A double iteratively reweighted algorithm for solving group sparse nonconvex optimization models

Wanqin Nie · Kai Tu · Minglu Ye · Shuqin Sun*

Abstract In this paper, we propose a double iteratively reweighted algorithm to solve nonconvex and nonsmooth optimization problems, where both the objectives and constraint functions are formulated by concave compositions to promote group-sparse structures. At each iteration, we combine convex surrogate with first-order information to construct linearly constrained subproblems to handle the concavity of model. The corresponding subproblems are then solved by alternating direction method of multipliers to satisfy the specific stop criteria. In particular, under mild assumptions, we prove that our algorithm guarantees the feasibility of each subsequent iteration point, and the cluster point of the resulting feasible sequence is shown to be a stationary

Wanqin Nie

E-mail: wanqinnie@hotmail.com

School of Mathematical Sciences, China West Normal University, Nanchong, Sichuan 637009, China

Kai Tu

E-mail: kaitu_02@163.com

School of Mathematical Sciences, Shenzhen University, Shenzhen 518061, China

The Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen 518060, China

Shenzhen Key Laboratory of Advanced Machine Learning and Applications, School of Mathematical Sciences, Shenzhen University, Shenzhen 518000, China

Minglu Ye

E-mail: yml2002cn@aliyun.com

Key Laboratory of Optimization Theory and Applications at China West Normal University of Sichuan Province, School of Mathematical Sciences, China West Normal University, Nanchong, Sichuan 637009, China

Shuqin Sun

E-mail: sunshuqinsusan@163.com

School of Mathematical Sciences, China West Normal University, Nanchong, Sichuan 637009, China

Key Laboratory of Optimization Theory and Applications at China West Normal University of Sichuan Province, Nanchong, Sichuan 637009, China

point. Additionally, we extend the group sparse optimization model, pioneer the application of the double iterative reweighted algorithm to solve constrained group sparse models (which exhibits superior efficiency), and incorporate a generalized Bregman distance to characterize the algorithm's termination conditions. Preliminary numerical experiments show the efficiency of the proposed method.

Keywords Group Sparsity · Iteratively Reweighted · ADMM · Non-convex

Mathematics Subject Classification (2020) 49M37 · 49N10 · 65K10 · 90C26

1 Introduction

Sparse optimization is a fundamental branch of mathematical optimization centered on seeking solutions with sparse structures (i.e., solutions containing a large proportion of zero or near-zero elements). In essence, sparse optimization techniques have seen remarkable expansion across diverse applications such as image processing [6,29], signal processing [32,33], machine learning [1], and recommendation systems [16], to name a few. Among key sparse optimization models, compressed sensing [3] enables accurate reconstruction of sparse or compressible signals with far fewer measurements than traditional sampling; matrix completion [16,39] recovers low-rank matrices from limited observed entries, widely used in recommendation systems to predict unobserved user preferences; and group sparse optimization [9,20] extends sparsity to variable groups, facilitating structured high-dimensional data analysis in fields like bioinformatics and multimedia processing.

Studying group sparse optimization is key as it addresses the structured, group-level sparsity of variables that traditional sparse optimization overlooks. Recently, the group sparse optimization model has gained wide research attention due to new application scenarios. It preserves anatomical structures in medical image denoising via group-level sparsity [9,21], and ensures consistent variable selection across gene clusters in biological data regression [24,30]. Compared with sparse optimization problems, group sparse optimization models have the following advantages: (i) It faithfully models variable correlations, yielding superior interpretability because the inherent group structure is preserved, showing through the performance in image reconstruction [21]. (ii) It demonstrates greater robustness against noise because group-level statistics are less susceptible to erratic changes in individual variables [5,11]. (iii) It simplifies post-processing by reducing computational complexity [38].

Based on group sparse optimization models, we first recall some unconstrained models. Its core in [9] is an unconstrained group sparse optimization model based on $\ell_{p,q}$ regularization, formulated as

$$\min_{x \in \mathbb{R}^n} F(x) = \|Ax - b\|^2 + \lambda \|x\|_{p,q}^q,$$

where x has a predefined group structure, $\|x\|_{p,q}$ is the $\ell_{p,q}$ norm inducing group sparsity with $p \geq 1$ and $0 \leq q \leq 1$, and $\lambda > 0$ is the regularization parameter. To solve this model, the authors propose the proximal gradient method (PGM-GSO). The model is applied to simulated data (for group sparse signal recovery, with $\ell_{p,1/2}$ regularization showing superior performance) and real gene transcriptional regulation data. Yuan and Lin [35] proposed an unconstrained group sparse model that uses least squares as the loss function and the $\ell_{2,1}$-norm regularization term to impose group-level sparsity. Fan and Li [7] put forward the smoothly clipped absolute deviation (SCAD) regularization theory based on nonconcave penalized likelihood. This theory was later extended to the unconstrained group SCAD regularization model.

Subsequently, we introduce several constrained group sparse optimization problems. For convex-constrained group sparse regularization problems with convex but non-smooth loss functions, [40] have proposed constructing models by relaxing group sparse penalty terms using group capped-ℓ_1, and designed the group smoothing proximal gradient (GSPG) algorithm for solution. To solve group sparsity-constrained minimization problems, [14] have transformed them into equivalent weighted $\ell_{p,q}$ norm constrained models. Based on the properties of Lagrangian duality, homotopy technology is adopted to handle parameter adjustment, ensuring that the algorithm output is the stationary point of the original problem. In [21], the authors consider the following special group sparse problem:

$$\min \sum_{i=1}^{m} \psi(\|x_i\|), \quad \text{s.t.} \quad \|Ax - b\| \leq \sigma, \ Bx \leq h, \tag{1}$$

where x is partitioned into m non-overlapping groups $\{x_i\}_{i=1}^{m}$ with each $x_i \in \mathbb{R}^{n_i}$ and $\sum_{i=1}^{m} n_i = n$, the group-wise sparsity induces vanishing entries, ψ is a capped folded concave function. They propose smoothing penalty algorithm (**SPA** for short) to solve this problem and show that the sequences generated by the algorithm converge to the directional stationary points. Based on the smoothing technique and penalty strategy, **SPA** needs to apply nonmonotone proximal gradient to solve a unconstrained nonconvex sub-problem until a specific criteria holds.

In the research progress of the models, scholars have found that incorporating nonconvex relaxation terms into the objective function, such as the SCAD penalty [7, 15, 34] and minimax concave penalty (MCP) [12, 36] can better promote sparsity and reduce estimation bias. These properties make them particularly suitable for high-dimensional data analysis and variable selection tasks. In contrast, the use of nonconvex loss functions in the constraint component, including Tukey loss [19], Huber loss [10], the 0-1 loss [27], and its surrogate losses [13] enhances the robustness of the model, enabling effective handling of noise and outliers. Investigating problem models that integrate nonconvex relaxation functions with nonconvex loss functions holds significant importance in sparse optimization and complex data modeling. Despite the increased computational challenges associated with solving nonconvex problems, such models

offer superior sparsity, robustness, and model flexibility. Consequently, the following nonconvex sparse optimization problem was considered in [26]:

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^n \psi(|x_i|), \quad \text{s.t.} \quad \sum_{i=1}^m \phi\left(\left(a_i^T x - b_i\right)^2\right) \le \sigma, \tag{2}$$

which incorporates a nonconvex relaxation in the objective function and a nonconvex loss function in the constraint part. A double iterative reweighting algorithm was devised by the authors to solve this problem, in which the subproblem was approximated and solved using a convex subproblem solver.

Given the aforementioned necessity of employing group sparse optimization, we extend the problem (2) to address the group sparisty of decision variables. That is, we consider the following group sparse optimization problem:

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^q \psi(\|x_{\mathcal{G}_i}\|), \quad \text{s.t.} \quad \sum_{i=1}^m \phi\left(\left(a_i^T x - b_i\right)^2\right) \le \sigma, \tag{3}$$

where $\{\mathcal{G}_1, \cdots, \mathcal{G}_q\}$ forms a partition of the index set $[n] = \{1, 2, \cdots, n\}$ and satisfies $\mathcal{G}_i \cap \mathcal{G}_j = \emptyset$ for all $i \ne j$ and $\cup_{i=1}^q \mathcal{G}_i = [n]$, $x_{\mathcal{G}_i} \in \mathbb{R}^{|\mathcal{G}_i|}$ denotes the i-th block of x indexed by $\mathcal{G}_i$. Then we present $x_{\mathcal{G}} = \left(\|x_{\mathcal{G}_1}\|, \cdots, \|x_{\mathcal{G}_q}\|\right)^T$, which facilitates the subsequent discussion. Furthermore, $m \ll n$, a_i^T represents the i-th row of the matrix $A \in \mathbb{R}^{m \times n}$; $b \in \mathbb{R}^m$, $\sigma > 0$ and two functions $\psi, \phi : \mathbb{R}_+ \to \mathbb{R}_+$ satisfy the following assumptions:

Assumption 1 (i) *The function ψ is continuous and strictly concave over $\mathbb{R}_+$ with $\psi(0) = 0$ and $\lim_{t \to \infty} \psi(t) = \infty$. ψ is differentiable on $(0, \infty)$, and $\psi'(t) > 0$ for all $t \in (0, \infty)$. Furthermore, $\lim_{t \downarrow 0} \psi'(t)$ exists and satisfies $\lim_{t \downarrow 0} \psi'(t) \in (0, \infty)$.*

(ii) The function ϕ is continuous and concave over $\mathbb{R}_+$ with $\phi(0) = 0$. It is differentiable on $(0, \infty)$, and $\phi'(t) \ge 0$ for all $t > 0$. Furthermore, $\lim_{t \downarrow 0} \phi'(t)$ exists and satisfies $\lim_{t \downarrow 0} \phi'(t) \in (0, \infty)$.

(iii) The matrix A is full row rank, and $\sigma \in \left(0, \sum_{i=1}^m \phi\left(b_i^2\right)\right)$.

By making the above assumptions, we ensure that the feasible set of (3) does not contain $x = 0$. Moreover, since ψ and ϕ are differentiable and concave, their right derivatives ψ'_+ and ϕ'_+ are continuous. This continuity proves useful in our subsequent analysis. It can also be observed that problem (3) reduces to problem (2) in [26] when $q = n$ given they both satisfy Assumption 1. However, due to the existence of gourp structure, the method in [26] can't directly apply to solve the new model (3). The main contributions of this work are as follows:

- We extend the group sparse optimization model by introducing non-convex inequality constraints, enhancing its robustness in practical scenarios. When recovering group structured signals via the double iterative reweighted algorithm, problem (3) outperforms problem (2) in recovery performance.

