q}^q + \|y_T\|_{2,q}^q + 2\|y_{T^c}\|_{2,q}^q \\ &\le \|(x-y)_T\|_{2,q}^q + (\|x\|_{2,q}^q - \|y\|_{2,q}^q + 2\|y_{T^c}\|_{2,q}^q). \quad \square \end{aligned}$$

With preparations above, we can show the following result.

Theorem 4.3. *Let t be an integer with $2t \le n$, and $d = Ax^o + \xi$ with $\|\xi\| = \varepsilon$. Then the following statements hold.*
(i) *(global recovery bound) Assume A satisfies* GRIP *of order $2t$ with $\delta'_{2t} \in (0,1)$ and*

$$(q, \delta'_{2t}) \quad \in \quad \{(q,\delta'_{2t}) \,:\, C_{\delta'} < 1\} \cup \{(q,\delta'_{2t}) \,:\, C_{\delta'} > 1, b(q, 1/C_{\delta'}) < 1\}, \tag{21}$$

where $C_{\delta'} = \frac{(1+\sqrt{2})\delta'_{2t}}{2(1-\delta'_{2t})}$. Then for all $x \in \mathbf{R}^N$ satisfying $F(x) \le F(x^o)$, we have

$$\|x - x^o\|_{2,q}^q \le c_1 \frac{\varepsilon^2}{\alpha} + c_2 \sigma_{t,q}(x^o) + c_3 t^{1-\frac{q}{2}} \varepsilon^q + c_4 t^{1-\frac{q}{2}} F(x^o)^{\frac{q}{2}}, \tag{22}$$

where c_1, c_2, c_3, c_4 are positive constants depending on A, q.

(ii) *(local recovery bound) Let x be a local minimizer of $F(x)$. Under the assumptions in* (i) *and that the columns of $A^T_{supp(x)}$ are linearly independent, we have*

$$\|x - x^o\|^q_{2,q} \leq c_1 \frac{\varepsilon^2}{\alpha} + c_2 \sigma_{t,q}(x^o) + c_3 t^{1-\frac{q}{2}} \varepsilon^q + c_5 t^{1-\frac{q}{2}} n^{\frac{q}{2}} \alpha^{\frac{q}{2-q}}; \tag{23}$$

or, under the assumptions that the columns of $A_{\mathrm{supp}(x)}$ are linearly independent and $supp(x^o) \subseteq supp(x)$, we have

$$\|x - x^o\| \leq c'_1 \varepsilon + c'_2 \alpha^{\frac{1}{2-q}}, \tag{24}$$

where c_5, c'_1, c'_2 are positive constants depending on A, q.

Proof. *(i).* For better presentation, we set $\eta := x - x^o$,

$$\mathbb{G} := \{\text{indices of the first } t \text{ largest groups of } x^o\} \subset \mathbb{I}, \quad T = \bigcup_{i \in \mathbb{G}} \mathcal{G}_i \subset \mathbb{J},$$

and partition the complement of T as $T^c = T_1 \cup T_2 \cup \cdots$, where

$$\mathbb{G}_1 := \{\text{indices of the } t \text{ largest groups of } \eta \text{ in } T^c\ \} \subset \mathbb{I}, \quad T_1 = \bigcup_{i \in \mathbb{G}_1} \mathcal{G}_i \subset \mathbb{J}, \tag{25}$$

$$\mathbb{G}_2 := \{\text{indices of the next } t \text{ largest groups of } \eta \text{ in } T^c\ \} \subset \mathbb{I}, \quad T_2 = \bigcup_{i \in \mathbb{G}_2} \mathcal{G}_i \subset \mathbb{J},$$

$$\vdots$$

By Lemma 4.2, we obtain

$$\|\eta_{T^c}\|^q_{2,q} \leq \|\eta_T\|^q_{2,q} + (\ \|x\|^q_{2,q} - \|x^o\|^q_{2,q} + 2\|x^o_{T^c}\|^q_{2,q}\ ).$$

It follows from $F(x) \leq F(x^o)$ that

$$\|x\|^q_{2,q} \leq \frac{1}{\alpha} F(x) \leq \frac{1}{\alpha} F(x^o) = \frac{\varepsilon^2}{\alpha} + \|x^o\|^q_{2,q}.$$

Thus, we get

$$\|\overline{\eta_{T^c}}\|^q_{2,q} \leq \|\overline{\eta_T}\|^q_{2,q} + \frac{\varepsilon^2}{\alpha} + 2\sigma_{t,q}(x^o), \tag{26}$$

and

$$\|\eta\|^q_{2,q} = \|\overline{\eta_T}\|^q_{2,q} + \|\overline{\eta_{T^c}}\|^q_{2,q} \leq 2\|\overline{\eta_T}\|^q_{2,q} + \frac{\varepsilon^2}{\alpha} + 2\sigma_{t,q}(x^o). \tag{27}$$

Step 1: find a preliminary bound for $\|\overline{\eta_T}\|_{2,q}$.
We first observe that

$$\begin{aligned}\|\overline{\eta_T}\|^2 + \|\overline{\eta_{T_1}}\|^2 &= \|\overline{\eta_T} + \overline{\eta_{T_1}}\|^2 \leq \frac{1}{(1-\delta'_{2t})} \|A(\overline{\eta_T} + \overline{\eta_{T_1}})\|^2 \\ &= \frac{1}{(1-\delta'_{2t})} \big\langle A(\eta - \overline{\eta_{T_2}} - \overline{\eta_{T_3}} - \cdots), A(\overline{\eta_T} + \overline{\eta_{T_1}}) \big\rangle\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{(1-\delta'_{2t})}\Big\{ \langle A\eta, A(\overline{\eta_T}+\overline{\eta_{T_1}})\rangle + \sum_{k\geq 2}\big[\, \langle A(-\overline{\eta_{T_k}}), A\overline{\eta_T}\rangle + \langle A(-\overline{\eta_{T_k}}), A(\overline{\eta_{T_1}})\rangle \,\big]\Big\} \\
[\text{ Equ. } (12)\,] &\leq \frac{1}{(1-\delta'_{2t})}\Big\{ \|A\eta\|\cdot\|A(\overline{\eta_T}+\overline{\eta_{T_1}})\| + \delta'_{2t}\sum_{k\geq 2}\big[\, \|\overline{\eta_{T_k}}\|\cdot\|\overline{\eta_T}\| + \|\overline{\eta_{T_k}}\|\cdot\|\overline{\eta_{T_1}}\| \,\big]\Big\} \\
&\leq \frac{1}{(1-\delta'_{2t})}\Big\{ \|A\eta\|\cdot\sqrt{1+\delta'_{2t}}\|\overline{\eta_T}+\overline{\eta_{T_1}}\| + \delta'_{2t}\sum_{k\geq 2}\|\overline{\eta_{T_k}}\|\cdot(\|\overline{\eta_T}\|+\|\overline{\eta_{T_1}}\|)\Big\} \\
&\leq \left(\frac{\sqrt{1+\delta'_{2t}}}{(1-\delta'_{2t})}\|A\eta\| + \frac{\delta'_{2t}}{(1-\delta'_{2t})}\sum_{k\geq 2}\|\overline{\eta_{T_k}}\|\right)(\|\overline{\eta_T}\|+\|\overline{\eta_{T_1}}\|),
\end{aligned}
$$

where [Equ. (12)] denotes Equation (12) as the reason for next step. We use this reasoning notation throughout the paper, when needed. Denote

$$\zeta := \frac{\sqrt{1+\delta'_{2t}}}{1-\delta'_{2t}}\|A\eta\|, \qquad \beta := \frac{\delta'_{2t}}{1-\delta'_{2t}}, \qquad \Sigma := \sum_{k\geq 2}\|\overline{\eta_{T_k}}\|.$$

Then we have

$$\left[\|\overline{\eta_T}\| - \frac{\zeta+\beta\Sigma}{2}\right]^2 + \left[\|\overline{\eta_{T_1}}\| - \frac{\zeta+\beta\Sigma}{2}\right]^2 \leq \frac{(\zeta+\beta\Sigma)^2}{2},$$

which implies

$$\|\overline{\eta_T}\| \leq \frac{1+\sqrt{2}}{2}(\zeta+\beta\Sigma), \quad \text{and} \quad \|\overline{\eta_{T_1}}\| \leq \frac{1+\sqrt{2}}{2}(\zeta+\beta\Sigma). \tag{28}$$

It follows from (8) and $q<2$ that

$$\|\overline{\eta_T}\|_{2,q} \leq t^{\frac{1}{q}-\frac{1}{2}}\|\overline{\eta_T}\|_2 \leq \frac{1+\sqrt{2}}{2}t^{\frac{1}{q}-\frac{1}{2}}\zeta + C_{\delta'}t^{\frac{1}{q}-\frac{1}{2}}\Sigma. \tag{29}$$

Step 2: find the bound of Σ in three cases.

Take $k\geq 1$. Then, by (25), we have $\forall\, i\in\mathbb{G}_k, \|\eta_{\mathcal{G}_i}\| \geq \max\{\|\eta_{\mathcal{G}_l}\|, l\in\mathbb{G}_{k+1}\} = \|\overline{\eta_{T_{k+1}}}\|_{2,\infty}$, which indicates

$$\|\overline{\eta_{T_k}}\|_{2,q} = \Big(\sum_{i\in\mathbb{G}_k}\|\eta_{\mathcal{G}_i}\|^q\Big)^{\frac{1}{q}} \geq t^{\frac{1}{q}}\|\overline{\eta_{T_{k+1}}}\|_{2,\infty} \overset{[\text{Equ. }(8)]}{\geq} t^{\frac{1}{q}-\frac{1}{2}}\|\overline{\eta_{T_{k+1}}}\|_{2,2} = t^{\frac{1}{q}-\frac{1}{2}}\|\overline{\eta_{T_{k+1}}}\|_2.$$

Therefore,

$$\Sigma = \sum_{k\geq 2}\|\overline{\eta_{T_k}}\| \leq t^{\frac{1}{2}-\frac{1}{q}}\sum_{k\geq 1}\|\overline{\eta_{T_k}}\|_{2,q} \overset{[\text{Equ. }(7)]}{\leq} t^{\frac{1}{2}-\frac{1}{q}}\Big(\sum_{k\geq 1}\|\overline{\eta_{T_k}}\|_{2,q}^q\Big)^{\frac{1}{q}} = t^{\frac{1}{2}-\frac{1}{q}}\|\overline{\eta_{T^c}}\|_{2,q}. \tag{30}$$

The bound for Σ in (30) always holds. However, when $C_{\delta'}>1$, we have a better bound. Since b is a continuous function and $b(q, 1/C_{\delta'})<1$, there must exists $e>0$ such that

$$0 < \frac{1}{(1+e)C_{\delta'}} < 1, \quad \text{and} \quad b\Big(q, \frac{1}{(1+e)C_{\delta'}}\Big) < 1.$$

If $\zeta \geq e\beta\Sigma$, then

$$\Sigma \le \frac{\zeta}{e\beta}. \tag{31}$$

If $\zeta < e\beta\Sigma$, by (28), one has

$$\|\overline{\eta_{T_1}}\| \le (1+e)\frac{(1+\sqrt{2})\delta'_{2t}}{2(1-\delta'_{2t})}\Sigma, \quad \text{i.e.,} \quad \Sigma = \sum_{k\ge 2}\left(\sum_{i\in\mathbb{G}_k}\|\eta_{\mathcal{G}_i}\|^2\right)^{\frac{1}{2}} \ge \frac{1}{(1+e)C_{\delta'}}\left(\sum_{i\in\mathbb{G}_1}\|\eta_{\mathcal{G}_i}\|^2\right)^{\frac{1}{2}}.$$

Recall Lemma 4.1. It implies

$$\Sigma \le \frac{1}{(1+e)C_{\delta'}} b\Big(q, \frac{1}{(1+e)C_{\delta'}}\Big) t^{\frac{1}{2}-\frac{1}{q}} \|\overline{\eta_{T^c}}\|_{2,q}. \tag{32}$$

Now we have obtained the bound (30) (31) (32) for Σ. Thus, we are ready to give the bound for $\|\eta\|_{2,q}^q$, except for the bound of ζ left in step 4.

Step 3(a): find the bound of $\|\eta\|_{2,q}^q$ when $C_{\delta'} < 1$.

Since $C_{\delta'} < 1$, combining (29) and (30) gives

$$\|\overline{\eta_T}\|_{2,q} \le \frac{1+\sqrt{2}}{2} t^{\frac{1}{q}-\frac{1}{2}}\zeta + C_{\delta'}\|\overline{\eta_{T^c}}\|_{2,q}. \tag{33}$$

Substituting (26) into it, we have

$$\|\overline{\eta_T}\|_{2,q}^q \le \left(\frac{1+\sqrt{2}}{2}\right)^q t^{1-\frac{q}{2}}\zeta^q + (C_{\delta'})^q \left(\|\overline{\eta_T}\|_{2,q}^q + \frac{\varepsilon^2}{\alpha} + 2\sigma_{t,q}(x^o)\right). \tag{34}$$

Therefore, we obtain

$$\|\overline{\eta_T}\|_{2,q}^q \le \frac{1}{1-\gamma}\left(\frac{1+\sqrt{2}}{2}\right)^q t^{1-\frac{q}{2}}\zeta^q + \frac{\gamma}{1-\gamma}\left(\frac{\varepsilon^2}{\alpha} + 2\sigma_{t,q}(x^o)\right),$$

where $\gamma := (C_{\delta'})^q < 1$. Finally, plugging it into (27) gives

$$\|\eta\|_{2,q}^q \le \frac{1+\gamma}{1-\gamma}\cdot\frac{\varepsilon^2}{\alpha} + 2\frac{1+\gamma}{1-\gamma}\sigma_{t,q}(x^o) + \frac{2}{1-\gamma}\left(\frac{1+\sqrt{2}}{2}\right)^q t^{1-\frac{q}{2}}\zeta^q. \tag{35}$$

Step 3(b): find the bound of $\|\eta\|_{2,q}^q$ when $C_{\delta'} > 1$ and $\zeta \ge e\beta\Sigma$.