- The double iterative reweighted algorithm is first applied to solve constrained group sparse models, demonstrating superior efficiency compared to existing methods. When recovering the same group structured signal via the same model using different group sparse optimization algorithms, the algorithm proposed in this paper outperforms the one in [21]. We note that while both this paper and [26] use the iteratively reweighted algorithm, our model (a generalization of [26] under group sparsity) lets our algorithm overcome the problem of solving the proximal operator in the inner loop. This challenge does not exist in [26].
- A generalized Bregman distance is incorporated to characterize termination conditions for nonlinear non-strongly convex subproblems, improving the applicability of the algorithm. Although increase the difficulty of analyzing the convergence of the designed algorithm and verifying whether the subproblems meet the termination criteria, we have successfully overcome these challenges.

The rest of this paper is organized as follows. In Section 2, we introduce notations and preliminary materials to be used in subsequent analysis. In Section 3, we present the algorithm flow of problem (3), the construction process of the convex subproblem, and the progress of ADMM for solving the subproblem. Additionally, in this section, we also use Theorem 1 to illustrate that the ADMM satisfies the three termination conditions set by the algorithm. In Section 4, we establish the convergence analysis of the algorithm to validate its feasibility. Finally, in Section 5, we verify the effectiveness of the algorithm through numerical comparison experiments conducted in MATLAB.

2 Notation and preliminaries

Throughout this paper, we denote $\mathbb{R}^n$ as the n-dimensional Euclidean space, $\mathbb{R}^n_+$ as the nonnegative orthant of $\mathbb{R}^n$ and let $\mathbb{N}$ denote the set of natural numbers. For a given vector $x \in \mathbb{R}^n$, x_i denotes its i-th component. The notation $\sqrt{x}$ is used to represent the vector whose i-th component is $\sqrt{x_i}$, and $\|x\|$ denotes the ℓ_2-norm of x. We let $\text{Diag}(x)$ be the $n \times n$ diagonal matrix with the i-th diagonal entry equal to x_i. For any two vectors $u, v \in \mathbb{R}^n$, $u \circ v$ and $\langle u, v \rangle$ denote their Hadamard (entry-wise) product, and inner product, respectively. For any $x, y \in \mathbb{R}^n$, they have the properties $\|x\| = \|x_{\mathcal{G}}\|$ and $\|x_{\mathcal{G}} - y_{\mathcal{G}}\| \le \|(x - y)_{\mathcal{G}}\|$, which ensures that Proposition 1(ii) holds.

We say that an extended-real-valued function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is proper if its domain $\text{dom} f = \{x \in \mathbb{R}^n : f(x) < \infty\}$ is nonempty. A proper function is said to be closed if it is lower semicontinuous. For a proper function f, the regular subdifferential and Mordukhovich (limiting) subdifferential [25, Definition 8.3] of f at $\bar{x} \in \text{dom} f$ are defined respectively as

$$\widehat{\partial} f(\bar{x}) = \left\{ v \in \mathbb{R}^n : \liminf_{x \to \bar{x}, x \neq \bar{x}} \frac{f(x) - f(\bar{x}) - v^T(x - \bar{x})}{\|x - \bar{x}\|} \ge 0 \right\},$$

and

$$\partial f(\bar{x}) = \left\{ v \in \mathbb{R}^n : \exists x^k \to \bar{x} \text{ and } v^k \in \widehat{\partial} f\left(x^k\right) \text{ s.t. } f(x^k) \to f(\bar{x}) \text{ and } v^k \to v \right\}.$$

By convention, we set $\widehat{\partial} f(x) = \partial f(x) = \emptyset$ if $x \notin \text{dom } f$, and define the domain of ∂f as $\text{dom}\partial f = \{x \in \mathbb{R}^n : \partial f(x) \neq \emptyset\}$. When f is proper and convex, the limiting subdifferential of f at $x \in \text{dom } f$ reduces to the classical subdifferential in convex analysis, i.e.,

$$\partial f(x) = \left\{ \xi \in \mathbb{R}^n : \xi^T(y-x) \leq f(y) - f(x) \ \ \forall y \in \mathbb{R}^n \right\};$$

see [25, Proposition 8.12]. For a nonempty set B, the indicator function δ_B is defined as

$$\delta_B(x) = \begin{cases} 0, & \text{if } x \in B, \\ \infty, & \text{if } x \notin B. \end{cases}$$

The normal cone (resp., regular normal cone) of B at $x \in B$ is defined as $N_B(x) = \partial \delta_B(x)$ (resp., $\widehat{N}_B(x) = \widehat{\partial} \delta_B(x)$), and the distance from any $x \in \mathbb{R}^n$ to B is defined as $\text{dist}(x, B) = \inf_{y \in B} \|x - y\|$, and if $B = \emptyset$, we take $d(x, B) = \infty$ by convention. If f is proper and closed, then the proximal mapping of f at $x \in \mathbb{R}^n$ with scaling parameter $\lambda > 0$ is defined as

$$\text{prox}_{\lambda f}(x) = \arg\min_{u \in \mathbb{R}^n} \left\{ f(u) + \frac{1}{2\lambda} \|x - u\|^2 \right\}. \tag{4}$$

The Bregman distance [2,31], also known as the Bregman divergence, has been extensively studied by scholars and is defined as follows.

Definition 1 (Bregman distance) Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a function that is continuously differentiable and strictly convex. For any two points $x \in \mathbb{R}^n$ and $x_0 \in \mathbb{R}^n \cap \text{dom } f$, the Bregman distance from x_0 to x is defined as:

$$D_f(x, x_0) = f(x) - f(x_0) - \langle \nabla f(x_0), x - x_0 \rangle.$$

Using this definition, we define the Bregman distance from a point x to a set $C \subset \mathbb{R}^n$ as $D_f(C, x)$ in the form:

$$D_f(C, x) = \inf_{y \in C} D_f(y, x).$$

If $C = \emptyset$, we take $D_f(C, x) = \infty$ by convention. If $f = \|\cdot\|^2$, then the Bregman distance reduces to the square of the Euclidean distance, that is, $D_f(C, x) = \inf_{y \in C} \|y - x\|^2 = \text{dist}^2(x, C)$. To proceed with the convergence analysis in Section 4, we apply [28, Theorem 2.4] with the choices $S = \mathbb{R}^n$ and $y^k = 0$ for all k and obtain the following Fact 1 concerning Bregman distance.

Fact 1 Let f be defined in Definition 1. Given the sequence $\{x^n\}_{n \in \mathbb{N}} \subseteq \mathbb{R}^n$, if

$$\lim_{n \to \infty} D^f(x^n, 0) = 0,$$

then we have that $\lim_{n \to \infty} x^n = 0$.

Consider problem (3) under Assumption 1. We introduce the following notations for the convenience of the forthcoming discussion. For any $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m_+$, and $z \in \mathbb{R}^q_+$, we define

$$\Phi(y) = \sum_{i=1}^m \phi(y_i), \ \Phi'_+(y) = \left(\phi'_+(y_1), \phi'_+(y_2), \cdots, \phi'_+(y_m)\right)^T,$$

$$\Psi(z) = \sum_{i=1}^q \psi(z_i), \ \Psi'_+(z) = \left(\psi'_+(z_1), \psi'_+(z_2), \cdots, \psi'_+(z_q)\right)^T, \quad \text{and} \quad (5)$$

$$G(x) = x_{\mathcal{G}} = \left(\|x_{\mathcal{G}_1}\|, \|x_{\mathcal{G}_2}\|, \cdots, \|x_{\mathcal{G}_q}\|\right)^T,$$

where $\mathcal{G}_i \ (i = 1, 2, \cdots, q)$ is defined in (3).

We next present the definition of the stationary point of problem (3).

Definition 2 A point $x \in \mathbb{R}^n$ is called a stationary point of problem (3) if there exists $\lambda \in \mathbb{R}_+$ such that the following conditions hold for (x, λ):

$$\lambda \left(\Phi\left((Ax - b) \circ (Ax - b)\right) - \sigma\right) = 0, \tag{6}$$

$$\Phi\left((Ax - b) \circ (Ax - b)\right) \leq \sigma, \tag{7}$$

$$0 \in \Psi'_+ \left(x_{\mathcal{G}}\right) \circ \partial G(x) + 2\lambda \sum_{i=1}^m \phi'_+ \left(\left(a_i^T x - b_i\right)^2\right) \left(a_i^T x - b_i\right) a_i, \tag{8}$$

The following fact reveals that any local minimzer of problem (3) is the stationary point under Mangasarian-Fromovitz constraint qualification (9). This follows by sequentially applying arguments analogous to those in [26, Proposition 2.1] and then [26, Proposition 2.2] under Assumption 1, exploiting the local Lipschitz continuity of $\psi \circ \| \cdot \|$.

Fact 2 *Suppose that Assumption 1 holds. If there exists $y \in \mathbb{R}^m$ such that $y = Ax - b$ and $\sum_{i=1}^m \phi\left(y_i^2\right) = \sigma$, where $\sigma \notin \left\{k\bar{\phi}: k = 1, \cdots, m\right\}$, with $\bar{\phi} = \sup_{t \in \mathbb{R}_+} \phi(t) \in (0, \infty]$. Then Mangasarian-Fromovitz constraint qualification holds, that is*

$$\Phi\left((Ax - b) \circ (Ax - b)\right) = \sigma \implies \sum_{i=1}^m \phi'_+ \left((a_i^T x - b_i)^2\right) \left(a_i^T - b_i\right) a_i \neq 0. \tag{9}$$

Furthermore, any local minimizer of problem (3) is its stationary point.

3 Double iteratively reweighted algorithm for group sparsity models

In this section, we present our algorithm to solve problem (3) whose feasible set is defined by $\mathfrak{F} = \{x \in \mathbb{R}^n : \Phi\left((Ax - b) \circ (Ax - b)\right) \leq \sigma\}$. We begin by

outlining the idea of algorithm design based on the assumptions in Fact 2. Taking advantage of the concavity of ϕ and ψ, given x^k, we have that

$$\sum_{i=1}^{q} \psi\left(\|x_{\mathcal{G}_i}\|\right) \leq \sum_{i=1}^{q} \psi\left(\|x_{\mathcal{G}_i}^k\|\right) + \sum_{i=1}^{q} \psi'_+\left(\|x_{\mathcal{G}_i}^k\|\right)\left(\|x_{\mathcal{G}_i}\| - \|x_{\mathcal{G}_i}^k\|\right), \quad \text{and}$$

$$\sum_{i=1}^{m} \phi\left(\left(a_i^T x - b_i\right)^2\right) \leq \Phi\left((Ax^k - b) \circ (Ax^k - b)\right)$$
$$+ \sum_{i=1}^{m} \phi'_+\left(\left(a_i^T x^k - b_i\right)^2\right)\left[\left(a_i^T x - b_i\right)^2 - \left(a_i^T x^k - b_i\right)^2\right].$$

The above two inequalities inspire us to formulate the following relaxed subproblem:

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{q} \psi'_+\left(\|x_{\mathcal{G}_i}^k\|\right) \cdot \|x_{\mathcal{G}_i}\|$$
$$\text{s.t. } \Phi\left((Ax^k - b) \circ (Ax^k - b)\right) \tag{10}$$
$$+ \sum_{i=1}^{m} \phi'_+\left(\left(a_i^T x^k - b_i\right)^2\right)\left[\left(a_i^T x - b_i\right)^2 - \left(a_i^T x^k - b_i\right)^2\right] \leq \sigma,$$

which, by using the notations in (5), can be reformulated as

$$\min_{x \in \mathbb{R}^n} \|\omega_{\mathcal{G}}^k \circ G(x)\|_1, \quad \text{s.t.} \quad A_k x - u = b^k, \quad u \in U^k, \tag{11}$$

where

$$\omega_{\mathcal{G}}^k = \Psi'_+\left(x_{\mathcal{G}}^k\right), \quad A_k = \text{Diag}\left(v^k\right) A \text{ with } v^k = \sqrt{\Phi'_+(y^k \circ y^k)} \text{ and } y^k = Ax^k - b,$$
$$b^k = v^k \circ b, \quad \sigma_k = \sigma + \|A_k x^k - b^k\|^2 - \Phi\left(y^k \circ y^k\right), \quad U^k = \{u : \|u\|^2 \leq \sigma_k\}. \tag{12}$$

Let $\mathfrak{F}_k$ denote the feasible set of (11), that is, $\mathfrak{F}_k = \{x \in \mathbb{R}^n : \|A_k x - b^k\|^2 \leq \sigma_k\}$. Suppose further that x^k is a feasible point of problem (3). Then from (10) and (12), we can see that $x^k \in \mathfrak{F}_k$. This ensures nonemptiness of $\mathfrak{F}_k$. Moreover, since the objective in (11) is level-bounded which is implied by the positivity of $\omega_{\mathcal{G}}^k$, the solution set of subproblem (11) is nonempty. Therefore, the subproblem (11) is well defined if $x^k \in \mathfrak{F}$.

Note that the functions in subproblem (11) are convex. Thus, if subproblem (11) is valid, it can be inexactly solved by some first order algorithms efficiently, such as alternating direction method of multipliers (ADMM) that we can see later. Here we adopt the Bregman distance to measure one of the termination conditions of the subproblem computation instead of the usual European distance used in [26]. However, the resulting approximate point maybe infeasible to problem (11). In order to guarantee the feasibility, we also need to retract the approximate solution to $\mathfrak{F}_k$.

Based on above reason, we subsequently present our Algorithm 1 (**GIR** for short) below for solving problem (3) under the assumptions in Fact 2.

Algorithm 1 Doubly iteratively reweighted algorithm.

1: **Input:** Take a positive diminishing sequence $\{\tau_k\}_{k\in\mathbb{N}}$, and a summable positive sequence $\{\mu_k\}_{k\in\mathbb{N}}$. Set $x^0 = A^\dagger b$ and $k = 0$.

2: **while** the stopping criteria does not hold **do**

3:    Compute $\omega_{\mathcal{G}}^k$, y^k, υ^k, A_k, b^k and σ_k as defined in (12).

4:    Find a pair $\left(\widetilde{x}^{k+1}, \widetilde{u}^{k+1}\right)$ by inexactly solving subproblem (11) such that the following three conditions hold:

$$D^f\left(\partial\left(\omega_{\mathcal{G}}^k \circ G\right)\left(\tilde{x}^{k+1}\right) + A_k^T N_{U^k}\left(\tilde{u}^{k+1}\right),\ 0\right) \le \epsilon_k, \qquad (13)$$

$$\|A_k\tilde{x}^{k+1} - b^k - \tilde{u}^{k+1}\| \le \epsilon_k, \qquad (14)$$

$$\|\omega_{\mathcal{G}}^k \circ G(P_k(\tilde{x}^{k+1}))\|_1 \le \|\omega_{\mathcal{G}}^k \circ x_{\mathcal{G}}^k\|_1 + \mu_k, \qquad (15)$$

where $\epsilon_k \in \left(0, \min\left\{\sigma_k, \sqrt{\sigma_k}, \tau_k\right\}\right]$,

$$P_k(x) = \begin{cases} x, & \text{if } \|A_k x - b^k\|^2 \le \sigma_k, \\ \left(1 - \frac{\sqrt{\sigma_k}}{\|A_k x - b^k\|}\right) A^\dagger b + \frac{\sqrt{\sigma_k}}{\|A_k x - b^k\|}x, & \text{otherwise.} \end{cases}$$

5:    Set $x^{k+1} = P_k\left(\tilde{x}^{k+1}\right)$, and $k = k + 1$.

6: **end while**

Remark 1 We have the following comments for Algorithm 1.

(i) Due to the definition of $\mathfrak{F}$ and $A^\dagger$, we have that $x^0 = A^\dagger b \in \mathfrak{F}$.

(ii) From [17], we have

$$\partial\left(\omega_{\mathcal{G}}^k \circ G\right)\left(x^{k+1}\right) = \omega_{\mathcal{G}}^k \circ \partial G\left(x^{k+1}\right) = J_1 \times J_2 \times \cdots \times J_q. \qquad (16)$$

where for any $i \in \{1, \cdots, q\}$,

$$J_i = \begin{cases} \left(\omega_{\mathcal{G}}^k\right)_i \cdot \frac{x_{\mathcal{G}_i}^{k+1}}{\|x_{\mathcal{G}_i}^{k+1}\|}, & \text{if } x_{\mathcal{G}_i}^{k+1} \ne 0, \\ B\left(0, (\omega_{\mathcal{G}}^k)_i\right), & \text{otherwise.} \end{cases}$$

(iii) In view of the concavity of ϕ, we have $P_k(x) \in \mathfrak{F}_k \subseteq \mathfrak{F}$ for all $x \in \mathbb{R}^n$.

(iv) It follows from [26, Lemma 3.1] that for each k, $0 < \sigma_k \le \sigma$.

3.1 Subproblem computation

In this subsection, we apply proximal ADMM to solve the inner subproblem (11). As outlined in Algorithm 1, if $x^k \in \mathfrak{F}$, then subproblem (11) is well defined by virtue of having a nonempty solution set. Thus in view of Remark 1, which shows that $x^0 \in \mathfrak{F}$ and all generated iterates x^{k+1} remain in $\mathfrak{F}$, we can see that Algorithm 1 is well-defined by showing that the tuple $\left(\widetilde{x}^{k+1}, \widetilde{u}^{k+1}\right)$ generated by the subproblem solver satisfies the inexact criteria (13), (14) and (15).

Note that subproblem (11) is a convex problem with two block variables, decision varibale x and the slack variable u, and one linear constraint. These

structures motivate us apply proximal ADMM to solve it. Let L_β be the augmented lagrangian function of (11), that is,

$$L_\beta(x, u; z) = \|\omega_{\mathcal{G}}^k \circ G(x)\|_1 + \delta_{U^k}(u) - z^T(A_k x - u - b^k) + \frac{\beta}{2}\|A_k x - u - b^k\|^2,$$

where z is the Lagrange multiplier, and β denotes the penalty coefficient. Then the basic process of solving (11) in Algorithm 1 via the proximal ADMM is shown in Algorithm 2.

Algorithm 2 Proximal ADMM for (11).

1: **Input:** Take $\beta > 0$, $r \in \left(0, \frac{1+\sqrt{5}}{2}\right)$, $\rho > 0$, and $\left(x^{k,0}, u^{k,0}, z^{k,0}\right)$. Set $l = 0$.

2: **while** the point $\left(x^{k,l}, u^{k,l}\right)$ does not satisfy (13), (14) and (15) **do**

3:     Compute

$$x^{k,l+1} \in \underset{x \in \mathbb{R}^n}{\text{Arg min}} \left\{ L_\beta\left(x, u^{k,l}; z^{k,l}\right) + \frac{1}{2}\left(x - x^{k,l}\right)^T\left(\rho I - \beta A_k^T A_k\right)\left(x - x^{k,l}\right) \right\}.$$

4:     Update $u^{k,l+1} = \underset{u \in \mathbb{R}^m}{\arg\min}\left\{ L_\beta(x^{k,l+1}, u; z^{k,l}) \right\}$, that is,

$$u^{k,l+1} = \frac{\sqrt{\sigma_k}\left(A_k x^{k,l+1} - b^k - \frac{1}{\beta}z^{k,l}\right)}{\max\left\{\sqrt{\sigma_k}, \|A_k x^{k,l+1} - b^k - \frac{1}{\beta}z^{k,l}\|\right\}}.$$

5:     Calculate $z^{k,l+1} = z^{k,l} - r\beta\left(A_k x^{k,l+1} - b^k - u^{k,l+1}\right)$ and set $l = l + 1$.

6: **end while**

Remark 2 We point out that in Algorithm 2, $\rho = \bar{L}\beta > 0$ where

$$\bar{L} = \max_i \left\{ \phi'_+\left(\left(a_i^T x^k - b_i\right)^2\right) \right\} \lambda_{\max}\left(A^T A\right).$$

Since A has full row rank, it follows that $\lambda_{\max}\left(A^T A\right) > 0$. Then we have $\bar{L} \geq \lambda_{\max}(A_k^T A_k)$ and $\rho I - \beta A_k^T A_k \succeq 0$.