Under the assumption, combining (29) and (31) yields

$$\|\overline{\eta_T}\|_{2,q} \le t^{\frac{1}{q}-\frac{1}{2}}\|\overline{\eta_T}\|_2 \le \frac{1+\sqrt{2}}{2}\left(1+\frac{1}{e}\right)t^{\frac{1}{q}-\frac{1}{2}}\zeta,$$

which indicates

$$\|\eta\|_{2,q}^q \le \frac{\varepsilon^2}{\alpha} + 2\sigma_{t,q}(x^o) + 2\left(\frac{1+\sqrt{2}}{2}\right)^q\left(1+\frac{1}{e}\right)^q t^{1-\frac{q}{2}}\zeta^q. \tag{36}$$

Step 3(c): find the bound of $\|\eta\|_{2,q}^q$ when $C_{\delta'} > 1$ and $\zeta < e\beta\Sigma$.

Under the assumption, combining (29) and (32) yields

$$\|\overline{\eta_T}\|_{2,q} \le \frac{1+\sqrt{2}}{2}t^{\frac{1}{q}-\frac{1}{2}}\zeta + \frac{1}{1+e}b\Big(q, \frac{1}{(1+e)C_{\delta'}}\Big)\|\overline{\eta_{T^c}}\|_{2,q}.$$

Substituting (26) into it gives

$$\|\overline{\eta_T}\|_{2,q}^q \le \left(\frac{1+\sqrt{2}}{2}\right)^q t^{1-\frac{q}{2}}\zeta^q + \frac{1}{(1+e)^q} b\Big(q, \frac{1}{(1+e)C_{\delta'}}\Big)^q \left(\|\overline{\eta_T}\|_{2,q}^q + \frac{\varepsilon^2}{\alpha} + 2\sigma_{t,q}(x^o)\right), \tag{37}$$

and thus,

$$\|\overline{\eta_T}\|_{2,q}^q \le \frac{1}{1-\gamma'}\left(\frac{1+\sqrt{2}}{2}\right)^q t^{1-\frac{q}{2}}\zeta^q + \frac{\gamma'}{1-\gamma'}\left(\frac{\varepsilon^2}{\alpha} + 2\sigma_{t,q}(x^o)\right),$$

where $\gamma' := \frac{1}{(1+e)^q} b\Big(q, \frac{1}{(1+e)C_{\delta'}}\Big)^q < 1$. Finally,

$$\|\eta\|_{2,q}^q \le \frac{1+\gamma'}{1-\gamma'}\cdot\frac{\varepsilon^2}{\alpha} + 2\frac{1+\gamma'}{1-\gamma'}\sigma_{t,q}(x^o) + \frac{2}{1-\gamma'}\left(\frac{1+\sqrt{2}}{2}\right)^q t^{1-\frac{q}{2}}\zeta^q. \tag{38}$$

Step 4: find the bound of ζ^q.
Note that $d = Ax^o + \xi$. Then,

$$\|Ax - d\|^2 = \|Ax - (Ax^o + \xi)\|^2 = \|A\eta - \xi\|^2 \ge (\|A\eta\| - \varepsilon)^2,$$

implying

$$\|A\eta\| \le \varepsilon + \|Ax - d\|.$$

Since $F(x) \le F(x^o)$, we have $F(x^o) \ge F(x) \ge \|Ax - d\|^2$, and

$$\|A\eta\| \le \varepsilon + \sqrt{F(x^o)}.$$

Therefore,

$$\zeta^q = \left(\frac{\sqrt{1+\delta'_{2t}}}{1-\delta'_{2t}}\|A\eta\|\right)^q \le \left(\frac{\sqrt{1+\delta'_{2t}}}{1-\delta'_{2t}}\right)^q \left(\varepsilon^q + F(x^o)^{\frac{q}{2}}\right). \tag{39}$$

Step 5: conclusion.
From step 3 and 4, one can see that if $F(x) \le F(x^o)$, there exist $c_1, c_2, c_3, c_4 > 0$ such that

$$\|\eta\|_{2,q}^q \le c_1\frac{\varepsilon^2}{\alpha} + c_2\sigma_{t,q}(x^o) + c_3 t^{1-\frac{q}{2}}\varepsilon^q + c_4 t^{1-\frac{q}{2}} F(x^o)^{\frac{q}{2}}.$$

This completes the proof of (i).
(ii) Let x be a local minimizer. Denote $\mathbb{G} = \text{gsupp}(x)$ and $S = \text{supp}(x)$. Since x is a local minimizer, it satisfies the first order necessary condition:

$$2A_S^T(A_S x_S - d) + \alpha w_S = 0,$$

where $w_j = q\|x_{\mathcal{G}_i}\|^{q-2}x_j$ for any $j \in S \cap \mathcal{G}_i, i \in \mathbb{G}$.

Assume that the columns of A_S^T are linearly independent and the assumptions in (i) hold. Like $\varrho(A)$ in (10), we introduce

$$\varrho'(A) := \min\left\{\frac{\|A_{S'}^T y\|^2}{\|y\|^2} : y \in \mathbf{R}^M, S' \subseteq \mathbb{J} \text{ such that } \det\left(A_{S'}A_{S'}^T\right) \ne 0\right\} > 0.$$

Back to the step 4 in the proof of (i), we have,

$$\begin{aligned}\varrho'(A)\|(Ax-d)\|^2 &\leq \|A_S^T(Ax-d)\|^2\\ &= \frac{\alpha}{2}\|w_S\|\\ &= \frac{\alpha^2q^2}{4}\sum_{x_{\mathcal{G}_i}\neq 0}\|x_{\mathcal{G}_i}\|^{2q-2}\\ [\text{ Theorem 3.1 }]\quad &\leq \frac{\alpha^2q^2n}{4}\left(\frac{\alpha q(1-q)}{2\|A\|_2^2}\right)^{\frac{2q-2}{2-q}}.\end{aligned}$$

It then follows that,

$$\|A\eta\| \leq \varepsilon + \frac{q}{2\sqrt{\varrho'(A)}}\left(\frac{q(1-q)}{2\|A\|_2^2}\right)^{\frac{q-1}{2-q}} n^{\frac{1}{2}}\alpha^{\frac{1}{2-q}},$$

and

$$\zeta^q \leq \left(\frac{\sqrt{1+\delta'_{2t}}}{1-\delta'_{2t}}\right)^q\left[\varepsilon^q + \left(\frac{q}{2\sqrt{\varrho'(A)}}\left(\frac{q(1-q)}{2\|A\|_2^2}\right)^{\frac{q-1}{2-q}}\right)^q n^{\frac{q}{2}}\alpha^{\frac{q}{2-q}}\right]. \tag{40}$$

Similar to the proof of (i), there exists $c_5 > 0$ such that

$$\|\eta\|_{2,q}^q \leq c_1\frac{\varepsilon^2}{\alpha} + c_2\sigma_{t,q}(x^o) + c_3t^{1-\frac{q}{2}}\varepsilon^q + c_5t^{1-\frac{q}{2}}n^{\frac{q}{2}}\alpha^{\frac{q}{2-q}}.$$

Now, we assume that the columns of A_S are linearly independent and $\text{supp}(x^o) \subseteq \text{supp}(x)$. Since $\text{supp}(x^o) \subseteq S$, we have $d = A_S x_S^o + \xi$ and

$$2A_S^T(A_Sx_S - A_Sx_S^o - \xi) + \alpha w_S = 0.$$

It follows that

$$x_S - x_S^o = (A_S^TA_S)^{-1}A_S^T\xi - \frac{\alpha}{2}(A_S^TA_S)^{-1}w_S.$$

Thus,

$$\begin{aligned}\|x_S - x_S^o\| &\leq \|(A_S^TA_S)^{-1}A_S^T\xi\| + \frac{\alpha}{2}\|(A_S^TA_S)^{-1}w_S\|\\ &\leq \|(A_S^TA_S)^{-1}A_S^T\|\cdot\|\xi\| + \frac{\alpha}{2}\|(A_S^TA_S)^{-1}\|\cdot\|w_S\|\\ &\leq \frac{\|A\|}{\varrho(A)}\varepsilon + \frac{\alpha q}{2\varrho(A)}n^{\frac{1}{2}}\left(\frac{\alpha q(1-q)}{2\|A\|_2^2}\right)^{\frac{q-1}{2-q}}.\end{aligned}$$

Therefore, there exist $c'_1, c'_2 > 0$ such that

$$\|x - x^o\| \leq c'_1\varepsilon + c'_2\alpha^{\frac{1}{2-q}}.$$

□

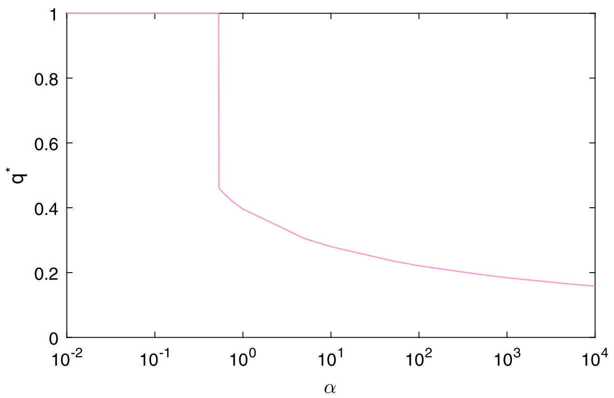

Fig. 1. The optimal q^* to minimize the recovery bound $R(q;\alpha)$ in (41) in terms of α under the setting of Remark 4.4.

Remark 4.4. Given $A \in \mathbf{R}^{M\times N}$ and noise level ε. We can see that the recovery bound, the right hand of (23), is related to the regularization parameter α, and q. Then, we define the recovery bound in (23) as a function in terms of α and q:

$$R(\alpha, q) := c_1(q)\frac{\varepsilon^2}{\alpha} + c_2(q)\sigma_{t,q}(x^o) + c_3(q)t^{1-\frac{q}{2}}\varepsilon^q + c_5(q)t^{1-\frac{q}{2}}n^{\frac{q}{2}}\alpha^{\frac{q}{2-q}}, \tag{41}$$

where $\alpha > 0, q \in (0, 1)$. For any given q, we can obtain that there exists an optimal α^* such that

$$\alpha^* = \arg\min_{\alpha} R(\alpha; q) = \left(\frac{(2-q)c_1(q)\varepsilon^2}{qc_5(q)t^{1-\frac{q}{2}}n^{\frac{q}{2}}}\right)^{\frac{2-q}{2}},$$

since $\partial_\alpha R(\alpha; q) < 0$ when $\alpha \to 0$; $\partial_\alpha R(\alpha; q) > 0$ when $\alpha \to \infty$; and $\partial_\alpha R(\alpha; q) = 0$ has only one solution. However, for any given α, an optimal q^* to minimize $R(q; \alpha)$ is hard to obtain. Thus, we show the optimal q^* in terms of α for a specific example in Fig. 1. We set $n = 120, t = 12, \delta'_{2t} = 0.4, \varepsilon = 0.1, \varrho'(A) = 0.1, \|A\|_2 = 2, \sigma_{t,q}(x^o) = 0.001$. As can be seen, the optimal q^* is decreasing in terms of α. When α is small, the optimal q^* is nearly 1; as α approaches 0.5, the optimal q^* drops rapidly to 0.46; then, q^* continues to decrease. Unfortunately, we have to admit that due to many inequalities used in our proof, this recovery bound is hard to provide a good reference of optimal parameters; see also [15,18] and references therein.

Remark 4.5. In Theorem 4.3, $C_{\delta'} < 1$ means $\delta'_{2t} < 0.4531$. The feasible set of $(q, \delta'_{2t}) \in (0, 1) \times (0, 1)$ satisfying (21) is depicted in Fig. 2. We can see that whatever δ'_{2t} is, there exists some $q \in (0, 1)$ such that any group sparse signal can be recovered approximately from its noisy measurements via solving $\ell_{2,q}$ regularized minimization. This result is much stronger than the recovery theory for Group Lasso ($\ell_{2,1}$) minimization in [26] which holds only when $\delta'_{2t} < 0.414$. Thus, the nonconvex minimization method can enhance performance of group sparsity recovery.

In the proof of the recovery bound of local minimizer, the lower bound theory of local minimizers helps to make $\|A\eta\|$ bounded by a term of parameter α, instead of function value $F(x^o)$ which is always unknown. Since the $\ell_{2,q}$ regularization is nonconvex, the global minimizer is barely obtained, and thus this local recovery bound is meaningful. Furthermore, we can see from the recovery bound that less noise and smaller sparsity level of x^o both lead to a smaller distance between minimizers of $F(x)$ and x^o.

For an intuitive understanding of our recovery bound, we give the following corollary.

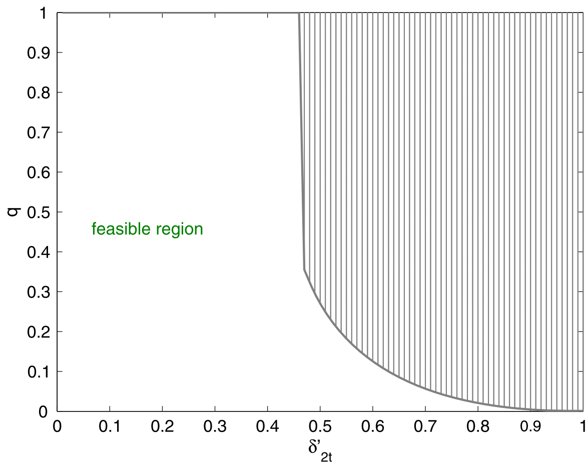

Fig. 2. The feasible region of $(q, \delta'_{2t}) \in (0, 1) \times (0, 1)$ satisfying (21).