Remark 3 We present the details on the computation of $x^{k,l+1}$. Using the definition of L_β and (12), the detailed update rule for $x^{k,l+1}$ is

$$x^{k,l+1} \in \underset{x \in \mathbb{R}^n}{\text{Arg min}} \sum_{i=1}^q \left\{ \frac{(w_{\mathcal{G}}^k)_i}{\rho} \cdot \|x_{\mathcal{G}_i}\| + \frac{1}{2}\|x_{\mathcal{G}_i} - v_{\mathcal{G}_i}^{k,l}\|^2 \right\},$$

where $v^{k,l} = x^{k,l} - \frac{\beta}{\rho}A_k^T\left(A_k x^{k,l} - b^k - u^{k,l} - \frac{z^{k,l}}{\beta}\right)$. Using the proximal mapping of $\|\cdot\|$, we obtain that $x^{k,l+1} = \left(\left(x_{\mathcal{G}_1}^{k,l+1}\right)^T, \cdots, \left(x_{\mathcal{G}_q}^{k,l+1}\right)^T\right)^T$, where

for $i = 1, 2, \cdots, q$,

$$
x_{\mathcal{G}_i}^{k,l+1} = \begin{cases} \left(1 - \dfrac{\frac{(w_{\mathcal{G}}^k)_i}{\rho}}{\|v_{\mathcal{G}_i}^{k,l}\|}\right) \cdot v_{\mathcal{G}_i}^{k,l}, & \text{if } \|x_{\mathcal{G}_i}^{k,l}\| \geq \dfrac{(w_{\mathcal{G}}^k)_i}{\rho}, \\ 0, & \text{otherwise.} \end{cases}
$$

Theorem 1 *Suppose that Assumption 1 hold, and that the sequences $\{x^{k,l}\}_{l\in\mathbb{N}}$ and $\{u^{k,l}\}_{l\in\mathbb{N}}$ are generated by Algorithm 2. Then for sufficiently large l, the pair $\left(x^{k,l}, u^{k,l}\right)$ satisfies conditions (13)-(15).*

Proof Before we proceed with the discussion, we first point out from [8, Theorem B.1] that the sequence $\{x^{k,l}\}_{l\in\mathbb{N}}$ converges to an optimal solution $x^{k,\star}$ of subproblem (11), and letting $l \to \infty$, we also have

$$
\|x^{k,l} - x^{k,l+1}\| \to 0, \ \|z^{k,l} - z^{k,l+1}\| \to 0, \ \|u^{k,l} - u^{k,l+1}\| \to 0. \tag{17}
$$

Furthermore, by the definitions of $\omega_{\mathcal{G}}$ and $G(\cdot)$, we get

$$
\|\omega_{\mathcal{G}}^k \circ G(x^{k,l})\|_1 \to \|\omega_{\mathcal{G}}^k \circ G(x^{k,\star})\|_1. \tag{18}
$$

In fact, using the optimality conditions during the update of x and u in steps 3-4 of Algorithm 2, we derive that

$$
\begin{cases} 0 \in \omega_{\mathcal{G}}^k \circ \partial G(x^{k,l+1}) - A_k^T z^{k,l} + \beta A_k^T \left(A_k x^{k,l+1} - b^k - u^{k,l}\right) \\ \quad + \left(\rho I - \beta A_k^T A_k\right)\left(x^{k,l+1} - x^{k,l}\right), \\ 0 \in N_{U^k}(u^{k,l+1}) + z^{k,l} - \beta \left(A_k x^{k,l+1} - b^k - u^{k,l+1}\right), \\ z^{k,l+1} = z^{k,l} - r\beta \left(A_k x^{k,l+1} - b^k - u^{k,l+1}\right), \end{cases} \tag{19}
$$

the first two relations in (19) implies that

$$
-\beta A_k^T\left(u^{k,l+1} - u^{k,l}\right) - \left(\rho I - \beta A_k^T A_k\right)\left(x^{k,l+1} - x^{k,l}\right) \in A_k^T N_{U^k}(u^{k,l+1}) + \omega_{\mathcal{G}}^k \circ \partial G(x^{k,l+1}).
$$

Combing this relation with (17), Remark 2 and the definition ϵ_k, for all sufficiently large l, we have

$$
\begin{aligned} &\| - \beta A_k^T \left(u^{k,l+1} - u^{k,l}\right) - \left(\rho I - \beta A_k^T A_k\right)\left(x^{k,l+1} - x^{k,l}\right) \| \\ &\leq \| - \beta A_k^T\|\|u^{k,l+1} - u^{k,l}\| + \|\rho I - \beta A_k^T A_k\|\|x^{k,l+1} - x^{k,l}\| \leq \epsilon_k. \end{aligned}
$$

Using the definition of D_f, this further implies that for all sufficiently large l,

$$
\begin{aligned} &D_f\left(\omega_{\mathcal{G}}^k \circ \partial G(x^{k,l+1}) + A_k^T N_{U^k}(u^{k,l+1}), \ 0\right) \\ &\leq D_f\left(-\beta A_k^T\left(u^{k,l+1} - u^{k,l}\right) - \left(\rho I - \beta A_k^T A_k\right)\left(x^{k,l+1} - x^{k,l}\right), \ 0\right) \leq \epsilon_k, \end{aligned}
$$

which confirms that criterion (13) holds. For criteria (14), from the second relation in (17) and the identity $\|A_k x^{k,l+1} - b^k - u^{k,l+1}\| = \frac{1}{r\beta}\|z^{k,l} - z^{k,l+1}\|$ (from the third relation in (19)), we have that for all sufficiently large l, $\|A_k x^{k,l+1} - b^k - u^{k,l+1}\| \leq \epsilon_k$ holds. This means (14) holds.

It remains to verify criterion (15). By use of (17), (18) and the definition of P_k in Algorithm 1, we have that $\lim_{l \to \infty} \|\omega_{\mathcal{G}}^k \circ \left(P_k(x^{k,l+1})\right)_{\mathcal{G}}\|_1 = \lim_{l \to \infty} \|\omega_{\mathcal{G}}^k \circ x_{\mathcal{G}}^{k,l+1}\|_1 = \lim_{l \to \infty} \|\omega_{\mathcal{G}}^k \circ x_{\mathcal{G}}^{k,l}\|_1 = \|\omega_{\mathcal{G}}^k \circ x_{\mathcal{G}}^{k,\star}\|_1$. This together with $\mu_k > 0$ implies that for sufficiently large l,

$$\|\omega_{\mathcal{G}}^k \circ \left(P_k(x^{k,l+1})\right)_{\mathcal{G}}\|_1 < \|\omega_{\mathcal{G}}^k \circ x_{\mathcal{G}}^{k,\star}\|_1 + \mu_k.$$

Note that x^k is feasible for the subproblem (11) and $x^{k,\star}$ is an optimal solution of (11), so $\|\omega_{\mathcal{G}}^k \circ x_{\mathcal{G}}^{k,\star}\|_1 \leq \|\omega_{\mathcal{G}}^k \circ x_{\mathcal{G}}^k\|_1$. Thus, we obtain that

$$\|\omega_{\mathcal{G}}^k \circ \left(P_k(x^{k,l+1})\right)_{\mathcal{G}}\|_1 + \mu_k < \|\omega_{\mathcal{G}}^k \circ x_{\mathcal{G}}^k\|_1 + \mu_k,$$

which yields that criteria (15) holds.

4 Convergence analysis

We establish the following proposition through a similar proof as demonstrated in [26, Proposition 3.1]. Here we omit the proof for brevity.

Proposition 1 *Suppose Assumptions 1 and the conditions in Fact 2 hold. Consider Algorithm 1 to solve problem (3). Let $\{x^k\}_{k \in \mathbb{N}}$ and $\{\tilde{x}^k\}_{k \in \mathbb{N}}$, be generated by Algorithm 1. Then the following statements hold:*

(i) *The sequences $\{x^k\}_{k \in \mathbb{N}}$, $\{\tilde{x}^k\}_{k \in \mathbb{N}}$, $\{x_{\mathcal{G}}^k\}_{k \in \mathbb{N}}$, $\{\tilde{x}_{\mathcal{G}}^k\}_{k \in \mathbb{N}}$ are bounded, and for all k, the following inequality holds*

$$\Psi\left(x_{\mathcal{G}}^{k+1}\right) - \Psi\left(x_{\mathcal{G}}^k\right) \leq \mu_k.$$

Furthermore, the sequence $\left\{\Psi\left(x_{\mathcal{G}}^k\right)\right\}_{k \in \mathbb{N}}$ is convergent.

(ii) *There exists constant $M > 0$ such that for all k,*

$$\|x_{\mathcal{G}}^{k+1} - \tilde{x}_{\mathcal{G}}^{k+1}\| \leq \|\left(x^{k+1} - \tilde{x}^{k+1}\right)_{\mathcal{G}}\| = \|x^{k+1} - \tilde{x}^{k+1}\| \leq \sqrt{\epsilon_k}M.$$

(iii) $\lim_{k \to \infty} \|\tilde{x}_{\mathcal{G}}^{k+1} - x_{\mathcal{G}}^k\| = 0.$

Remark 4 By Proposition 1 (i), the sequences $\{\omega_{\mathcal{G}}^k\}_{k \in \mathbb{N}}$, $\{A_k\}_{k \in \mathbb{N}}$, $\{b^k\}_{k \in \mathbb{N}}$ and $\{\sigma_k\}_{k \in \mathbb{N}}$ generated by Algorithm 1 are all bounded, and hence the sequence $\{A_k\tilde{x}^{k+1}\}_{k \in \mathbb{N}}$ is also bounded.

To show the convergence analysis of Algorithm 1, we first note from (13) that for each k, there exists $\xi^k \in \mathbb{R}^n$ with $D^f\left(\xi^k, 0\right) \leq \epsilon_k$ such that

$$\xi^k \in \omega_{\mathcal{G}}^k \circ \partial G\left(\tilde{x}^{k+1}\right) + A_k^T N_{U^k}\left(\tilde{u}^{k+1}\right),$$

which, combined with $\lim_{k \to \infty} \epsilon_k = 0$, implies $\lim_{k \to \infty} D^f\left(\xi^k, 0\right) = 0$. This together with Fact 1 yields

$$\lim_{k \to \infty} \|\xi^k\| = 0. \tag{20}$$

Using the definition of normal cone and the fact that $\sigma_k > 0$, we further derive the existence of $\tilde{\lambda}_k \geq 0$ satisfying

$$\xi^k \in \omega_{\mathcal{G}}^k \circ \partial G(\tilde{x}^{k+1}) + \tilde{\lambda}_k A_k^T \tilde{u}^{k+1} \quad \text{and} \quad \tilde{\lambda}_k \left(\|\tilde{u}^{k+1}\|^2 - \sigma_k \right) = 0. \qquad (21)$$

Define

$$\tilde{v}^{k+1} = A_k \tilde{x}^{k+1} - b^k - \tilde{u}^{k+1}. \qquad (22)$$

By (14), we have $\|\tilde{v}^{k+1}\| \leq \epsilon_k \leq \min \left\{ \sigma_k, \sqrt{\sigma_k} \right\}$. Furthermore, from (21) it follows that

$$\tilde{x}^{k+1} \in \operatorname*{Arg\,min}_{x \in \mathbb{R}^n} \left\{ (\omega_{\mathcal{G}}^k)^T G(x) - (\xi^k)^T x + \frac{1}{2} \tilde{\lambda}_k \left(\|A_k x - b^k - \tilde{v}^{k+1}\|^2 \right) \right\}, \quad (23)$$

and

$$\tilde{\lambda}_k \left(\|A_k \tilde{x}^{k+1} - b^k - \tilde{v}^{k+1}\|^2 - \sigma_k \right) = 0. \qquad (24)$$

We are now ready to establish the subsequential convergence of the iterates generated by Algorithm 1, following an argument analogous to that in [26, Theorem 3.1]. For completeness and self-containment, we present the full proof below.

Theorem 2 *Consider* (3) *under Assumptions* 1, *Facts* 1 *and* 2. *Let* $\{x^k\}_{k \in \mathbb{N}}$, $\{\tilde{x}^k\}_{k \in \mathbb{N}}$, $\{x_{\mathcal{G}}^k\}_{k \in \mathbb{N}}$ *and* $\{\tilde{x}_{\mathcal{G}}^k\}_{k \in \mathbb{N}}$ *be generated by Algorithm* 1, *and let* $\{\tilde{\lambda}_k\}_{k \in \mathbb{N}}$ *be defined in* (21). *Then the following statements hold:*

(i) $\liminf\limits_{k \to \infty} \tilde{\lambda}_k > 0$.

(ii) $\lim\limits_{k \to \infty} \phi_+' \left(\left(a_i^T x^k - b_i \right)^2 \right) \left(a_i^T \left(\tilde{x}^{k+1} - x^k \right) \right) = 0$ *for all* i.

(iii) $\Phi \left((Ax^* - b) \circ (Ax^* - b) \right) = \sigma$ *for every accumulation point* x^* *of* $\{x^k\}_{k \in \mathbb{N}}$.

(iv) *The sequence* $\left\{ \tilde{\lambda}_k \right\}_{k \in \mathbb{N}}$ *is bounded.*

(v) *Every accumulation point of* $\{x^k\}_{k \in \mathbb{N}}$ *is a stationary point of* (3).

Proof We compelte the proof one by one.

(i): If, on the contrary, $\liminf\limits_{k \to \infty} \tilde{\lambda}_k = 0$, then there exists a subsequence $\{k_t\}_{k_t \in \mathbb{N}}$ of $\{k\}_{k \in \mathbb{N}}$ such that $\lim\limits_{t \to \infty} \tilde{\lambda}_{k_t} = 0$. Furthermore, by Proposition 1(i) and (iii), we also have $\lim\limits_{t \to \infty} x_{\mathcal{G}}^{k_t} = x_{\mathcal{G}}^*$ and $\lim\limits_{t \to \infty} \tilde{x}_{\mathcal{G}}^{k_t+1} = x_{\mathcal{G}}^*$. Let $\{\xi^{k_t}\}_{k_t \in \mathbb{N}}$ satisfy

$$\xi^{k_t} \in \omega_{\mathcal{G}}^{k_t} \circ \partial G(\tilde{x}^{k_t+1}) + \tilde{\lambda}_{k_t} A_{k_t}^T \tilde{u}^{k_t+1}. \qquad (25)$$

Then by the upper semicontinuity of ∂G and taking limits as $t \to \infty$ in (25), we have from (20) and Remark 4 that

$$0 \in \Psi_+' \left(x_{\mathcal{G}}^* \right) \circ \partial G(x^*).$$

Thus, by using (16) and the nonnegativity of ψ_+', we further have $x^* = 0$, which means $\lim\limits_{t \to \infty} x^{k_t} = 0$. Since $x^{k_t} \in \mathfrak{F}$ in view of Remark 1 (iii), we have

from the closedness of $\mathfrak{F}$ that $0 \in \mathfrak{F}$, which is a contradiction to Assumption 1(iii). Thus, $\liminf\limits_{k \to \infty} \tilde{\lambda}_k > 0$.

(ii): Using the same argument as in the proof of [26, Theorem 3.1(ii)], we have that for each k, there exists $t_k \in [0, 1]$ such that $\hat{x}^k = t_k x^k + (1 - t_k) A^\dagger b$ satisfies $\lim\limits_{k \to \infty} t_k = 1$ and $\lim\limits_{k \to \infty} \|\hat{x}^k - x^k\| = 0$. This together with

$$0 \le \|\hat{x}_{\mathcal{G}}^k - x_{\mathcal{G}}^k\| \le \|\left(\hat{x}^k - x^k\right)_{\mathcal{G}}\| = \|\hat{x}^k - x^k\|,$$

gives

$$\lim_{k \to \infty} \|\hat{x}_{\mathcal{G}}^k - x_{\mathcal{G}}^k\| = 0. \tag{26}$$

Furthermore, we see from the definition of $\hat{x}^k$ and A_k that

$$\|A_k \hat{x}^k - b^k - \tilde{v}^{k+1}\|^2 - \sigma_k \le 0. \tag{27}$$

We now claim that

$$(\omega_{\mathcal{G}}^k)^T \left(\tilde{x}_{\mathcal{G}}^{k+1} - \hat{x}_{\mathcal{G}}^k\right) \le (\xi^k)^T \left(\tilde{x}^{k+1} - \hat{x}^k\right) - \frac{1}{2}\tilde{\lambda}_k \|A_k \tilde{x}^{k+1} - A_k \hat{x}^k\|^2. \tag{28}$$

In fact, it follows that

$$\begin{aligned}
&(\omega_{\mathcal{G}}^k)^T \tilde{x}_{\mathcal{G}}^{k+1} - (\xi^k)^T \tilde{x}^{k+1} \\
&= (\omega_{\mathcal{G}}^k)^T \tilde{x}_{\mathcal{G}}^{k+1} - (\xi^k)^T \tilde{x}^{k+1} + \frac{1}{2}\tilde{\lambda}_k \left(\|A_k \tilde{x}^{k+1} - b^k - \tilde{v}^{k+1}\|^2 - \sigma_k\right) \\
&\le (\omega_{\mathcal{G}}^k)^T \hat{x}_{\mathcal{G}}^k - (\xi^k)^T \hat{x}^k + \frac{1}{2}\tilde{\lambda}_k \left(\|A_k \hat{x}^k - b^k - \tilde{v}^{k+1}\|^2 - \sigma_k\right) \\
&\quad - \frac{1}{2}\tilde{\lambda}_k \|A_k \tilde{x}^{k+1} - A_k \hat{x}^k\|^2 \\
&\le (\omega_{\mathcal{G}}^k)^T \hat{x}_{\mathcal{G}}^k - (\xi^k)^T \hat{x}^k - \frac{1}{2}\tilde{\lambda}_k \|A_k \tilde{x}^{k+1} - A_k \hat{x}^k\|^2,
\end{aligned}$$

where the equality follows from (24), the first inequality follows from (23) and the second inequality holds because of (27). Rearranging the above inequality yields the assertation.

Now, by the concavity of ψ and the definition of $\{\omega_{\mathcal{G}}^k\}_{k \in \mathbb{N}}$ in (12), we have

$$\begin{aligned}
\Psi\left(x_{\mathcal{G}}^{k+1}\right) &\le \Psi\left(x_{\mathcal{G}}^k\right) + (\omega_{\mathcal{G}}^k)^T \left(x_{\mathcal{G}}^{k+1} - \tilde{x}_{\mathcal{G}}^{k+1} + \tilde{x}_{\mathcal{G}}^{k+1} - \hat{x}_{\mathcal{G}}^k + \hat{x}_{\mathcal{G}}^k - x_{\mathcal{G}}^k\right) \\
&\le \Psi\left(x_{\mathcal{G}}^k\right) + (\omega_{\mathcal{G}}^k)^T \left(x_{\mathcal{G}}^{k+1} - \tilde{x}_{\mathcal{G}}^{k+1}\right) + (\xi^k)^T \left(\tilde{x}^{k+1} - \hat{x}^k\right) \\
&\quad - \frac{1}{2}\tilde{\lambda}_k \|A_k \tilde{x}^{k+1} - A_k \hat{x}^k\|^2 + (\omega_{\mathcal{G}}^k)^T \left(\hat{x}_{\mathcal{G}}^k - x_{\mathcal{G}}^k\right),
\end{aligned}$$

where the last inequality uses (28). Rearranging the above inequality gives that

$$\begin{aligned}
0 &\le \frac{1}{2}\tilde{\lambda}_k \|A_k \tilde{x}^{k+1} - A_k \hat{x}^k\|^2 \\
&\le \Psi\left(x_{\mathcal{G}}^k\right) - \Psi\left(x_{\mathcal{G}}^{k+1}\right) + (\omega_{\mathcal{G}}^k)^T\left(x_{\mathcal{G}}^{k+1} - \tilde{x}_{\mathcal{G}}^{k+1}\right) + (\xi^k)^T\left(\tilde{x}^{k+1} - \hat{x}^k\right) + (\omega_{\mathcal{G}}^k)^T\left(\hat{x}_{\mathcal{G}}^k - x_{\mathcal{G}}^k\right).
\end{aligned} \tag{29}$$

Note that the convergence of $\left\{\Psi(x_{\mathcal{G}}^k)\right\}_{k\in\mathbb{N}}$ (Proposition 1(i)) gives

$$\lim_{k\to\infty}\left(\Psi(x_{\mathcal{G}}^k)-\Psi(x_{\mathcal{G}}^{k+1})\right)=0,$$

while (20) and the boundedness of $\{\tilde{x}^{k+1}\}_{k\in\mathbb{N}}$ and $\{\hat{x}^k\}_{k\in\mathbb{N}}$ imply

$$\lim_{k\to\infty}(\xi^k)^T\left(\tilde{x}^{k+1}-\hat{x}^k\right)=0.$$

Combining these two displays, the boundedness of $\{A_k\}_{k\in\mathbb{N}}$ and $\{\omega_{\mathcal{G}}^k\}_{k\in\mathbb{N}}$, (26), Proposition 1(ii) and Theorem 2(i) with (29), we obtain that

$$\lim_{k\to\infty}\left\|A_k\left(\tilde{x}^{k+1}-\hat{x}^k\right)\right\|=0.$$

By using the boundedness of $\{A_k\}_{k\in\mathbb{N}}$ and (26), we further have

$$\lim_{k\to\infty}\left\|A_k\left(\tilde{x}^{k+1}-x^k\right)\right\|=0.$$

This together with the use of the notations in (12) gives

$$\lim_{k\to\infty}\sqrt{\phi'_+\left(\left(a_i^Tx^k-b_i\right)^2\right)}\left(a_i^T\left(\tilde{x}^{k+1}-x^k\right)\right)=0\quad\forall i.$$

Thus, combined with the boundedness of $\{x^k\}_{k\in\mathbb{N}}$ and the continuity of ϕ'_+, it yields

$$\lim_{k\to\infty}\phi'_+\left(\left(a_i^Tx^k-b_i\right)^2\right)\left(a_i^T\left(\tilde{x}^{k+1}-x^k\right)\right)=0\quad\forall i.$$

(iii): The proof of this part is the same as that of [26, Theorem 3.1(iii)], and thus is omitted herein.