Corollary 4.6. *(Recovery bound for noiseless data) Let t be an integer with $2t \le n$ and $d = Ax^o$ with $\|x^o\|_{2,0} \le t$. Assume that A satisfies* GRIP *of order $2t$ with $\delta'_{2t} \in (0, 1)$ and (q, δ'_{2t}) satisfies (21). Then for all $x \in \mathbf{R}^N$ with $F(x) \le F(x^o)$, we have*

$$\|x - x^o\|_{2,q}^q \le \tilde{c}_1 t^{1-\frac{q}{2}} F(x^o)^{\frac{q}{2}},$$

where $\tilde{c}_1$ is a positive constant depending on A, q. Furthermore, let x be a local minimizer of $F(x)$. Under the same assumptions and that the columns of $A^T_{supp(x)}$ are linearly independent, we have

$$\|x - x^o\|_{2,q}^q \le \tilde{c}_2 t^{1-\frac{q}{2}} n^{\frac{q}{2}} \alpha^{\frac{q}{2-q}};$$

On the other hand, under the assumptions that the columns of $A_{\mathrm{supp}(x)}$ are linearly independent and $supp(x^o) \subseteq supp(x)$, we have

$$\|x - x^o\| \le \tilde{c}' \alpha^{\frac{1}{2-q}},$$

where $\tilde{c}_2, \tilde{c}'$ are positive constants depending on A, q.

Remark 4.7. Since an optimal solution x^* of $F(x)$ is also a local minimizer of $F(x)$ and satisfies $F(x^*) \le F(x^o)$, the recovery bound for global minimizers is obvious.

From Corollary 4.6, we can see that the sparse solution x^o can be recovered near perfectly by the solutions of $F(x)$, as long as α is sufficiently small. Applying Lemma 2.1, we have the order of our recovery bound

$$\|x - x^o\|^2 = \mathcal{O}(\alpha^{\frac{2}{2-q}}), \tag{42}$$

which equals the one under group restricted eigenvalue condition of A assumptions (see Theorem 9 in [15]). When q approaches 1, this bound approaches the classical recovery bound $\mathcal{O}(\alpha^2)$ for group Lasso ($\ell_{2,1}$ minimization) under GRIP assumptions [34,35].

Remark 4.8. For non-group case $f(x)$ in (17), the columns of $A_{\mathrm{supp}(x^*)}$ where x^* is a local minimizer are linearly independent [33]. Thus, a local recovery bound, like (24), of $f(x)$ can be easily obtained under $\mathrm{supp}(x^o) \subseteq \mathrm{supp}(x^*)$.

5. From a stationary point to a local minimizer

Due to the nonconvexity of $F(x)$, most algorithms can only find a stationary point of $F(x)$. Although a local minimizer is a stationary point, the inverse is usually not true. Thus, it is important to find when a stationary point is a local minimizer.

We begin with an easy but instructive result.

Proposition 5.1. *For any $d \in \mathbf{R}^M, \alpha > 0, A \in \mathbf{R}^{M\times N}$, 0 is a strict local minimizer of $F(x)$.*

Proof. It follows from

$$\begin{aligned} F(x) - F(0) &= \|Ax - d\|^2 + \alpha\|x\|_{2,q}^q - \|d\|^2 \\ &\geq -2d^T Ax + \alpha\|x\|_{2,q}^q \\ &= \sum_{i\in\mathbb{I}, x_{\mathcal{G}_i}\neq 0} \left[\alpha\|x_{\mathcal{G}_i}\|^q - 2(d^T A)_{\mathcal{G}_i} x_{\mathcal{G}_i}\right] \\ &\geq \sum_{i\in\mathbb{I}, x_{\mathcal{G}_i}\neq 0} \left[\alpha\|x_{\mathcal{G}_i}\|^q - 2\|(d^T A)_{\mathcal{G}_i}\| \cdot \|x_{\mathcal{G}_i}\|\right] \\ &= \sum_{i\in\mathbb{I}, x_{\mathcal{G}_i}\neq 0} \|x_{\mathcal{G}_i}\| \left(\alpha\|x_{\mathcal{G}_i}\|^{q-1} - 2\|(d^T A)_{\mathcal{G}_i}\|\right) \\ &> 0, \end{aligned}$$

when x is sufficiently close to 0. □

Remark 5.2. In a similar way, we can prove that 0 is a trivial local minimizer of $F^p(x)$ defined in (14).

Since 0 is a trivial local minimizer of $F(x)$, initialization with zero in algorithms is not suitable.

Theorem 5.3. *Suppose that x^* is a stationary point of $F(x)$ with the columns of $A_{\mathrm{supp}(x^*)}$ linearly independent. If for all $i \in \mathrm{gsupp}(x^*)$,*

$$\|x^*_{\mathcal{G}_i}\| > \left(\frac{\alpha q(1-q)}{2\varrho(A)}\right)^{\frac{1}{2-q}}, \tag{43}$$

then x^* is a strict local minimizer of $F(x)$.

Proof. We denote

$$\mathbb{G}^* := \mathrm{gsupp}(x^*), \qquad \tilde{S} := \bigcup_{i\in\mathbb{G}^*} \mathcal{G}_i, \qquad \tilde{n} := \#\mathbb{G}^*, \qquad \tilde{y}^* := x^*_{\tilde{S}},$$

with $\left((\tilde{y}^*_{\tilde{\mathcal{G}}_1})^T, \ldots, (\tilde{y}^*_{\tilde{\mathcal{G}}_{\tilde{n}}})^T\right)^T$ as a group structure of $\tilde{y}^* \in \mathbf{R}^{\tilde{N}}$, i.e., $\forall\, i = 1, 2, \ldots, \tilde{n}$,

$$\exists\, l \in \mathbb{I}, \quad \text{such that} \quad \tilde{y}^*_{\tilde{\mathcal{G}}_i} = x^*_{\mathcal{G}_l},$$

where $\tilde{N} = \#\tilde{S}$, $\tilde{N}_i = \#\tilde{\mathcal{G}}_i$. Note that $\left((\tilde{y}^*_{\tilde{\mathcal{G}}_1})^T, \ldots, (\tilde{y}^*_{\tilde{\mathcal{G}}_{\tilde{n}}})^T\right)^T$ is different from $\left((y^*_{\bar{\mathcal{G}}_1})^T, \ldots, (y^*_{\bar{\mathcal{G}}_{\bar{n}}})^T\right)^T$ in the proof of Theorem 3.1. Since x^* is a stationary point of $F(x)$, it is not hard to check that $\tilde{y}^*$ is a stationary point of $\tilde{H}(y)$ defined as

$$\tilde{H}(y) := \|A_{\tilde{S}}y - d\|^2 + \alpha \sum_{i=1}^{\tilde{n}} \|y_{\tilde{\mathcal{G}}_i}\|^q.$$

Similar to (16), we have $\tilde{H}(y)$ is smooth at $\tilde{y}^*$ with

$$\nabla^2 \tilde{H}(\tilde{y}^*) = 2A_{\tilde{S}}^T A_{\tilde{S}} + \alpha \begin{pmatrix} \tilde{M}^1 + \tilde{D}^1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \tilde{M}^{\tilde{n}} + \tilde{D}^{\tilde{n}} \end{pmatrix},$$

where $\tilde{M}^i, \tilde{D}^i \in \mathbf{R}^{\tilde{N}_i \times \tilde{N}_i}$,

$$\tilde{M}^i = q(q-2)\|\tilde{y}^*_{\tilde{\mathcal{G}}_i}\|^{q-4}\ \tilde{y}^*_{\tilde{\mathcal{G}}_i}(\tilde{y}^*_{\tilde{\mathcal{G}}_i})^T, \qquad \tilde{D}^i = q\|\tilde{y}^*_{\tilde{\mathcal{G}}_i}\|^{q-2} I.$$

Clearly, for any $z \in \mathbf{R}^{\tilde{N}}$,

$$\begin{aligned} z^T \nabla^2 \tilde{H}(\tilde{y}^*) z &= 2\|A_{\tilde{S}}z\|^2 + \sum_{i=1}^{\tilde{n}} \left[-\alpha q(2-q)\|\tilde{y}^*_{\tilde{\mathcal{G}}_i}\|^{q-4}(z^T_{\tilde{\mathcal{G}}_i}\tilde{y}^*_{\tilde{\mathcal{G}}_i})^2 + \alpha q \|\tilde{y}^*_{\tilde{\mathcal{G}}_i}\|^{q-2} z^T_{\tilde{\mathcal{G}}_i} z_{\tilde{\mathcal{G}}_i} \right] \\ &\geq 2\|A_{\tilde{S}}z\|^2 + \sum_{i=1}^{\tilde{n}} \left[-\alpha q(2-q)\|\tilde{y}^*_{\tilde{\mathcal{G}}_i}\|^{q-2}\|z_{\tilde{\mathcal{G}}_i}\|^2 + \alpha q \|\tilde{y}^*_{\tilde{\mathcal{G}}_i}\|^{q-2}\|z_{\tilde{\mathcal{G}}_i}\|^2 \right] \\ &\geq 2\varrho(A)\|z\|^2 - \sum_{i=1}^{\tilde{n}} \alpha q(1-q)\|\tilde{y}^*_{\tilde{\mathcal{G}}_i}\|^{q-2}\|z_{\tilde{\mathcal{G}}_i}\|^2 \\ [\text{ Equ. (43) }] \quad &> 2\varrho(A)\|z\|^2 - \alpha q(1-q)\left[\left(\frac{\alpha q(1-q)}{2\varrho(A)} \right)^{\frac{1}{2-q}} \right]^{q-2} \|z\|^2 \\ &= 0. \end{aligned}$$

Hence, $\nabla^2 \tilde{H}(\tilde{y}^*)$ is positive definite and $\tilde{y}^*$ is a strict local minimizer of $\tilde{H}(y)$.

Let $v \in \mathbf{R}^N$ be sufficiently close to 0. Then,

$$\begin{aligned} F(x^* + v) &= \|A(x^* + v) - d\|^2 + \alpha \|x^* + v\|^q_{2,q} \\ &= \|A_{\tilde{S}^c} v_{\tilde{S}^c} - (d - A_{\tilde{S}}(x^* + v)_{\tilde{S}})\|^2 + \alpha \sum_{i \in (\mathbb{G}^*)^c} \|v_{\mathcal{G}_i}\|^q + \alpha \sum_{i \in \mathbb{G}^*} \|(x^* + v)_{\mathcal{G}_i}\|^q. \end{aligned}$$

The second step is due to $x^*_{\tilde{S}^c} = 0$. When $v_{\tilde{S}^c} \neq 0$, we can regard $(d - A_{\tilde{S}}(x^* + v)_{\tilde{S}}), \alpha \sum_{i \in \mathbb{G}^*} \|(x^* + v)_{\mathcal{G}_i}\|^q$ in last step above as constant. Then, applying Proposition 5.1 with $N = \#\tilde{S}^c$ gives

$$\begin{aligned} F(x^* + v) &> \|A_{\tilde{S}^c} \cdot 0 - (d - A_{\tilde{S}}(x^* + v)_{\tilde{S}})\|^2 + \alpha \sum_{i \in (\mathbb{G}^*)^c} \|0\|^q + \alpha \sum_{i \in \mathbb{G}^*} \|(x^* + v)_{\mathcal{G}_i}\|^q \\ &= \|A_{\tilde{S}}(\tilde{y}^* + v_{\tilde{S}}) - d\|^2 + \alpha \sum_{i \in \mathbb{G}^*} \|(x^* + v)_{\mathcal{G}_i}\|^q \\ &= \tilde{H}(\tilde{y}^* + v_{\tilde{S}}) \\ &\geq \tilde{H}(\tilde{y}^*) \\ &= F(x^*). \end{aligned}$$

When $v_{\tilde{S}^c} = 0$, we can deduce

$$F(x^* + v) = \tilde{H}(\tilde{y}^* + v_{\tilde{S}}) > \tilde{H}(\tilde{y}^*) = F(x^*),$$

from that $\tilde{y}^*$ is a strict local minimizer of $\tilde{H}(y)$.

This shows that x^* is a strict local minimizer of $F(x)$. □

Remark 5.4. According to the theorem above and Theorem 3.1, we have that if $F(x)$ has a local minimum at x^*, then

$$\forall\, i \in \text{gsupp}(x^*), \quad \|x^*_{\mathcal{G}_i}\| \geq \left(\frac{\alpha q(1-q)}{2\|A\|^2}\right)^{\frac{1}{2-q}}. \tag{44}$$

Conversely, if x^* is a stationary point of $F(x)$ with the columns of $A_{\text{supp}(x^*)}$ linearly independent and

$$\forall\, i \in \text{gsupp}(x^*), \quad \|x^*_{\mathcal{G}_i}\| > \left(\frac{\alpha q(1-q)}{2\varrho(A)}\right)^{\frac{1}{2-q}}, \tag{45}$$

then x^* is a strict local minimizer of $F(x)$.

Denote $\bar{\iota} > 0$ as the bound in (45), and $\underline{\iota} > 0$ as the bound in (44). Clearly, for a stationary point, if its minimal ℓ_2 norm of nonzero groups is greater than $\bar{\iota}$, it is a local minimizer; if its minimal ℓ_2 norm of nonzero groups is less than $\underline{\iota}$, it is definitely not a local minimizer; if the norm is between $\underline{\iota}$ and $\bar{\iota}$, we are not sure whether it is a local minimizer.

6. Algorithms

In this section, we present two IRLS based algorithms to minimize $F(x)$. Generally, in each iteration of IRLS method, one needs to solve

$$x^{k+1} = \arg\min_{x \in \mathbf{R}^N} \|Ax - d\|^2 + \frac{\alpha}{2}\sum_{j=1}^{N}(w_j^k x_j)^2,$$

where w^k is the weight vector dependent on x^k. The first algorithm in [27] is a generalization of classical IRLS with smoothing algorithm for the $\ell_{2,q}$ regularized minimization. It generates a sequence which has convergent subsequences, and the cluster points are critical points of smoothed $F(x)$. The second is our novel IRLS with thresholding algorithm. It not only has global convergence with the limit being a critical point of $F(x)$, but also greatly improves the computation efficiency. From the comparisons of the two algorithms, we show the superiority of thresholding technique.

6.1. IRLS with smoothing algorithm

The function $F(x)$ is non-Lipschitz since the ℓ_q quasi-norm is non-Lipschitz around 0. Thus, the classical IRLS algorithm starts with the following smoothed objective function:

$$F_\epsilon(x) := \|Ax - d\|^2 + \alpha\sum_{i=1}^{n}(\|x_{\mathcal{G}_i}\|^2 + \epsilon)^{\frac{q}{2}}, \tag{46}$$

where $\epsilon > 0$ is the smoothing parameter. If x^* is a local minimizer of $F_\epsilon(x)$, then it satisfies the first-order necessary condition as follows:

$$2A^T(Ax^* - d) + \alpha q\left[\,(\|x^*_{\mathcal{G}_i}\|^2 + \epsilon)^{\frac{q}{2}-1}x^*_{\mathcal{G}_i}\,\right]_{1\le i\le n} = 0.$$

We first provide a bound analysis of $F_\epsilon(x)$.

Proposition 6.1. *There exists $\epsilon_0 > 0$ such that when $0 < \epsilon < \epsilon_0$, any local minimizer x^* of $F_\epsilon(x)$ satisfies*

$$\forall\, i \in \mathbb{I}, \qquad \textit{either} \quad \|x^*_{\mathcal{G}_i}\| \le \theta_1 \quad \textit{or} \quad \|x^*_{\mathcal{G}_i}\| \ge \theta_2, \tag{47}$$

where $\theta_1 < \theta_2$ are positive constants.

Proof. Note that $F_\epsilon(x)$ is smooth everywhere. Following the process in (16), the second-order condition indicates

$$\nabla^2 F_\epsilon(x^*) = 2A^TA + \alpha \begin{pmatrix} M^1 + D^1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & M^n + D^n \end{pmatrix} \succeq 0,$$

where $M^i, D^i \in \mathbf{R}^{N_i \times N_i}$,

$$M^i = q(q-2)(\|x^*_{\mathcal{G}_i}\|^2 + \epsilon)^{\frac{q}{2}-2}\, x^*_{\mathcal{G}_i}(x^*_{\mathcal{G}_i})^T, \qquad D^i = q(\|x^*_{\mathcal{G}_i}\|^2 + \epsilon)^{\frac{q}{2}-1} I.$$

Take $i \in \mathbb{I}$. We then wish to find an upper or lower bound for nonzero $x^*_{\mathcal{G}_i}$. Define $z \in \mathbf{R}^N$ as a unite vector whose group support is $\{i\}$:

$$z_{\mathcal{G}_l} = \begin{cases} \dfrac{x^*_{\mathcal{G}_i}}{\|x^*_{\mathcal{G}_i}\|}, \text{if } \ l = i, \\ 0, \qquad \text{otherwise.} \end{cases}$$

Therefore,

$$\begin{aligned} 0 &\le z^T \nabla^2 F_\epsilon(x^*) z \\ &= 2\|Az\|^2 + \alpha q(q-2)(\|x^*_{\mathcal{G}_i}\|^2 + \epsilon)^{\frac{q}{2}-2}((z_{\mathcal{G}_i})^T x^*_{\mathcal{G}_i})^2 + \alpha q(\|x^*_{\mathcal{G}_i}\|^2 + \epsilon)^{\frac{q}{2}-1}(z_{\mathcal{G}_i})^T z_{\mathcal{G}_i} \\ &\le 2\|A\|_2^2 + \alpha q(\|x^*_{\mathcal{G}_i}\|^2 + \epsilon)^{\frac{q}{2}-2}\left[(q-1)\|x^*_{\mathcal{G}_i}\|^2 + \epsilon\right]. \end{aligned}$$

We introduce a function $\phi : [0, +\infty) \to R$ as follows

$$\phi(t) = 2\|A\|_2^2 + \alpha q\,(t + \epsilon)^{\frac{q}{2}-2}\left[(q-1)t + \epsilon\right].$$

Then, we solve $\phi(t) \ge 0$. One can see that $\phi(0) > 0$, and the derivative of $\phi(t)$ is

$$\phi'(t) = \alpha q(1 - q/2)(t + \epsilon)^{\frac{q}{2}-3}[(1-q)t - 3\epsilon].$$

Thus, $\phi(t)$ is decreasing in $[0, 3\epsilon/(1-q))$ and increasing in $(3\epsilon/(1-q), +\infty)$. If $\phi(3\epsilon/(1-q)) < 0$, equivalently, $\epsilon < \epsilon_0$ where $\epsilon_0 := \left(\frac{\alpha q}{\|A\|_2^2}\right)^{2/(2-q)}\left(\frac{4-q}{1-q}\right)^{(q-4)/(2-q)}$, there exist θ_1, θ_2 such that $0 < \theta_1^2 < 3\epsilon/(1-q) < \theta_2^2$ and $\{t \ge 0 \ : \ \phi(t) \ge 0\} = [0, \theta_1^2] \cup [\theta_2^2, +\infty)$. That is, $\|x^*_{\mathcal{G}_i}\|^2 \in (0, \theta_1^2] \cup [\theta_2^2, +\infty)$, indicating the result. □

From the above proposition, we can see that there is no lower bound for local minimizers of the smoothed model. That is, it may allow some small values in its solution, making the solution approximately sparse but not exactly sparse. This result is not surprising, since the $\ell_{2,q}$ "norm" is smoothed and there is less penalty for values close to 0. The IRLS with smoothing algorithm for $\ell_{2,q}$ regularized minimization is summarized in Algorithm 1. By using ϵ, large values in the weight vector w_k are avoided (see [27] for more details).

Algorithm 1 IRLS with smoothing (IRLS-sm) [27].

Input: x^0 such that $Ax^0 = b, \epsilon_0 = 1, \beta \in (0,1)$ and estimated sparsity level s.
for $k = 0, 1, 2, \ldots$ **do**
Set the weight vector $w^k \in \mathbf{R}^N$:

$$\forall\, j \in \mathbb{J}, \qquad w_j^k = q(\|x_{\mathcal{G}_i}^k\|^2 + \epsilon^k)^{\frac{q}{2}-1}, \text{ if } j \in \mathcal{G}_i.$$

Solve the following linear system for x^{k+1}:

$$2A^T(Ax^{k+1} - d) + \alpha\, \mathrm{diag}(w^k)x^{k+1} = 0. \tag{48}$$

Update $\epsilon^{k+1} = \min(\epsilon^k, \beta \cdot \mathrm{r}(x^{k+1})_{s+1})$ where $\mathrm{r}(x^{k+1})$ is the rearrangement of the group norms of x^{k+1} in decreasing order.
end for

6.2. IRLS with thresholding algorithm

Motivated by the sparsity of local minimizers of $F(x)$ as shown in Theorem 3.1, we can adopt a threshold on nonzero groups of x^k, making x^k sparse. Meanwhile, we keep the zero groups of x^k in the following iterations. That is, at the $(k+1)$th iteration, we solve

$$\begin{aligned} x^{k+1} = \arg \min_{x \in \mathbf{R}^N} \quad & \|Ax - d\|^2 + \frac{\alpha}{2} \sum_{i \in \mathbb{G}^k} \sum_{j \in \mathcal{G}_i} (w_j^k x_j)^2, \\ \text{s.t.} \quad & x_{\mathcal{G}_i} = 0, \ \ \forall\, i \notin \mathbb{G}^k, \end{aligned} \tag{49}$$

where $\mathbb{G}^k = \mathrm{gsupp}(x^k)$. Our algorithm is summarized in Algorithm 2.

Algorithm 2 IRLS with thresholding (IRLS-th).

Input: threshold $\tau > 0$, $x^0 \in \mathbf{R}^N$ such that $\|x_{\mathcal{G}_i}^0\| \geq \tau, \forall\, i \in \mathbb{I}$;
for $k = 0, 1, 2, \ldots$ **do**
Set: $\mathbb{G}^k = \mathrm{gsupp}(x^k)$, $S^k = \cup_{i \in \mathbb{G}^k} \mathcal{G}_i$, $A^k = A_{S^k}, N^k = \# S^k$ and the weight vector $w^k = w_{_{S^k}}$ where $w \in \mathbf{R}^N$:

$$\forall\, j \in S^k, \qquad w_j = q\|x_{\mathcal{G}_i}^k\|^{q-2}, \text{ if } j \in \mathcal{G}_i.$$

Solve the following linear system for $\tilde{x}^{k+1} \in \mathbf{R}^{N^k}$:

$$2(A^k)^T(A^k\tilde{x}^{k+1} - d) + \alpha\, \mathrm{diag}(w^k)\tilde{x}^{k+1} = 0. \tag{50}$$

Set: $x_{S^k}^{k+1} = \tilde{x}^{k+1}, x_{(S^k)^c}^{k+1} = 0.$
Thresholding: $\forall\, i \in \mathbb{I}, x_{\mathcal{G}_i}^{k+1} = 0, \quad \text{if} \quad \|x_{\mathcal{G}_i}^{k+1}\| < \tau.$
end for

In Algorithm 2, since $\|x_{\mathcal{G}_i}^k\| \geq \tau$, we can avoid large values in w^k. Meanwhile, the constraints in (49) help to eliminate the number of variables, thus reducing the dimension of linear equations solved in each iteration. We mention that the support-shrinking strategy therein was also derived in [28,36,37], but for reweighted ℓ_1 variants with two-loop algorithmic structure for different signal and image reconstruction problems. Similar to these IRL1 variants, we have finite convergence property of the group support in our one-loop Algorithm 2.

Lemma 6.2. *In Algorithm 2, the group support set sequence* $\{\mathbb{G}^k\}$ *satisfies* $\forall\, k \geq 0, \mathbb{G}^{k+1} \subseteq \mathbb{G}^k$, *and thus converges. Specifically, there exists* $K \geq 0$ *and* $\mathbb{G}^* \subset \mathbb{I}$ *such that* $\forall\, k \geq K$,

$$G^k = \mathbb{G}^*. \tag{51}$$

6.2.1. Convergence and error bound analysis of Algorithm 2

In this part, we analyze the convergence of x^k generated by Algorithm 2, and give an error bound for its limit point. The convergence of x^k relies on the Kurdyka-Łojasiewicz (KL) property of $F(x)$. We establish our result by proving three conditions critical to the convergence under KL framework, i.e., that x^k is bounded, the function value sequence $F(x^k)$ has sufficient descent and the subdifferential sequence $\partial F(x^k)$ is bounded relatively. See Appendix for more details. Note that our algorithm does not have a proximal term.

Lemma 6.3. *For all $x, y \in \mathbf{R}^N$ with $x \neq 0$,*

$$\|x\|^q - \|y\|^q - q\|x\|^{q-2}(x-y)^T y \geq \frac{q}{2}\|x\|^{q-2}\|x-y\|^2. \tag{52}$$

Proof. It is obvious by modifying the proof of Lemma 2.3 in [31], where we replace the scalar product with vector inner product. □

We give some notations at first. Let $k \geq 0$. Note that $\tilde{x}^{k+1}$ in (50) belongs to $\mathbf{R}^{N^k}$. We denote $\overline{\tilde{x}^{k+1}} \in \mathbf{R}^N$ as an extension of $\tilde{x}^{k+1}$:

$$\overline{\tilde{x}^{k+1}}_{S^k} = \tilde{x}^{k+1}, \qquad \overline{\tilde{x}^{k+1}}_{(S^k)^c} = 0, \tag{53}$$

that is, its support $\text{supp}\left(\overline{\tilde{x}^{k+1}}\right) \subseteq S^k$. Also, we denote

$$r^{k+1} = \overline{\tilde{x}^{k+1}} - x^{k+1}, \tag{54}$$

so its support $\text{supp}(r^{k+1}) \subseteq S^k \backslash S^{k+1}$ and its group support $\text{gsupp}(r^{k+1}) \subseteq \mathbb{G}^k \backslash \mathbb{G}^{k+1}$. Then, we have

$$Ax^k = A^k x^k_{S^k}, \quad Ax^{k+1} = A\left(\overline{\tilde{x}^{k+1}} - r^{k+1}\right), \quad \text{and} \quad A\overline{\tilde{x}^{k+1}} = A^k \tilde{x}^{k+1}, \tag{55}$$

When $\mathbb{G}^{k+1} = \mathbb{G}^k$, one has $r^{k+1} = 0$ and $Ax^{k+1} = A^k \tilde{x}^{k+1}$.

Lemma 6.4. *(Boundedness, and sufficient descent of objective function value) Let $\{x^k\}$ be the sequence generated from Algorithm 2. The following statements hold:*

(i) *there exists $C > 0$ such that*

$$C \geq \|x^k_{\mathcal{G}_i}\| \geq \tau, \qquad \forall\, i \in \mathbb{G}^k, \quad k \geq 0; \tag{56}$$

(ii) *there exist $a > 0, \tilde{\tau} > 0$ such that*

$$F(x^k) - F(x^{k+1}) \geq a\|x^k - x^{k+1}\|^2, \tag{57}$$

holds for all $k \geq K$ where K is defined in Lemma 6.2; meanwhile, (57) holds for all $k \geq 0$ when the threshold $\tau \leq \tilde{\tau}$.

Proof. *(i).* Let $k \geq 0$. Note that $\mathbb{G}^{k+1} \subseteq \mathbb{G}^k$ according to Lemma 6.2.