(iv): Suppose, to the contrary, that $\left\{\tilde{\lambda}_k\right\}_{k\in\mathbb{N}}$ is unbounded. Then there exists a subsequence $\left\{\tilde{\lambda}_{k_t}\right\}_{k_t\in\mathbb{N}}$ such that $\lim_{t\to\infty}\tilde{\lambda}_{k_t}=\infty$. In view of the boundedness of $\{x^k\}_{k\in\mathbb{N}}$ that is derived from Proposition 1(i), by passing to a further subsequence if necessary, we can assume $\lim_{t\to\infty}x^{k_t}=x^*$ for some $x^*\in\mathbb{R}^n$ without loss of generality. Now, by use of the first relationship in (21), we obtain

$$\frac{\xi^{k_t}}{\tilde{\lambda}_{k_t}}\in\frac{\omega_{\mathcal{G}}^{k_t}\circ\partial G(\tilde{x}^{k_t+1})}{\tilde{\lambda}_{k_t}}+A_{k_t}^T\tilde{u}^{k_t+1}.\tag{30}$$

Since $\partial G(\cdot)$ is contained in the closed unit ball of $\mathbb{R}^n$ and $\left\{\omega_{\mathcal{G}}^{k_t}\right\}_{k_t\in\mathbb{N}}$ is bounded in view of Remark 4, by passing the limit as $t\to\infty$ on both sides of (30), we conclude from (20) that $\lim_{t\to\infty}A_{k_t}^T\tilde{u}^{k_t+1}=0$. This together with the definition of $\tilde{v}^k$ for all k in (22) implies $\lim_{t\to\infty}\left[A_{k_t}^T\left(A_{k_t}\tilde{x}^{k_t+1}-b^{k_t}\right)-A_{k_t}^T\tilde{v}^{k_t+1}\right]=0$. Since

$\{A_{k_t}\}_{k_t \in \mathbb{N}}$ is bounded and $\lim\limits_{t \to \infty} \tilde{v}^{k_t+1} = 0$ by virtue of Remark 4 and (14) respectively, we further derive

$$\lim_{t \to \infty} A_{k_t}^T \left(A_{k_t} \tilde{x}^{k_t+1} - b^{k_t} \right) = 0. \tag{31}$$

In view of (12), we rewrite (31) as $\lim\limits_{t \to \infty} \sum\limits_{i=1}^{m} \phi'_+ \left(\left(a_i^T x^{k_t} - b_i \right)^2 \right) \left(a_i^T \tilde{x}^{k_t+1} - b_i \right) a_i = 0$. This together with Theorem 2(ii) gives

$$\begin{aligned}
0 &= \lim_{t \to \infty} \sum_{i=1}^{m} \phi'_+ \left(\left(a_i^T x^{k_t} - b_i \right)^2 \right) \left(a_i^T \tilde{x}^{k_t+1} - b_i \right) a_i \\
&= \lim_{t \to \infty} \sum_{i=1}^{m} \phi'_+ \left(\left(a_i^T x^{k_t} - b_i \right)^2 \right) \left(a_i^T x^{k_t} - b_i \right) a_i \\
&= \sum_{i=1}^{m} \phi'_+ \left(\left(a_i^T x^* - b_i \right)^2 \right) \left(a_i^T x^* - b_i \right) a_i.
\end{aligned} \tag{32}$$

However, by Theorem 2(iii), it follows that $\varPhi \left((Ax^* - b) \circ (Ax^* - b) \right) = \sigma$. Together with Fact 2, this implies $\sum\limits_{i=1}^{m} \phi'_+ \left(\left(a_i^T x^* - b_i \right)^2 \right) \left(a_i^T x^* - b_i \right) a_i \neq 0$, which contradicts (32). Thus, the sequence $\left\{ \tilde{\lambda}_k \right\}_{k \in \mathbb{N}}$ is bounded.

(v): Let x^* be any accumulation point of $\{x^k\}_{k \in \mathbb{N}}$. Then there exists a subsequence $\{x^{k_t}\}_{k_t \in \mathbb{N}} \subset \{x^k\}_{k \in \mathbb{N}}$ such that $\lim\limits_{t \to \infty} x^{k_t+1} = x^*$. This together with Proposition 1(ii) shows

$$\lim_{t \to \infty} \tilde{x}^{k_t+1} = x^*. \tag{33}$$

In view of the boundedness of $\{x^k\}_{k \in \mathbb{N}}$ and $\{\bar{\lambda}_k\}_{k \in \mathbb{N}}$, by passing to a subsequence if necessary, we have without loss of generality that

$$\lim_{t \to \infty} x^{k_t} = \bar{x} \tag{34}$$

and $\lim\limits_{t \to \infty} \tilde{\lambda}_{k_t} = \lambda_*$ for some $\bar{x} \in \mathbb{R}^n$ and $\lambda_* \in \mathbb{R}$, respectively. Since x^* is an accumulation point of $\{x^k\}_{k \in \mathbb{N}}$, in view of Theorem 2(i) and (iii), one has that x^* and $\lambda_*/2$ satisfy the former two conditons (6) and (7) in Definition 2. That is, $\varPhi \left((Ax^* - b) \circ (Ax^* - b) \right) \leq \sigma$ and $\lambda_* \left(\varPhi \left((Ax^* - b) \circ (Ax^* - b) \right) - \sigma \right) = 0$. Now, to complete the proof, it suffices to show that $\left(x^*, \frac{\lambda_*}{2} \right)$ also satisfies (8).

In order to do so, we first have from Theorem 2(ii) that for each i,

$$\begin{aligned}
&\lim_{t \to \infty} \phi'_+ \left(\left(a_i^T x^{k_t} - b_i \right)^2 \right) \left(a_i^T \tilde{x}^{k_t+1} - b_i \right) \\
&= \lim_{t \to \infty} \phi'_+ \left(\left(a_i^T x^{k_t} - b_i \right)^2 \right) \left(a_i^T x^{k_t} - b_i \right).
\end{aligned} \tag{35}$$

Define $I = \left\{ i : \phi'_+ \left(\left(a_i^T \bar{x} - b_i \right)^2 \right) > 0 \right\}$. Then for each $i \in I$, it follows from (33)-(35) and the continuity of ϕ'_+ that

$$a_i^T x^* = a_i^T \bar{x}. \tag{36}$$

On the other hand, for each $i \notin I$ (i.e., $\phi'_+ \left(\left(a_i^T \bar{x} - b_i \right)^2 \right) = 0$), we claim that

$$\phi'_+ \left(\left(a_i^T x^* - b_i \right)^2 \right) = 0.$$

In fact, if on the contrary that there exists $i_0 \notin I$ such that $\phi'_+ \left(\left(a_{i_0}^T x^* - b_{i_0} \right)^2 \right) > 0$, then by using the concavity of ϕ, one has that $\left(a_i^T \bar{x} - b_i \right)^2$ is a maximizer of ϕ for all $i \notin I$. Thus, the following two inequalities hold:

$$\phi \left(\left(a_{i_0}^T x^* - b_{i_0} \right)^2 \right) < \phi \left(\left(a_{i_0}^T \bar{x} - b_{i_0} \right)^2 \right), \quad \text{and}$$
$$\phi \left(\left(a_i^T x^* - b_i \right)^2 \right) \leq \phi \left(\left(a_i^T \bar{x} - b_i \right)^2 \right) \ \forall i \notin I \cup \{i_0\}.$$

In view of above display, we further have

$$\Phi \left((Ax^* - b) \circ (Ax^* - b) \right) = \sum_{i=1}^m \phi \left(\left(a_i^T x^* - b_i \right)^2 \right)$$
$$= \phi \left(\left(a_{i_0}^T x^* - b_{i_0} \right)^2 \right) + \sum_{i \neq i_0, i \notin I} \phi \left(\left(a_i^T x^* - b_i \right)^2 \right) + \sum_{i \in I} \phi \left(\left(a_i^T \bar{x} - b_i \right)^2 \right)$$
$$\leq \phi \left(\left(a_{i_0}^T x^* - b_{i_0} \right)^2 \right) + \sum_{i \neq i_0, i \notin I} \phi \left(\left(a_i^T \bar{x} - b_i \right)^2 \right) + \sum_{i \in I} \phi \left(\left(a_i^T \bar{x} - b_i \right)^2 \right)$$
$$< \phi \left(\left(a_{i_0}^T \bar{x} - b_{i_0} \right)^2 \right) + \sum_{i \neq i_0, i \notin I} \phi \left(\left(a_i^T \bar{x} - b_i \right)^2 \right) + \sum_{i \in I} \phi \left(\left(a_i^T \bar{x} - b_i \right)^2 \right)$$
$$= \sum_{i=1}^m \phi \left(\left(a_i^T \bar{x} - b_i \right)^2 \right) = \Phi \left((A\bar{x} - b) \circ (A\bar{x} - b) \right),$$

where the second equality follows from (36). However, since x^* and $\bar{x}$ are both accumulation points of $\{x^k\}$, the above display contadicts

$$\Phi \left((Ax^* - b) \circ (Ax^* - b) \right) = \Phi \left((A\bar{x} - b) \circ (A\bar{x} - b) \right) = \sigma$$

which is deduced from Theorem 2(iii). Therefore, we obtain the claim and conclude that

$$\begin{cases} a_i^T x^* = a_i^T \bar{x}, & \text{if } \phi'_+ \left(\left(a_i^T \bar{x} - b_i \right)^2 \right) > 0, \\ \phi'_+ \left(\left(a_i^T \bar{x} - b_i \right)^2 \right) = \phi'_+ \left(\left(a_i^T x^* - b_i \right)^2 \right) = 0, & \text{if } \phi'_+ \left(\left(a_i^T \bar{x} - b_i \right)^2 \right) = 0. \end{cases} \tag{37}$$

Now, with (21), the definition of $\tilde{v}^{k+1}$ in (22) and the notations in (12), one has

$$
\begin{aligned}
\xi^{k_t} &\in \omega_{\mathcal{G}}^{k_t} \circ \partial G(\tilde{x}^{k_t+1}) + \tilde{\lambda}_{k_t} A_{k_t}^T \tilde{u}^{k_t+1} \\
&= \Psi'_+(x_{\mathcal{G}}^{k_t}) \circ \partial G(\tilde{x}^{k_t+1}) + \tilde{\lambda}_{k_t} A_{k_t}^T \left(A_{k_t} \tilde{x}^{k_t+1} - b^{k_t} - \tilde{v}^{k_t+1} \right) \\
&= \Psi'_+ \left(\tilde{x}_{\mathcal{G}}^{k_t+1} + x_{\mathcal{G}}^{k_t} - \tilde{x}_{\mathcal{G}}^{k_t+1} \right) \circ \partial G(\tilde{x}^{k_t+1}) \\
&\quad + \tilde{\lambda}_{k_t} \left(\sum_{i=1}^m \phi'_+ \left(\left(a_i^T x^{k_t} - b_i \right)^2 \right) \left(a_i^T \tilde{x}^{k_t+1} - b_i \right) a_i - A_{k_t}^T \tilde{v}^{k_t+1} \right).
\end{aligned}
$$

Recall (33), (34) and $\lim_{t \to \infty} \tilde{\lambda}_{k_t} = \lambda_*$. Then taking the limit as $t \to \infty$ on the leftmost and rightmost sides of above display, we obtain from Proposition 1(iii), (37), the boundedness of $\{A_k\}_{k \in \mathbb{N}}$, and the definition of $\tilde{v}^{k+1}$ that

$$
0 \in \Psi'_+(x_{\mathcal{G}}^*) \circ \partial G(x^*) + \lambda_* \sum_{i=1}^m \phi'_+ \left(\left(a_i^T x^* - b_i \right)^2 \right) \left(a_i^T x^* - b_i \right) a_i
$$

This implies that (8) holds when (x, λ) is replaced by $\left(x^*, \frac{\lambda_*}{2} \right)$. The proof is therefore completed.