When $\mathbb{G}^{k+1} = \mathbb{G}^k$, we have $r^{k+1} = 0$ and

$$F(x^k) - F(x^{k+1})$$

$$
\begin{aligned}
&=(\|Ax^k-d\|^2+\alpha\|x^k\|_{2,q}^q)-(\|Ax^{k+1}-d\|^2+\alpha\|x^{k+1}\|_{2,q}^q)\\
&=(\|A^kx^k_{S^k}-d\|^2-\|A^k\tilde{x}^{k+1}-d\|^2)+\alpha\sum_{i\in\mathbb{G}^k}(\|x^k_{\mathcal{G}_i}\|^q-\|x^{k+1}_{\mathcal{G}_i}\|^q)\\
&=\|A^kx^k_{S^k}-A^k\tilde{x}^{k+1}\|^2+2(A^k\tilde{x}^{k+1}-d)^T(A^kx^k_{S^k}-A^k\tilde{x}^{k+1})+\alpha\sum_{i\in\mathbb{G}^k}(\|x^k_{\mathcal{G}_i}\|^q-\|x^{k+1}_{\mathcal{G}_i}\|^q)\\
&\geq\left[2(A^k)^T(A^k\tilde{x}^{k+1}-d)\right]^T(x^k_{S^k}-\tilde{x}^{k+1})+\alpha\sum_{i\in\mathbb{G}^k}(\|x^k_{\mathcal{G}_i}\|^q-\|x^{k+1}_{\mathcal{G}_i}\|^q)\\
[\text{ Equ. (50) }]\quad&=[-\alpha\,\mathrm{diag}(w^k)\tilde{x}^{k+1}]^T(x^k_{S^k}-\tilde{x}^{k+1})+\alpha\sum_{i\in\mathbb{G}^k}(\|x^k_{\mathcal{G}_i}\|^q-\|x^{k+1}_{\mathcal{G}_i}\|^q)\\
&=\alpha\sum_{i\in\mathbb{G}^k}[\ \|x^k_{\mathcal{G}_i}\|^q-\|x^{k+1}_{\mathcal{G}_i}\|^q-q\|x^k_{\mathcal{G}_i}\|^{q-2}(x^k_{\mathcal{G}_i}-x^{k+1}_{\mathcal{G}_i})^Tx^{k+1}_{\mathcal{G}_i}\ ]\\
[\text{ Equ. (52) }]\quad&\geq\frac{\alpha q}{2}\sum_{i\in\mathbb{G}^k}\|x^k_{\mathcal{G}_i}\|^{q-2}\|x^{k+1}_{\mathcal{G}_i}-x^k_{\mathcal{G}_i}\|^2,
\end{aligned}
\tag{58}
$$

implying $F(x^k)-F(x^{k+1})\geq 0$ when $\mathbb{G}^{k+1}=\mathbb{G}^k$.

When $\mathbb{G}^{k+1}\subsetneq\mathbb{G}^k$, we have

$$
F(x^k)-F(x^{k+1})=\left[F(x^k)-F(\overline{\tilde{x}^{k+1}})\right]-\left[F(x^{k+1})-F(\overline{\tilde{x}^{k+1}})\right],
$$

where $\overline{\tilde{x}^{k+1}}$ is defined in (53). The next is to find a lower bound for $F(x^k)-F(\overline{\tilde{x}^{k+1}})$ and an upper bound for $F(x^{k+1}-F(\overline{\tilde{x}^{k+1}}))$. Similar to (58), one has

$$
\begin{aligned}
&F(x^k)-F(\overline{\tilde{x}^{k+1}})\\
\geq&\frac{\alpha q}{2}\sum_{i\in\mathbb{G}^k}\|x^k_{\mathcal{G}_i}\|^{q-2}\|\overline{\tilde{x}^{k+1}}_{\mathcal{G}_i}-x^k_{\mathcal{G}_i}\|^2\\
=&\sum_{i\in\mathbb{G}^{k+1}}\frac{\alpha q}{2}\|x^k_{\mathcal{G}_i}\|^{q-2}\|x^{k+1}_{\mathcal{G}_i}-x^k_{\mathcal{G}_i}\|^2+\sum_{i\in\mathbb{G}^k\backslash\mathbb{G}^{k+1}}\frac{\alpha q}{2}\|x^k_{\mathcal{G}_i}\|^{q-2}\|r^{k+1}_{\mathcal{G}_i}-x^k_{\mathcal{G}_i}\|^2.
\end{aligned}
\tag{59}
$$

Meanwhile, applying $r^{k+1}=\overline{\tilde{x}^{k+1}}-x^{k+1}$ yields

$$
\begin{aligned}
F(x^{k+1})-F(\overline{\tilde{x}^{k+1}})&=(\|Ax^{k+1}-d\|^2-\|A\overline{\tilde{x}^{k+1}}-d\|^2)+\alpha(\|x^{k+1}\|_{2,q}^q-\|\overline{\tilde{x}^{k+1}}\|_{2,q}^q)\\
&=\|Ar^{k+1}\|^2-2(A\overline{\tilde{x}^{k+1}}-d)^T(Ar^{k+1})-\alpha\sum_{\mathbb{G}^k\backslash\mathbb{G}^{k+1}}\|r^{k+1}_{\mathcal{G}_i}\|^q.
\end{aligned}
$$

It follows from $A\overline{\tilde{x}^{k+1}}=A^k\tilde{x}^{k+1}$ and $\mathrm{gsupp}\{r^{k+1}\}\subseteq\mathbb{G}^k\backslash\mathbb{G}^{k+1}$, that

$$
\|Ar^{k+1}\|^2\leq\sum_{i\in\mathbb{G}^k\backslash\mathbb{G}^{k+1}}\|A\|^2\|r^{k+1}_{\mathcal{G}_i}\|^2,
$$

and

$$
\begin{aligned}
-2(A\overline{\tilde{x}^{k+1}}-d)^T(Ar^{k+1})&=\sum_{i\in\mathbb{G}^k\backslash\mathbb{G}^{k+1}}-2\left[A^T_{\mathcal{G}_i}(A^k\tilde{x}^{k+1}-d)\right]^Tr^{k+1}_{\mathcal{G}_i}\\
[\text{ Equ. (50) }]\quad&=\sum_{i\in\mathbb{G}^k\backslash\mathbb{G}^{k+1}}\alpha q\|x^k_{\mathcal{G}_i}\|^{q-2}\|r^{k+1}_{\mathcal{G}_i}\|^2.
\end{aligned}
$$

It then follows that

$$F(x^{k+1}) - F(\overline{\tilde{x}^{k+1}}) \leq \sum_{i \in \mathbb{G}^k \backslash \mathbb{G}^{k+1}} \left[(\|A\|^2 + \alpha q \|x^k_{\mathcal{G}_i}\|^{q-2}\|)\|r^{k+1}_{\mathcal{G}_i}\|^2 - \alpha \|r^{k+1}_{\mathcal{G}_i}\|^q \right]. \tag{60}$$

Since $\|r^{k+1}_{\mathcal{G}_i}\| \leq \tau$ and $\|x^k_{\mathcal{G}_i}\| \geq \tau$ for $i \in \mathbb{G}^k$, we get

$$F(x^{k+1}) - F(\overline{\tilde{x}^{k+1}}) \leq \sum_{i \in \mathbb{G}^k \backslash \mathbb{G}^{k+1}} (\|A\|^2\tau^2 + \alpha q \tau^q) = \#(\mathbb{G}^k \backslash \mathbb{G}^{k+1}) \cdot (\|A\|^2\tau^2 + \alpha q \tau^q).$$

Recall (59) which indicates $F(x^k) - F(\overline{\tilde{x}^{k+1}}) \geq 0$. Therefore, one has

$$F(x^k) - F(x^{k+1}) \geq -\#(\mathbb{G}^k \backslash \mathbb{G}^{k+1}) \cdot (\|A\|^2\tau^2 + \alpha q \tau^q). \tag{61}$$

It follows from $F(x^k) - F(x^{k+1}) \geq 0$ when $\mathbb{G}^{k+1} = \mathbb{G}^k$ and (61) when $\mathbb{G}^{k+1} \subsetneq \mathbb{G}^k$ that

$$F(x^{k+1}) - F(x^0) \leq \sum_{l=0}^{k} \#(\mathbb{G}^l \backslash \mathbb{G}^{l+1}) \cdot (\|A\|^2\tau^2 + \alpha q \tau^q) \leq n(\|A\|^2\tau^2 + \alpha q \tau^q),$$

which means $F(x^{k+1}) \leq F(x^0) + n(\|A\|^2\tau^2 + \alpha q \tau^q)$, i.e., $F(x^{k+1})$ is bounded. Thus,

$$\|x^{k+1}\|^q_{2,q} \leq \frac{1}{\alpha} F(x^{k+1}) \leq \frac{1}{\alpha} F(x^0) + \frac{n}{\alpha}(\|A\|^2\tau^2 + \alpha q \tau^q).$$

Considering $\|x^0\|^q_{2,q} \leq \frac{1}{\alpha} F(x^0)$, there exists $C > 0$ dependent on x^0 such that $\tau \leq \|x^k_{\mathcal{G}_i}\| \leq C$, for any $i \in \mathbb{G}^k$. Thus statement (i) can be obtained.

(ii). By Lemma 6.2, if $k \geq K$, we have $\mathbb{G}^k = \mathbb{G}^{k+1} = \mathbb{G}^*$. Recall (58). We can see that when $k \geq K$ $F(x^k)$ is decreasing, and

$$F(x^k) - F(x^{k+1}) \geq \frac{\alpha q}{2} C^{q-2} \sum_{i \in \mathbb{G}^*} \|x^{k+1}_{\mathcal{G}_i} - x^k_{\mathcal{G}_i}\|^2 = \frac{\alpha q}{2} C^{q-2} \|x^k - x^{k+1}\|^2.$$

Let $a = \frac{\alpha q}{4} C^{q-2}$, then (57) is true for all $k \geq K$.

Next, we will prove (57) for $k \geq 0$ under the assumption $\tau \leq \tilde{\tau} := \left(\frac{\alpha - \alpha q}{\|A\|^2 + \tilde{a}}\right)^{1/(2-q)}$ where $\tilde{a}$ is an arbitrary positive constant.

When $\mathbb{G}^{k+1} \subsetneq \mathbb{G}^k$, we take an $i \in \mathbb{G}^k \backslash \mathbb{G}^{k+1}$. Now, since

$$\begin{aligned} &(\|A\|^2 + \alpha q \|x^k_{\mathcal{G}_i}\|^{q-2} + \tilde{a})\|r^{k+1}_{\mathcal{G}_i}\|^{2-q} - \alpha \\ \leq &(\|A\|^2 + \tilde{a} + \alpha q \tau^{q-2})\tau^{2-q} - \alpha \\ = &(\|A\|^2 + \tilde{a})\tau^{2-q} + \alpha q - \alpha \\ \leq &0, \end{aligned}$$

one can see that

$$(\|A\|^2 + \alpha q \|x^k_{\mathcal{G}_i}\|^{q-2})\|r^{k+1}_{\mathcal{G}_i}\|^2 - \alpha \|r^{k+1}_{\mathcal{G}_i}\|^q \leq -\tilde{a}\|r^{k+1}_{\mathcal{G}_i}\|^2.$$

Substituting it into (60) gives

$$F(x^{k+1}) - F(\overline{\tilde{x}^{k+1}}) \leq -\tilde{a} \sum_{i \in \mathbb{G}^k \backslash \mathbb{G}^{k+1}} \|r^{k+1}_{\mathcal{G}_i}\|^2. \tag{62}$$

Note that $x^{k+1}_{\mathcal{G}_i} = 0$ and $2\|r^{k+1}_{\mathcal{G}_i} - x^k_{\mathcal{G}_i}\|^2 + 2\|r^{k+1}_{\mathcal{G}_i}\|^2 \ge \|x^k_{\mathcal{G}_i}\|^2$ for all $i \in \mathbb{G}^k \backslash \mathbb{G}^{k+1}$. Combining (62) with (59), we have

$$\begin{aligned}
&F(x^k) - F(x^{k+1}) \\
&\ge \sum_{i\in\mathbb{G}^{k+1}} \frac{\alpha q}{2}\|x^k_{\mathcal{G}_i}\|^{q-2}\|x^{k+1}_{\mathcal{G}_i} - x^k_{\mathcal{G}_i}\|^2 + \sum_{i\in\mathbb{G}^k\backslash\mathbb{G}^{k+1}} \left[\frac{\alpha q}{2}\|x^k_{\mathcal{G}_i}\|^{q-2}\|r^{k+1}_{\mathcal{G}_i} - x^k_{\mathcal{G}_i}\|^2 + \tilde{a}\|r^{k+1}_{\mathcal{G}_i}\|^2\right] \\
&\ge \sum_{i\in\mathbb{G}^{k+1}} \frac{\alpha q}{2}\|x^k_{\mathcal{G}_i}\|^{q-2}\|x^{k+1}_{\mathcal{G}_i} - x^k_{\mathcal{G}_i}\|^2 + \sum_{i\in\mathbb{G}^k\backslash\mathbb{G}^{k+1}} \min\left\{\frac{\alpha q}{4}\|x^k_{\mathcal{G}_i}\|^{q-2}, \frac{\tilde{a}}{2}\right\}\|x^k_{\mathcal{G}_i} - 0\|^2 \\
&\ge \sum_{i\in\mathbb{G}^k} \min\left\{\frac{\alpha q}{4}C^{q-2}, \frac{\tilde{a}}{2}\right\}\|x^{k+1}_{\mathcal{G}_i} - x^k_{\mathcal{G}_i}\|^2. \qquad (63)
\end{aligned}$$

Therefore, combining (58) when $\mathbb{G}^{k+1} = \mathbb{G}^k$ and (63) when $\mathbb{G}^{k+1} \subsetneq \mathbb{G}^k$, one can see that if $\tau \le \tilde{\tau}$, $F(x^k)$ is decreasing for all $k \ge 0$. Let $a = \min\left\{\frac{\alpha q}{4}C^{q-2}, \frac{\tilde{a}}{2}\right\}$. Then, we have

$$F(x^k) - F(x^{k+1}) \ge a\|x^k - x^{k+1}\|^2, \quad \forall\, k \ge 0.$$

The proof of statement (ii) is completed. □

Lemma 6.5. *(Relatively bounded subdifferential) For each $k \ge 0$, there exists $u^{k+1} \in \partial F(x^{k+1})$ such that*

$$\|u^{k+1}\| \le b\|x^k - x^{k+1}\|, \qquad (64)$$

where b is a positive constant.