5 Numerical experiments

In this section, we show the performance of Algorithm **GIR** via numerical experiments. Our codes are written and implemented in MATLAB 2019b, and all numerical experiments are performed on a 64-bit PC equipped with an Intel(R) Xeon(R) Silver 4210 CPU (2.20GHz) and 32GB of RAM.

We first illustrate the notations in Tables 3-5. The minimum and maximum values of the residuals $= (\Psi(\vartheta_{\mathcal{G}}^k) - \sigma)/\sigma$ (Res$_{\min}$ and Res$_{\max}$, respectively) are recorded, where ϑ^k denotes the approximate sparse solution generated by the corresponding algorithm. The objective function value of the recovered vector is recorded as Fval. We also record three kinds of data refer to the instances being "successfully solved": the average number of inner iterations (Iter$_s$), the average CPU time (CPU$_s$), and the average recovery error (RecErr$_s$). If instances that fail to meet "successfully solved" are recorded as Iter$_f$, CPU$_f$ and RecErr$_f$. We call a random instance is "successfully solved" if the recovery error meets the criterion $\frac{\|\vartheta^k - x_{\text{orig}}\|}{\max\{\|x_{\text{orig}}\|, 1\}} \leq 0.01$, in which x_{orig} is generated through MATLAB commands in the following Subsection 5.1.

5.1 Importance of new problem and algorithm

The introduction of a group sparsity structure elevates the importance of problem (3) and Algorithms 1 and 2, named as **GIR**$_{\text{ADMM}}$, which form a central focus of this subsection.

We consider the following constrained optimization problem:

$$\min_{x\in\mathbb{R}^n} \sum_{i=1}^{q} \log\left(1 + \|x_{\mathcal{G}_i}\|/\epsilon\right), \quad \text{s.t.} \quad \sum_{i=1}^{m} \log\left(1 + (a_i^T x - b_i)^2/\delta^2\right) \leq \sigma, \quad (38)$$

where $m \ll n$, $\epsilon > 0$, $\delta > 0$, $b \in \mathbb{R}^m$ and $\sigma > 0$. The row vectors $a_1, a_2, \cdots, a_m$ of matrix A are linearly independent. Here, ψ and ϕ correspond to the log penalty function and Cauchy loss function, respectively. We can see that this model incorporates a log penalty function [4] in the objective function, which can effectively induce variable sparsity while enhancing model interpretability, reducing complexity, and stabilizing the optimization process. By integrating the Cauchy loss function [18] into the constraints, it significantly improves the model's robustness and strengthens resistance to outliers and noise.

We note that (38) is a special case of problem (3), which satisfies Assumptions 1. Thus, every accumulation point of the generated sequence $\{x^k\}_{k\in\mathbb{N}}$ by Algorithm 1 is a stationary point, as guaranteed by Theorem 2.

For the numerical comparsion, we generate an $m \times n$ matrix A with independent and identically distributed (i.i.d.) standard Gaussian entries. Next, we generate an original block-sparse signal $x_{\text{orig}} \in \mathbb{R}^n$ using the following MATLAB commands:

```
I = randperm(n/J); I = I(s+1 : end);
x0 = randn(J,n/J); x0(:, I) = 0;
x_{orig} = reshape(x0, n, 1);
```

Here, J denotes the size of each block, and s is the number of nonzero blocks. We further define $\eta = 0.005 * \text{randn}(m, 1)$, where the noise entries follow an i.i.d standard Cauchy distribution, and set $b = Ax_{\text{orig}} + \eta$. Finally, we specify $J = 2$, $\epsilon = 0.1$ and $\sigma = 1.2 \sum_{i=1}^{m} \log\left(1 + (\eta_i)^2/\delta^2\right)$ with $\delta = 0.05$.

In the numerical experiments, we set the three-element tuple $(m, n, s) = (540i, 2560i, 80i)$ for $i \in \{2, 4, 6, 8, 10\}$. Based on the generation of x_{orig}, we have $\text{nnz}(x_{\text{orig}}) = 80 * 2i$. For each i, 30 random instances are generated by the above commands. We report the mean value of L (L), the average CPU time for computing L (Time$_L$), the average CPU time for computing $A^\dagger b$ (Time$_{\text{slater}}$), and the average CPU time for performing QR decomposition of A^T (Time$_{\text{QR}}$) in Table 1.

Table 1 Record of average value and average CPU time

(m, n, s)	L	Time$_L$	Time$_{\text{QR}}$	Time$_{\text{slater}}$
$(1080, 5120, 160)$	1.08e+04	0.2	0.4	0.0
$(2160, 10240, 320)$	2.18e+04	1.0	1.3	0.0
$(3240, 15360, 480)$	3.26e+04	2.5	2.9	0.0
$(4320, 20480, 640)$	4.35e+04	5.2	5.2	0.1
$(5400, 25600, 800)$	5.44e+04	6.6	8.5	0.1

Note that if $q = n$ in problem (38), then it reduces to

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^n \log\left(1 + |x_i|/\epsilon\right), \quad \text{s.t.} \quad \sum_{i=1}^m \log\left(1 + (a_i^T x - b_i)^2/\delta^2\right) \leq \sigma. \qquad (39)$$

This problem is a special case of problem (2) without group structure, which is from [26]. Sun and Pong [26] also propose the double iterative reweighted algorithm to solve problem (2), whose subproblem is also can be solved by ADMM. These strategies are named as $\mathbf{IR}_{\text{ADMM}}$ for short.

For $\mathbf{GIR}_{\text{ADMM}}$ and $\mathbf{IR}_{\text{ADMM}}$, we have the following common settings. The initial point $x^0 = A^\dagger b$ is computed using the following MATLAB commands:

```
[Q,R] = qr(A',0); xfeas = Q*(R'\b);
```

We set $\tau_k = \max\left\{5^{-k-1}, 10^{-8}\right\}$ and $\mu_k = \max\left\{1.2^{-k-1}, 10^{-8}\right\}$ for Algorithm 1. The outer iteration terminates when the criterion $\frac{\|x^{k+1} - x^k\|}{\max\{\|x^k\|, 1\}} \leq 10^{-4}$ is met.

The other parameters of $\mathbf{IR}_{\text{ADMM}}$ for its inner problem is what the reference [26] proposed. In the implementation of $\mathbf{GIR}_{\text{ADMM}}$, at each iteration k, we set $\beta = \bar{L}^{-\frac{1}{2}}$, $\gamma = \frac{0.99(1+\sqrt{5})}{2}$, ρ and $\bar{L}$ as in Remark 2 for Algorithm 2. Note that $\mathbf{IR}_{\text{ADMM}}$ use the standard ℓ_2-norm square as distance function. Thus, we choose f in Definition 1 to be $f(x) = \|x\|^2$. This implies (13) equals to

$$\text{dist}\left(0, \partial\left(\omega_\mathcal{G}^k \circ G\right)(\tilde{x}^{k+1}) + A_k^T N_{U^k}(\tilde{u}^{k+1})\right) \leq \sqrt{\epsilon_k},$$

where $\epsilon_k = \min\left\{\bar{\epsilon}_k, \tau_k \bar{\varGamma}_l, \gamma\beta\bar{\epsilon}_k, \gamma\beta\tau_k(\|z^{k,l+1}\| + 1)\right\}$.

The numerical results are reported in Tables 2 and 3. In Table 2, Success(%) and nnz(%) are used to denote the success rate and the percentage of sparsity successfully recovered, respectively. From the value of nnz, we can see that problem (38) can obtain more sparser solution point than problem (39).

Table 2 Vector sparsity recovery success rate and recovery degree proportion

Dimension and sparsity	$\mathbf{GIR}_{\text{ADMM}}$		$\mathbf{IR}_{\text{ADMM}}$ in [26]	
(m, n, s)	Success(%)	nnz(%)	Success(%)	nnz(%)
$(1080, 5120, 160)$	100	100	100	63
$(2160, 10240, 320)$	100	100	100	17
$(3240, 15360, 480)$	100	100	100	3
$(4320, 20480, 640)$	100	100	100	20
$(5400, 25600, 800)$	100	100	100	0

We present the more comparation on $\mathbf{GIR}_{\text{ADMM}}$ and $\mathbf{IR}_{\text{ADMM}}$ in Table 3. It can be observed that $\mathbf{GIR}_{\text{ADMM}}$ consistently requires less CPU time and a smaller value of Fval than $\mathbf{IR}_{\text{ADMM}}$. The number of iterations needed for Algorithm $\mathbf{GIR}_{\text{ADMM}}$ is fewer than for $\mathbf{IR}_{\text{ADMM}}$. The results indicate that $\mathbf{GIR}_{\text{ADMM}}$ outperforms $\mathbf{IR}_{\text{ADMM}}$ in preserving the sparsity of the original vector across all cases.