Proof. Let $k \ge 0$. According to (13), we denote

$$\begin{cases}\forall\, i \in \mathbb{G}^{k+1}, & u^{k+1}_{\mathcal{G}_i} = 2(A_{\mathcal{G}_i})^T(Ax^{k+1} - d) + \alpha q\|x^{k+1}_{\mathcal{G}_i}\|^{q-2}x^{k+1}_{\mathcal{G}_i}, \\ \forall\, i \notin \mathbb{G}^{k+1}, & u^{k+1}_{\mathcal{G}_i} = 0.\end{cases}$$

For any $i \in \mathbb{G}^{k+1}$, since $2(A_{\mathcal{G}_i})^T(A^k\tilde{x}^{k+1} - d) + \alpha q\|x^k_{\mathcal{G}_i}\|^{q-2}x^{k+1}_{\mathcal{G}_i} = 0$, one has

$$\begin{aligned}
u^{k+1}_{\mathcal{G}_i} &= \alpha q\|x^{k+1}_{\mathcal{G}_i}\|^{q-2}x^{k+1}_{\mathcal{G}_i} + 2(A_{\mathcal{G}_i})^T(A^k\tilde{x}^{k+1} - d) - 2(A_{\mathcal{G}_i})^T Ar^{k+1}, \\
&= \alpha q\|x^{k+1}_{\mathcal{G}_i}\|^{q-2}x^{k+1}_{\mathcal{G}_i} - \alpha q\|x^k_{\mathcal{G}_i}\|^{q-2}x^{k+1}_{\mathcal{G}_i} - 2(A_{\mathcal{G}_i})^T Ar^{k+1}.
\end{aligned}$$

Then, we denote $\tilde{u} \in \mathbf{R}^N$ as

$$\begin{cases}\forall\, i \in \mathbb{G}^{k+1}, & \tilde{u}_{\mathcal{G}_i} := \alpha q\|x^{k+1}_{\mathcal{G}_i}\|^{q-2}x^{k+1}_{\mathcal{G}_i} - \alpha q\|x^k_{\mathcal{G}_i}\|^{q-2}x^{k+1}_{\mathcal{G}_i} = \alpha q\left(\|x^{k+1}_{\mathcal{G}_i}\|^{q-2} - \|x^k_{\mathcal{G}_i}\|^{q-2}\right)x^{k+1}_{\mathcal{G}_i}, \\ \forall\, i \notin \mathbb{G}^{k+1}, & \tilde{u}_{\mathcal{G}_i} = 0.\end{cases}$$

It follows that

$$\|u^{k+1}\| = \|u^{k+1}_{S^{k+1}}\| = \|\tilde{u}_{S^{k+1}} - 2(A_{S^{k+1}})^T Ar^{k+1}\| \le \|\tilde{u}_{S^{k+1}}\| + \|2(A_{S^{k+1}})^T Ar^{k+1}\|.$$

Take $i \in \mathbb{G}^{k+1}$. Since $\tau \le \|x^k_{\mathcal{G}_i}\|, \|x^{k+1}_{\mathcal{G}_i}\| \le C$ and $\phi(t) = t^{q-2}$ is Lipschitz-continuous on $[\tau, C]$ with Lipschitz constant l_τ, we obtain

$$\|\tilde{u}_{\mathcal{G}_i}\| \le \alpha qCl_\tau \left|\, \|x^{k+1}_{\mathcal{G}_i}\| - \|x^k_{\mathcal{G}_i}\| \,\right| \le \alpha qCl_\tau\|x^{k+1}_{\mathcal{G}_i} - x^k_{\mathcal{G}_i}\| = \tilde{b}\|x^{k+1}_{\mathcal{G}_i} - x^k_{\mathcal{G}_i}\|,$$

followed by

$$\|\tilde{u}_{S^{k+1}}\| \leq \tilde{b}\|x^{k+1}_{S^{k+1}} - x^k_{S^{k+1}}\|,$$

where $\tilde{b} := \alpha q C l_\tau$.

Note that $\mathrm{gsupp}(r^{k+1}) \subseteq \mathbb{G}^k \backslash \mathbb{G}^{k+1}$. Since for all $i \in \mathbb{G}^k \backslash \mathbb{G}^{k+1}$, $\|r^{k+1}_{\mathcal{G}_i}\| \leq \tau, x^{k+1}_{\mathcal{G}_i} = 0$ and $\|x^k_{\mathcal{G}_i}\| \geq \tau$, we have

$$\|r^{k+1}\| \leq \|x^k_{S^k\backslash S^{k+1}}\| = \|x^{k+1}_{S^k\backslash S^{k+1}} - x^k_{S^k\backslash S^{k+1}}\|.$$

Therefore,

$$\begin{aligned}
\|u^{k+1}\| &\leq \tilde{b}\|x^{k+1}_{S^{k+1}} - x^k_{S^{k+1}}\| + 2\|A\|^2\|r^{k+1}\| \\
&\leq \max\{\tilde{b}, 2\|A\|^2\}\left(\|x^{k+1}_{S^{k+1}} - x^k_{S^{k+1}}\| + \|x^k_{S^k\backslash S^{k+1}}\|\right) \\
&\leq b\|x^{k+1} - x^k\|,
\end{aligned}$$

where $b = \sqrt{2}\max\{\tilde{b}, 2\|A\|^2\}$. □

Now, we are ready to prove the global convergence of our algorithm. Meanwhile, we also will provide an error bound for the limit point of sequence generated by our algorithm.

Theorem 6.6. *(Global convergence and error bound) Let $\{x^k\}$ be the sequence generated by Algorithm 2. The following statements hold:*

(i) *x^k converges to a stationary point x^* of $F(x)$.*
(ii) *Suppose x^o is t group sparse, i.e., $\|x^o\|_{2,0} \leq t$. Assume that $Ax^0 = d$ and A satisfies* GRIP *of order $2t$ with $\delta'_{2t} < 1$. When the threshold $\tau \leq \tilde{\tau}$ with $\tilde{\tau}$ defined in Lemma 6.4, we have*

$$\|x^* - x^o\| \leq C_1\sqrt{\alpha} + C_2\sqrt{\sigma_{t,2}(x^*)}, \tag{65}$$

where C_1, C_2 are positive constants depending on x^0, A, t, δ'_{2t}.

Proof. *(i).* In Lemma 6.4, since $\{x^k\}$ is bounded, there exists a subsequence $\{x^{k_l}\}$ converging to a point $x^* \in \mathbf{R}^N$. Recall (57) and (64). By Theorem 2.9 in [38] (see Lemma 9.4 for details), we obtain that x^k converges to a stationary point x^* of $F(x)$.

(ii). The following argument is a group extension of the one in Theorem 2.2 in [31]. When $\tau \leq \tilde{\tau}$, $F(x^k)$ is decreasing and

$$\|Ax^* - d\| \leq \sqrt{F(x^*)} \leq \sqrt{F(x^0)} = \sqrt{\alpha\|x^0\|^q_{2,q}}.$$

Let $\mathbb{G}^o$ be the index set of nonzero group of x^o with $T^o = \cup_{i\in\mathbb{G}^o}\mathcal{G}_i$. Let $\mathbb{G}$ be the index set of the first t largest groups of x^* with $T = \cup_{i\in\mathbb{G}}\mathcal{G}_i$. Note that $\|x^o\|_{2,0} \leq t$ and $x^* - x^o = \overline{(x^* - x^o)_{T\cup T^o}} + \overline{(x^* - x^o)_{(T\cup T^o)^c}}$. Then, we have

$$\begin{aligned}
\|x^* - x^o\| &\leq \|(x^* - x^o)_{T\cup T^o}\| + \|(x^* - x^o)_{(T\cup T^o)^c}\| \\
&\leq \frac{1}{\sqrt{1-\delta'_{2t}}}\|A\overline{(x^* - x^o)_{T\cup T^o}}\| + \|(x^* - x^o)_{(T\cup T^o)^c}\|
\end{aligned}$$

$$\begin{aligned}
&\le \frac{1}{\sqrt{1-\delta'_{2t}}}\|Ax^*-d\| + \left(\frac{\|A\|}{\sqrt{1-\delta'_{2t}}}+1\right)\|(x^*-x^o)_{(T\cup T^o)^c}\| \\
&= \frac{1}{\sqrt{1-\delta'_{2t}}}\|Ax^*-d\| + \left(\frac{\|A\|}{\sqrt{1-\delta'_{2t}}}+1\right)\|x^*_{(T\cup T^o)^c}\| \\
&\le \frac{1}{\sqrt{1-\delta'_{2t}}}\sqrt{\alpha\|x^0\|_{2,q}^q} + \left(\frac{\|A\|}{\sqrt{1-\delta'_{2t}}}+1\right)\sqrt{\sigma_{t,2}(x^*)},
\end{aligned}$$

which proves (65). □

We can see that the new IRLS with thresholding algorithm globally converges to a stationary point of $F(x)$; by contrast, IRLS with smoothing algorithm has only local convergence (subsequence convergence). Meanwhile, the limit is away from the true solutions by a factor of $\sqrt{\alpha}$ plus the tail $\sigma_{t,2}(x^*)$ of x^*. If both α and the tail are small, then x^* is close to the true sparse signal. In addition, if the threshold τ is equal or larger than the bound proven in Theorem 5.3, the limit of x^k is also a local minimizer of $F(x)$.

7. Numerical experiments

Here, we present some numerical results of our IRLS with thresholding (IRLS-th) algorithm, as well as comparisons with some state-of-the-art algorithms. All of the tests were performed using Windows 7 64-bit and Matlab 2018A, on a HP workstation Z840 with Intel Xeon CPU E5-2667.

Now, we give a general experiment setting. Unless otherwise noted, we use this setting in the whole section. Firstly, we set $M = 2^8, N = 2^{10}, n = 2^7$, and $N_i = 8$ for any $i \in \mathbb{I}$. That is, the original signal $x^o \in \mathbb{R}^N$ is split equally into 128 groups, with indices of its non-zero groups chosen randomly. The sparsity level is t/n where t is the number of nonzero groups in x^o, and is not fixed. The non-zero values in x^o are i.i.d. generated from standard Gaussian distribution. We also randomly generate an i.i.d. Gaussian matrix $A \in \mathbf{R}^{M\times N}$, and normalize it to satisfy $AA^T = I$. Then, the noisy simulated observation is

$$d = Ax^o + 0.001 \cdot \mathrm{randn}(M, 1).$$

For our IRLS-th algorithm, the regularization parameter $\alpha = 10^{-3}||A^T d||_\infty$, the threshold $\tau = 0.01$, exponent $q = 0.5$, initial vector $x^0 = \mathbf{1}$.

For all of the algorithms, the stopping criterion is

$$\frac{||x^k - x^{k-1}||}{||x^{k-1}||} < 10^{-4}.$$

The recovery result x^* is regarded as successful if

$$\frac{||x^* - x^o||}{||x^o||} < 0.01.$$

The success rate is the ratio of success over 100 trials.

7.1. Convergence of IRLS-th algorithm

In this subsection, we will show the convergence of IRLS-th algorithm. We set the sparsity level to 10%, i.e., we have 12 nonzero groups; see Fig. 3.

It can be seen from Fig. 3(b) that the group support size of x^k is decreasing and converges. As demonstrated in Fig. 3(c), the value of the objective function is monotonically decreasing and converges.

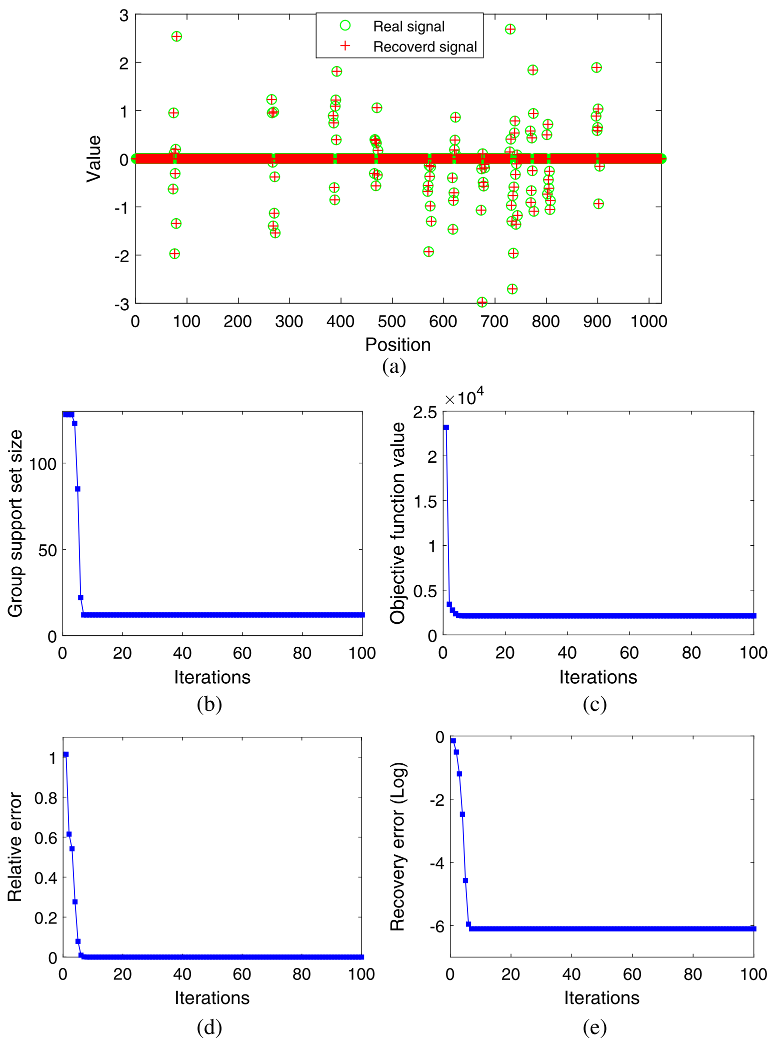

Fig. 3. Convergence of proposed IRLS-th algorithm for a group sparse recovery problem with sparsity level 10%.

By Fig. 3(d), the relative error is decreasing and converges. As shown in Fig. 3(e), the recovery error $Log(\frac{||x^k-x^o||_2}{||x^o||_2})$ at each iteration is decreasing, meaning that the approximation becomes more and more accurate and perfectly recover x^o in the end. This experiment clearly justifies the global convergence property of IRLS-th algorithm shown in Theorem 6.6.

7.2. Choice of τ

In this subsection, we discuss the choice of τ. Recall the lower bound theory in Theorem 3.1. The established lower bound is supposed to provide a reference value for τ. However, this bound is independent of observation d and local minimizers, and it might be too rough since we used a lot of inequalities to estimate it. Thus, for a specific problem, we may use a larger threshold to speed up the computation, while preserving the recovery accuracy.

Under our experimental settings, the theoretical lower bound is 0.0027. We therefore test our algorithm with different τ among $\{10^{-6}, 10^{-3}, 0.0027, 10^{-2}, 10^{-1}, 1\}$ in terms of the sparsity level. The success rates and average running time are plotted in Fig. 4. As can be seen, the success rates with different τ are equal when $\tau < 0.01$, but the average running time (only successful recovery is considered) is decreasing as τ increases. Therefore, in the following experiments, we use $\tau = 0.01$.

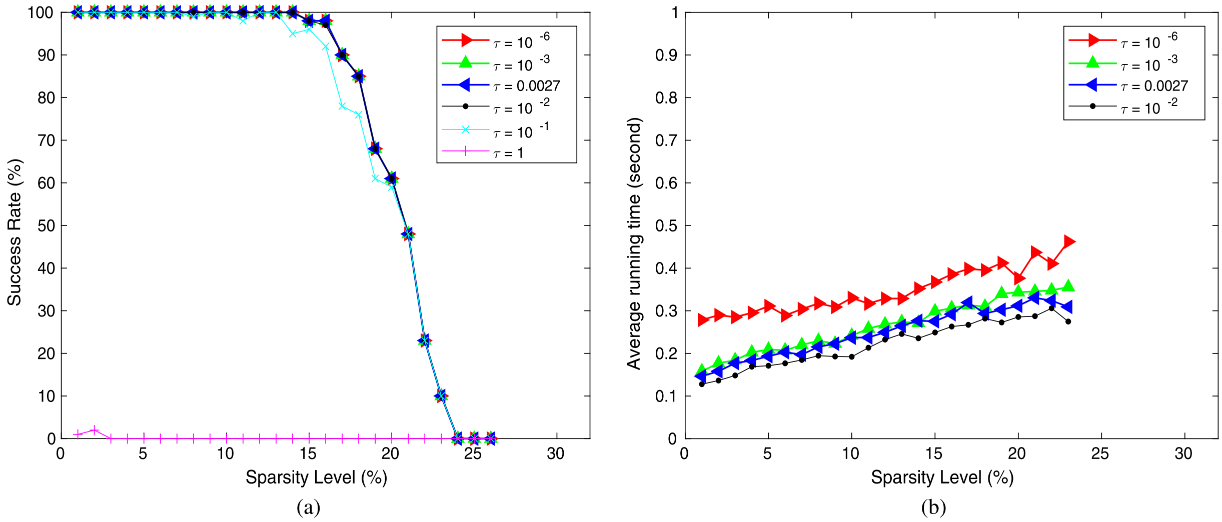

Fig. 4. The success rates and average running time of IRLS-th algorithm with different threshold τ.

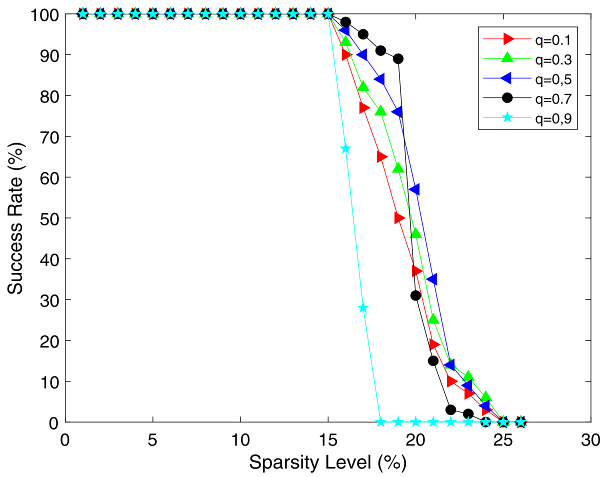

Fig. 5. The success rates of IRLS-th algorithm with different exponent q.

7.3. Choice of q

In this subsection, we will discuss the performances of our algorithm with different q among $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ in terms of the sparsity level. For each sparsity level, we generate 100 independent x^o. For each x^o, we run the IRLS-th algorithm with different q. The success rates of the results are shown in Fig. 5. We can see that when $q = 0.9$ with the objective function approximately convex, our algorithm performs worst. Similar to [15,31], when $q = 0.5$, it performs best. Therefore, in the following experiments, we choose $q = 0.5$.

7.4. Adaptability to group size

In this subsection, we test the sensitivity of our IRLS-th algorithm on the group size. With the group number $n = 128$, we change the group size N_i to be $2^2, 2^3, 2^4$ and 2^5. The corresponding (M, N) are $(2^7, 2^9)$, $(2^8, 2^{10})$, $(2^9, 2^{11})$, and $(2^{10}, 2^{12})$. For each pair of (M, N), we generate 100 independent x^o to test the success rate. By Fig. 6, we can see that when the group size becomes larger, the success rate increases. This is reasonable because larger groups contain more information for recovery.

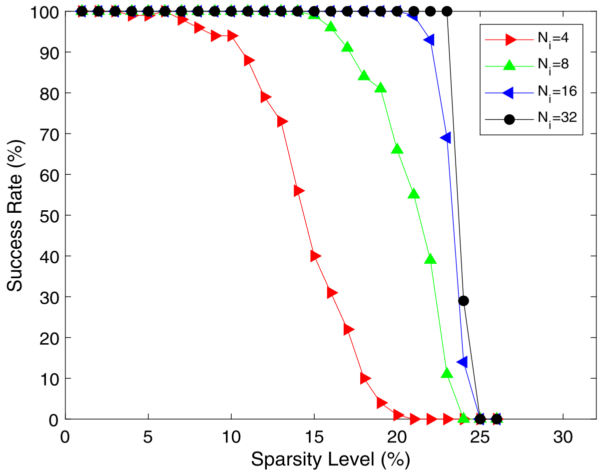

Fig. 6. The success rates of IRLS-th algorithm for recovery problems with different group size N_i.

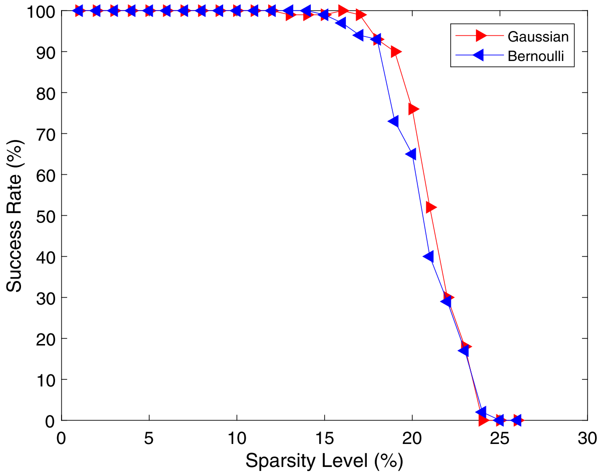

Fig. 7. The success rates of IRLS-th algorithm with different measurement matrix A.

7.5. Adaptability to different types of A

In this subsection, we test our algorithm with different measurement matrices A, i.e., random Gaussian matrices and random Bernoulli matrices. Also, A is normalized by $AA^T = I$ for the random Bernoulli matrices. The results are shown in Fig. 7. We can find that the IRLS-th algorithm could achieve almost the same success rate with different types of A.

7.6. Adaptability to different kinds of original signal x^o

In this subsection, we generate x^o from different kinds of distribution, i.e., the standard Gaussian distribution, uniform distribution and Bernoulli distribution (0.5). For each distribution, we generate 100 original signal x^o. The results are in Fig. 8. We can see that our algorithm would fit different kinds of signal with comparable success rates.

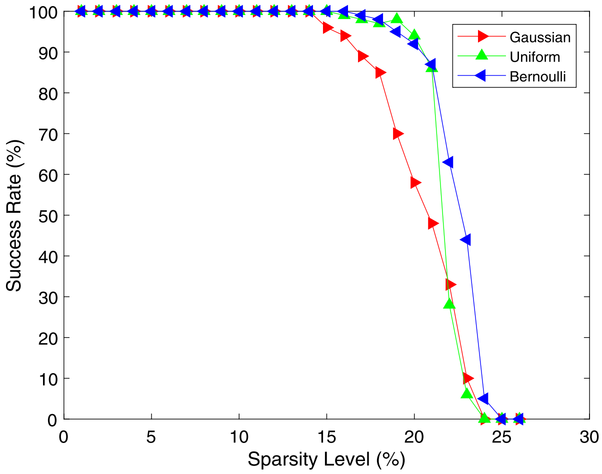

Fig. 8. The success rates of IRLS-th algorithm with different true signal x^o.

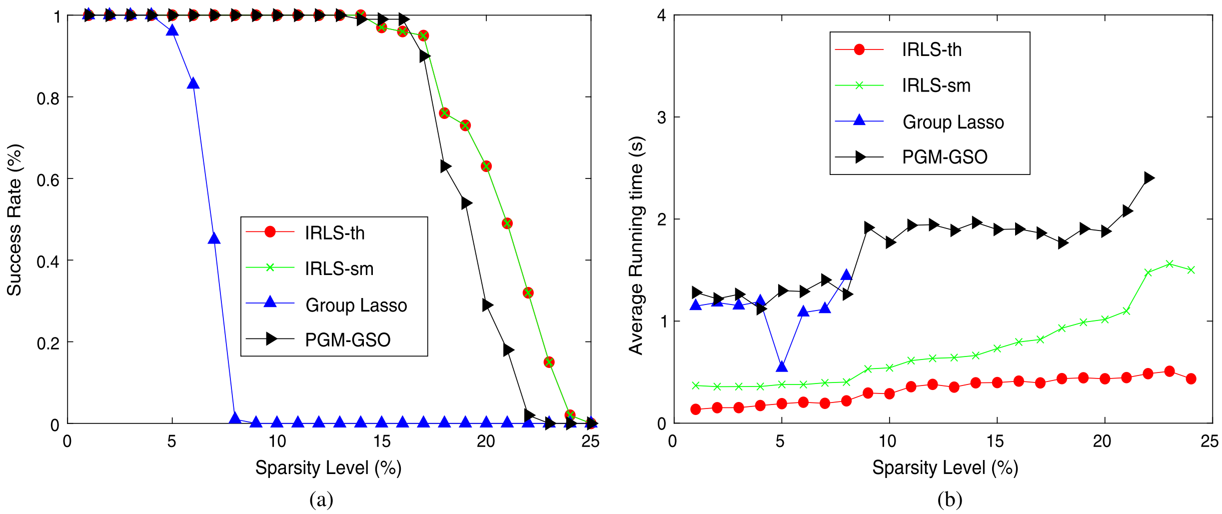

Fig. 9. Comparison of success rates and average running time between IRLS-th algorithm and state-of-the-art group sparse optimization algorithms.

7.7. Comparisons to some state-of-the-art algorithms

In this subsection, we compare our IRLS-th algorithm with several state-of-the-art group sparse optimization algorithms, i.e., IRLS with smoothing (IRLS-sm) algorithm [27], Group Lasso algorithm [4], and PGM-GSO algorithm [15]. The comparisons include the success rates and the average running time (only successful recovery is considered) in terms of sparsity level.

The results are shown in Fig. 9. Overall speaking, both IRLS algorithms perform better than PGM-GSO algorithm and Group Lasso algorithm. In Fig. 9(a), the success rate curves of IRLS-th algorithm and IRLS-sm algorithm coincide with each other, which means our IRLS-th algorithm perfectly keeps the excellent recovery ability of the classical IRLS-sm algorithm. From Fig. 9(b), we can see that our IRLS-th algorithm shows significant superiority in running time. It is at least 50% faster than IRLS-sm algorithm. When the sparsity level increases, the advantage is more obvious.

8. Conclusion

In this article, we have studied the $\ell_{2,q}$ regularized minimization model, including lower bound theory, recovery bound, algorithms and so on. Our work not only shows theoretical merits of the nonconvex group sparse optimization model, but also provides a new way to improve IRLS method in terms of global convergence and numerical performance.

Our study so far is only concerned with the group sparse signal recovery without overlapping group structure. While in many real applications, such as the gene expression data in bioinformatics, elements in different groups could potentially be overlapped. Besides, instead of the squared ℓ_2 fidelity term considered here, the fidelity term would be in other forms in applications with different types of measurement noise. All these extensions will be part of our future research.

9. Appendix

9.1. Subdifferential

The following is the definition of subdifferential.

Definition 9.1. (Subdifferential, [39]) Let $\phi : \mathbf{R}^N \to (-\infty, +\infty]$ be a proper and lower semicontinuous function. The domain of ϕ is defined as $\text{dom}\phi = \{x \in \mathbf{R}^N : \phi(x) < +\infty\}$. For a point $x \in \text{dom}\phi$,

1. the regular subdifferential of ϕ at x is defined as

$$\widehat{\partial}\phi(x) = \left\{u \in \mathbf{R}^N : \lim_{y \neq x} \inf_{y \to x} \frac{\phi(y) - \phi(x) - \langle u, y - x\rangle}{\|y - x\|} \geq 0\right\};$$

2. the subdifferential of ϕ at x is defined as

$$\partial\phi(x) = \{u \in \mathbf{R}^N : \exists\, x^k \to x, \phi(x^k) \to \phi(x) \text{ and } \widehat{\partial}\phi(x^k) \ni u^k \to u, \text{ as } k \to \infty\}.$$

9.2. Kurdyka-Łojasiewicz property

The foundational works on the Kurdyka-Łojasiewicz(KL) property are given by Łojasiewicz and Kurdyka [40,41]. Recently, there were great successes for the applications of KL property in optimization theory; see [42–45,38,46]. This subsection is a brief introduction of KL property, and most of the results are from [45].

For any subset $S \subseteq \mathbf{R}^N$ and any point $x \in \mathbf{R}^N$, the distance from x to D is defined by $\text{dist}(x, D) := \inf\{\|y - x\| : y \in D\}$. When $D = \emptyset$, we have that $\text{dist}(x, D) = +\infty$.

Definition 9.2. (KL property, [45])

1. The function $\phi : \mathbf{R}^N \to (-\infty, +\infty]$ is said to have the Kurdyka-Łojasiewicz property at $x^* \in \text{dom}\,\partial\phi := \{x \in \mathbf{R}^N : \partial\phi(x) \neq \emptyset\}$ if there exists $\eta \in (0, +\infty]$, a neighborhood U of x^*, and a continuous concave function $\psi : [0, \eta) \to (0, +\infty]$ such that
   - (i) $\psi(0) = 0$;
   - (ii) ψ is continuous differentiable on $(0, \eta)$;
   - (iii) for all $t \in (0, \eta)$, $\psi'(t) > 0$;
   - (iv) for all $x \in U \cap \{y \in \mathbf{R}^N : \phi(x^*) < \phi(y) < \phi(x^*) + \eta\}$, the Kurdyka-Łojasiewicz (KL) inequality holds:

$$\psi'(\phi(x) - \phi(x^*))\text{dist}(0, \partial\phi(x)) \geq 1.$$

2. If ϕ satisfies the KL property at each point of $\mathrm{dom}\partial\phi$, ϕ is called a KL function.

There is a rich class of KL functions defined based on o-minimal structure (see definition below).

Definition 9.3. (o-minimal structure on $\mathbf{R}$, Definition 4.1, [45]) Let $\mathcal{O} = \{\mathcal{O}_n\}_{n\in\mathbb{N}}$ such that each $\mathcal{O}_n$ is a collection of subsets of $\mathbf{R}^n$. The family $\mathcal{O}$ is an o-minimal structure on $\mathbf{R}$, if it satisfies the following axioms:

(i) Each $\mathcal{O}_n$ is a boolean algebra. Namely $\emptyset \in \mathcal{O}_n$ and for each A, B in $\mathcal{O}_n$, $A \cup B$, $A \cap B$, and $\mathbf{R}^n \setminus A$ belong to $\mathcal{O}_n$.
(ii) For all A in $\mathcal{O}_n$, $A \times \mathbf{R}$ and $\mathbf{R} \times A$ belong to $\mathcal{O}_{n+1}$.
(iii) For all A in $\mathcal{O}_{n+1}$, $\Pi(A) := \{(x_1, \dots, x_n) \in \mathbf{R}^n : (x_1, \dots, x_n, x_{n+1}) \in A\}$ belongs to $\mathcal{O}_n$.
(iv) For all $i \neq j$ in $\{1, 2, \dots, n\}$, $\{(x_1, \dots, x_n) \in \mathbf{R}^n : x_i = x_j\}$ belongs to $\mathcal{O}_n$.
(v) The set $\{(x_1, x_2) \in \mathbf{R}^2 : x_1 < x_2\}$ belongs to $\mathcal{O}_2$.
(vi) The elements of $\mathcal{O}_1$ are exactly finite unions of intervals.

Let $\mathcal{O}$ be an o-minimal structure on $\mathbf{R}$. We say that a set $A \subseteq \mathbf{R}^n$ is definable on $\mathcal{O}$ if $A \in \mathcal{O}_n$. A map $f : \mathbf{R}^n \to \mathbf{R}^m$ (resp. a real-extended-valued function $f : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$) is definable if its graph is a definable subset of $\mathbf{R}^n \times \mathbf{R}^m$ (resp. $\mathbf{R}^n \times \mathbf{R}$). It is known that any proper lower semicontinuous function definable on an o-minimal structure is a KL function (see Theorem 4.1 in [45]). Note that o-minimal structure has very stable properties as follows:

(1) finite sums of definable functions are definable;
(2) indicator functions of definable sets are definable;
(3) compositions of definable functions or mappings are definable.

A class of o-minimal structure is the log-exp structure (see Example 2.5, [47]). By this structure, the following functions are all definable:

1. semi-algebraic functions (see Definition 5 in [46]), such as real polynomial functions, and absolute value function $f : \mathbf{R} \to \mathbf{R}$ defined by $x \mapsto |x|$.
2. power function $f : \mathbf{R} \to \mathbf{R}$ defined by

$$x \mapsto \begin{cases} x^r, & x > 0 \\ 0, & x \leq 0, \end{cases}$$

where $r \in \mathbf{R}$.

For $f(x)$ and $F(x)$ defined in this paper, $\|Ax - d\|^2$ is semi-algebraic functions. Using properties (1) and (3) of definable functions, their regularization terms are both finite sums and compositions of definable functions. Thus, $f(x)$ and $F(x)$ are KL functions. See also [28,36].

9.3. Convergence of descent methods for KL functions

For KL functions, [38] has proven a very useful abstract convergence result for descent methods provided that three essential conditions are satisfied. See below.

Theorem 9.4. *(Theorem 2.9, [38]) Let $\phi : \mathbf{R}^N \to \mathbf{R} \cup \{+\infty\}$ be a proper lower semi-continuous function. Consider a sequence $\{x^k\}$ that satisfies the following conditions:*

- $\exists\, a \in \mathbf{R}^+, \quad s.t.\ \forall\, k \geq 1, \quad \phi(x^k) - \phi(x^{k+1}) \geq a\|x^k - x^{k+1}\|^2$;
- $\exists\, b \in \mathbf{R}^+, \quad s.t.\ \forall\, k \geq 1, \quad \exists\, u^{k+1} \in \partial\phi(x^{k+1})$ *such that* $\|u^{k+1}\| \leq b\|x^k - x^{k+1}\|$;
- $\exists$ *a subsequence* $\{x^{k_j}\}$ *and* x^*, $\quad s.t.\ x^{k_j} \to x^*$ *and* $\phi(x^{k_j}) \to \phi(x^*)$.

If ϕ *has the KL property at* x^*, *then* $\{x^k\}$ *converges to* x^*, *and* x^* *is a stationary point of* ϕ. *Moreover* $\{x^k\}$ *has a finite length, i.e.,*

$$\sum_{k=0}^{\infty} \|x^{k+1} - x^k\| < +\infty.$$

Acknowledgment

This work was supported by National Natural Science Foundation of China (Grants 11871035 and 11531013), and Recruitment Program of Global Young Experts.

References

[1] Ingrid Daubechies, Michel Defrise, Christine De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Commun. Pure Appl. Math. 57 (11) (2004) 1413–1457, A Journal Issued by the Courant Institute of Mathematical Sciences.
[2] Scott Shaobing Chen, David L. Donoho, Michael A. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (1) (2001) 129–159.
[3] Robert Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. 73 (3) (2011) 267–288.
[4] Ming Yuan, Yi Lin, Model selection and estimation in regression with grouped variables, J. R. Stat. Soc., Ser. B, Stat. Methodol. 68 (1) (2006) 49–67.
[5] Stephen Erickson, Chiara Sabatti, Empirical Bayes estimation of a sparse vector of gene expression changes, Stat. Appl. Genet. Mol. Biol. 4 (1) (2005) 22.
[6] Farzad Parvaresh, Haris Vikalo, Sidhant Misra, Babak Hassibi, Recovering sparse signals using sparse measurement matrices in compressed dna microarrays, IEEE J. Sel. Top. Signal Process. 2 (3) (2008) 275–285.
[7] M. Usman, C. Prieto, T. Schaeffter, P.G. Batchelor, k-t group sparse: a method for accelerating dynamic mri, Magn. Reson. Med. 66 (4) (2011) 1163.
[8] Dmitry Malioutov, Müjdat Cetin, Alan S. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. Signal Process. 53 (8) (2005) 3010–3022.
[9] Angshul Majumdar, Rabab K. Ward, Compressed sensing of color images, Signal Process. 90 (12) (2010) 3122–3127.
[10] Rick Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, IEEE Signal Process. Lett. 14 (10) (2007) 707–710.
[11] Rick Chartrand, Wotao Yin, Iteratively reweighted algorithms for compressive sensing, in: IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP 2008, IEEE, 2008, pp. 3869–3872.
[12] Zongben Xu, Xiangyu Chang, Fengmin Xu, Hai Zhang, $l_{1/2}$ regularization: a thresholding representation theory and a fast solver, IEEE Trans. Neural Netw. Learn. Syst. 23 (7) (2012) 1013.
[13] Tong Zhang, Analysis of multi-stage convex relaxation for sparse regularization, J. Mach. Learn. Res. 11 (Mar 2010) 1081–1107.
[14] Simon Foucart, Mingjun Lai, Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q < 1$, Appl. Comput. Harmon. Anal. 26 (3) (2009) 395–407.
[15] Yaohua Hu, Chong Li, Kaiwen Meng, Jing Qin, Xiaoqi Yang, Group sparse optimization via $\ell_{p,q}$ regularization, J. Mach. Learn. Res. 18 (1) (2017) 960–1011.
[16] Alain Rakotomamonjy, Rémi Flamary, Gilles Gasso, Stéphane Canu, $\ell_p - \ell_q$ penalty for sparse linear and sparse multiple kernel multitask learning, IEEE Trans. Neural Netw. 22 (8) (2011) 1307–1320.
[17] Emmanuel J. Candes, Terence Tao, Decoding by linear programming, IEEE Trans. Inf. Theory 51 (12) (2005) 4203–4215.
[18] Qiyu Sun, Recovery of sparsest signals via ℓ_q-minimization, Appl. Comput. Harmon. Anal. 32 (3) (2012) 329–341.
[19] Emmanuel J. Candes, Justin Romberg, Terence Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory 52 (2) (2006) 489–509.
[20] Emmanuel J. Candes, Justin K. Romberg, Terence Tao, Stable signal recovery from incomplete and inaccurate measurements, Commun. Pure Appl. Math. 59 (8) (2006) 1207–1223.
[21] Albert Cohen, Wolfgang Dahmen, Ronald Devore, Compressed sensing and best k-term approximation, J. Am. Math. Soc. 22 (1) (2009) 211–231.
[22] Emmanuel J. Candes, The restricted isometry property and its implications for compressed sensing, C. R. Math. 346 (9–10) (2008) 589–592.
[23] T. Tony Cai, Lie Wang, Guangwu Xu, Shifting inequality and recovery of sparse signals, IEEE Trans. Signal Process. 58 (3) (2009) 1300–1308.

[24] Peter J. Bickel, Ya'acov Ritov, B. Tsybakov Alexandre, Simultaneous analysis of lasso and Dantzig selector, Ann. Stat. 37 (4) (2009) 1705–1732.
[25] M. Eren Ahsen, Mathukumalli Vidyasagar, Error bounds for compressed sensing algorithms with group sparsity: a unified approach, Appl. Comput. Harmon. Anal. 43 (2) (2017) 212–232.
[26] Yonina C. Eldar, Moshe Mishali, Robust recovery of signals from a structured union of subspaces, IEEE Trans. Inf. Theory 55 (11) (2009) 5302–5316.
[27] Yao Wang, Jianjun Wang, Zongben Xu, On recovery of block-sparse signals via mixed norm minimization, EURASIP J. Adv. Signal Process. 2013 (1) (2013) 76.
[28] Yunhua Xue, Yanfei Feng, Chunlin Wu, An efficient and globally convergent algorithm for $\ell_{p,q}$-ℓ_r model in group sparse optimization, Commun. Math. Sci. 18 (1) (2020) 227–258.
[29] Ingrid Daubechies, Ronald DeVore, Massimo Fornasier, C. Sinan Güntürk, Iteratively reweighted least squares minimization for sparse recovery, Commun. Pure Appl. Math. 63 (1) (2010) 1–38.
[30] Mingjun Lai, Jingyue Wang, An unconstrained ℓ_q minimization with $0 < q \leq 1$ for sparse solution of underdetermined linear systems, SIAM J. Optim. 21 (1) (2011) 82–101.
[31] Mingjun Lai, Yangyang Xu, Wotao Yin, Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization, SIAM J. Numer. Anal. 51 (2) (2013) 927–957.
[32] Robert A. Adams, John JF Fournier, Sobolev Spaces, vol. 140, Elsevier, 2003.
[33] Xiaojun Chen, Fengmin Xu, Yinyu Ye, Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization, SIAM J. Sci. Comput. 32 (5) (2010) 2832–2852.
[34] Thomas Blumensath, Mike E. Davies, Iterative thresholding for sparse approximations, J. Fourier Anal. Appl. 14 (5–6) (2008) 629–654.
[35] Lukas Meier, Sara Van De Geer, Peter Bühlmann, The group lasso for logistic regression, J. R. Stat. Soc., Ser. B, Stat. Methodol. 70 (1) (2008) 53–71.
[36] Zhifang Liu, Chunlin Wu, Yanan Zhao, A new globally convergent algorithm for non-Lipschitz $\ell_p - \ell_q$ minimization, Adv. Comput. Math. 45 (3) (2019) 1369–1399.
[37] Chao Zeng, Rui Jia, Chunlin Wu, An iterative support shrinking algorithm for non-Lipschitz optimization in image restoration, J. Math. Imaging Vis. 61 (1) (Jan 2019) 122–139.
[38] Hedy Attouch, Jérôme Bolte, Benar Fux Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods, Math. Program. 137 (1) (Feb 2013) 91–129.
[39] R. Tyrrell Rockafellar, Roger J-B. Wets, Variational Analysis, vol. 317, Springer Science & Business Media, 2009.
[40] Stanislaw Łojasiewicz, Une propriété topologique des sous-ensembles analytiques réels, Les équations aux dérivées partielles 117 (1963) 87–89.
[41] Kurdyka Krzysztof, On gradients of functions definable in o-minimal structures, Ann. Inst. Fourier 48 (3) (1998) 769–783.
[42] Jérôme Bolte, Aris Daniilidis, Adrian Lewis, The łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems, SIAM J. Optim. 17 (4) (2007) 1205–1223.
[43] Jérôme Bolte, Aris Daniilidis, Adrian Lewis, Masahiro Shiota, Clarke subgradients of stratifiable functions, SIAM J. Optim. 18 (2) (2007) 556–572.
[44] Hédy Attouch, Jérôme Bolte, On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, Math. Program. 116 (1) (2009) 5–16.
[45] Hédy Attouch, Jérôme Bolte, Patrick Redont, Antoine Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality, Math. Oper. Res. 35 (2) (2010) 438–457.
[46] Jérôme Bolte, Shoham Sabach, Marc Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, Math. Program. 146 (1–2) (2014) 459–494.
[47] Lou Van den Dries, Chris Miller, et al., Geometric categories and o-minimal structures, Duke Math. J. 84 (2) (1996) 497–540.