Table 3 Numerical comparison of $\mathbf{GIR}_{\mathrm{ADMM}}$ and $\mathbf{IR}_{\mathrm{ADMM}}$

Dimension and sparsity	$\mathbf{GIR}_{\mathrm{ADMM}}$						$\mathbf{IR}_{\mathrm{ADMM}}$ in [26]					
(m, n, s)	Iter_s	CPU_s	Fval	RecErr_s	$\mathrm{Res}_{\min}$	$\mathrm{Res}_{\max}$	Iter_s	CPU_s	Fval	RecErr_s	$\mathrm{Res}_{\min}$	$\mathrm{Res}_{\max}$
$(1080, 5120, 160)$	1576	10.2	4.0e+02	3.2e-04	-3.9e-02	-2.2e-03	2835	19.4	6.2e+02	3.3e-04	-5.3e-03	6.0e-04
$(2160, 10240, 320)$	2231	51.8	7.9e+02	2.4e-04	-6.8e-03	-1.4e-05	3812	97.0	1.2e+03	2.4e-04	-1.5e-03	1.7e-05
$(3240, 15360, 480)$	2870	143.7	1.2e+03	1.9e-04	-2.6e-03	-3.6e-06	4635	246.1	1.9e+03	1.9e-04	-6.0e-04	-6.3e-06
$(4320, 20480, 640)$	3422	298.2	1.6e+03	1.7e-04	-1.8e-03	-2.2e-05	5400	525.4	2.5e+03	1.7e-04	-3.2e-04	-7.1e-06
$(5400, 25600, 800)$	3939	519.7	2.0e+03	1.5e-04	-1.0e-03	-1.1e-05	6212	947.6	3.1e+03	1.5e-04	-3.3e-04	-4.4e-06

5.2 Algorithm comparision

In this subsection, we first apply $\mathbf{GIR}_{\mathrm{ADMM}}$ and $\mathbf{SPA}$ to solve the following gourp sparse optimization problem

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{q} \log\left(1 + \|x_{\mathcal{G}_i}\|/\epsilon\right), \quad \text{s.t.} \quad \|Ax - b\| \leq \sqrt{\sigma}. \tag{40}$$

This problem is a special case of both (3) and (1). It reduces to (3) when $\phi = I$, and to (1) because its objective function can be reformulated as a capped folded concave function (see Appendix for details on this transformation and the proximal operator computation). Therefore, the $\mathbf{SPA}$ method can be applied to solve it.

In the implementation, we use the same way to generate A, b, σ in the Subsection 5.1, and we set $\epsilon = 0.3$. For $\mathbf{SPA}$, we set the parameter as $\nu = \epsilon\left(e^C - 1\right)$, $n_i = J = 2$, $x^0 = 1_n$, $\lambda_0 = 40$, $\mu_0 = \epsilon_0 = 1$, $\rho = 3$, $\theta = \frac{1}{\rho}$, $M = 3$, where ν and C are defined in Remark 5. We set the feasible point the same as that in $\mathbf{GIR}_{\mathrm{ADMM}}$. The entire algorithm that we set terminates when

$$\max\left\{\left(\|Ax - b\|^2 - \sigma\right)_+, 0.01\epsilon_k\right\} \leq 10^{-6}$$

is satisfied.

The corresponding numerical reuslts are presented in Table 4. Here we can see that although $\mathbf{SPA}$ requires less CPU time than $\mathbf{GIR}_{\mathrm{ADMM}}$ to solve the problem (40), it underperforms $\mathbf{GIR}_{\mathrm{ADMM}}$ in vector recovery, sparsity preservation and the value of Fval.

Table 4 The first numerical records of $\mathbf{GIR}_{\mathrm{ADMM}}$ and $\mathbf{SPA}$

Dimension and sparsity	$\mathbf{GIR}_{\mathrm{ADMM}}$						$\mathbf{SPA}$ in [21]					
(m, n, s)	Iter_s	CPU_s	Fval	RecErr_s	$\mathrm{Res}_{\min}$	$\mathrm{Res}_{\max}$	Iter_f	CPU_f	Fval	RecErr_f	$\mathrm{Res}_{\min}$	$\mathrm{Res}_{\max}$
$(1080, 5120, 160)$	1848	11.5	2.5e+02	3.3e-04	-10.0e-01	-4.7e+00	47	3.0	9.3e+02	8.9e-01	-4.0e-05	4.2e-7
$(2160, 10240, 320)$	2144	47.9	5.0e+02	2.4e-04	-10.0e-01	4.3e+00	32	6.5	1.9e+03	8.9e-01	-5.1e-05	1.9e-07
$(3240, 15360, 480)$	2137	102.9	7.5e+02	1.9e-04	-9.8e-01	9.1e+00	32	14.0	2.8e+03	8.9e-01	-1.9e-04	4.4e-07
$(4320, 20480, 640)$	1885	159.2	9.9e+02	1.7e-04	-10.0e-01	8.6e-01	30	23.0	3.7e+03	8.9e-01	-1.5e-05	4.0e-07
$(5400, 25600, 800)$	2020	260.7	1.2e+03	1.5e-04	-10.0e-01	3.8e+00	29	34.0	4.6e+03	8.9e-01	-1.9e-05	2.7e-07

Next, we conduct a comparative numerical study to evaluate the performance of the proposed $\mathbf{GIR}_{\mathrm{ADMM}}$ against $\mathbf{SPA}$ in solving problem (38), which is

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{q} \log\left(1 + \|x_{\mathcal{G}_i}\|/\epsilon\right), \quad \text{s.t.} \quad \sum_{i=1}^{m} \log\left(1 + (a_i^T x - b_i)^2/\delta^2\right) \leq \sigma. \quad (41)$$

The objective of this experiment is to benchmark our approach against a recognized method, assessing key performance indicators such as solution accuracy, and computational stability.

In the implementation, we also use the same way to generate A, b, σ as that in Subsection 5.1. The setting of $\mathbf{GIR}_{\mathrm{ADMM}}$ is also what Subsection 5.1 proposed. The parameter settings of $\mathbf{SPA}$ are the same as those used when solving (40). Moreover, we set the stopping criterion of $\mathbf{SPA}$ as

$$\max\left\{\left(\sum_{i=1}^{m} \log\left(1 + (a_i^T x - b_i)^2/\delta^2\right) - \sigma\right)_+, 0.01\epsilon_k\right\} \leq 10^{-6}.$$

Numerical results show that, for our Algorithm $\mathbf{GIR}_{\mathrm{ADMM}}$, the Success(%) and nnz(%) recorded still satisfy 100%, while $\mathbf{SPA}$ still fails to achieve recovery and cannot guarantee sparsity. The other experimental data are recorded in Table 5, which also demonstrates the excellent performance of $\mathbf{GIR}_{\mathrm{ADMM}}$.

Table 5 The second numerical records of $\mathbf{GIR}_{\mathrm{ADMM}}$ and $\mathbf{SPA}$

Dimension and sparsity	$\mathbf{GIR}_{\mathrm{ADMM}}$						$\mathbf{SPA}$ in [21]					
(m, n, s)	$\mathrm{Iter_s}$	$\mathrm{CPU_s}$	Fval	$\mathrm{RecErr_s}$	$\mathrm{Res_{min}}$	$\mathrm{Res_{max}}$	$\mathrm{Iter_f}$	$\mathrm{CPU_f}$	Fval	$\mathrm{RecErr_f}$	$\mathrm{Res_{min}}$	$\mathrm{Res_{max}}$
$(1080, 5120, 160)$	2431	15.7	2.5e+02	3.4e-04	-5.0e-03	-2.5e-05	23	1.5	9.3e+02	8.9e-01	-8.1e-06	-3.5e-11
$(2160, 10240, 320)$	3497	80.1	4.9e+02	2.4e-04	-5.3e-04	7.6e-08	22	4.6	1.9e+03	8.9e-01	-9.2e-06	-2.4e-11
$(3240, 15360, 480)$	4425	221.6	7.5e+02	1.9e-04	-2.0e-04	-2.3e-05	23	9.9	2.8e+03	8.9e-01	-1.2e-05	-4.8e-10
$(4320, 20480, 640)$	5207	451.6	9.9e+02	1.6e-04	-2.6e-03	-2.2e-07	22	16.0	3.7e+03	8.9e-01	-1.5e-05	-9.5e-10
$(5400, 25600, 800)$	5945	781.0	1.2e+03	1.5e-04	-3.1e-04	-2.5e-07	22	26.0	4.6e+03	8.9e-01	-1.8e-05	-7.3e-12

6 Conclusion

In this work, we propose the first application of the iteratively reweighted algorithm to solve a class of group sparse optimization problems. Both the objective function and the constraint of our problem involve non-convex functions, which broadens the scope of group sparse-related research models and enhances the robustness of the model. By appropriately reformulating the problem, the non-convex problem is transformed into a sequence of convex subproblems, which are then approximated using the well-established convex solver ADMM. In the algorithm design, we design the termination criteria of subproblem solver by using Bregman distance rather than European distance.

Furthermore, the update rule for the next iteration point such that all iteration points generated by the proposed algorithm remain within the feasible region.

However, due to the specific structure of the constructed subproblems, it is challenging to ensure that $x^{k+1} - x^k \to 0$. Additionally, analyzing the convergence rate of the algorithm for our research remains non-trivial. Consequently, future research directions may include: (1) analyzing the convergence rate of the proposed algorithm, (2) extending the application of the algorithm to practical, concrete scenarios, and (3) design alternative algorithms for solving the problem model in this paper, leveraging the explicit expression of the proximity operator of the log-sum penalty function for algorithm development.

Appendix

Inspired by [37, Section 6.1], we first establish that problems (40) and (41) are equivalent to an optimization problem with an additionally imposed bounded feasible set.

Lemma 1 *Under Assumption 1, the solution set of problems* (40) *and* (41) *are bounded.*

Proof Note that $A^\dagger b$ is a feasible point of problems (40) and (41). In view of the definition of G in Section 1, we set $\tilde{x}_\mathcal{G} = G(A^\dagger b)$ and define $C = \sum\limits_{i=1}^{q} \log\left(1 + \frac{(\tilde{x}_\mathcal{G})_i}{\epsilon}\right)$. The level-boundedness of the objective function in problems (40) and (41) implies that its solution set (denoted by S) is nonempty. For any $x \in S$, the following inequality holds:

$$\sum_{i=1}^{q} \log\left(1 + \frac{(x_\mathcal{G})_i}{\epsilon}\right) \leq \sum_{i=1}^{q} \log\left(1 + \frac{(\tilde{x}_\mathcal{G})_i}{\epsilon}\right).$$

Given that $\log\left(1 + \frac{(\cdot)}{\epsilon}\right)$ is nondecreasing over $\mathbb{R}_+$, we obtain:

$$C \geq \sum_{i=1}^{q} \log\left(1 + \frac{(x_\mathcal{G})_i}{\epsilon}\right) \geq \max_{1 \leq i \leq q} \log\left(1 + \frac{(x_\mathcal{G})_i}{\epsilon}\right) = \log\left(1 + \frac{\|x_\mathcal{G}\|_\infty}{\epsilon}\right).$$

This implies $\|x_\mathcal{G}\|_\infty \leq \epsilon\left(e^C - 1\right)$. The proof is completed by invoking the definition of $x_\mathcal{G}$.

Remark 5 Lemma 1 indicates $\sum\limits_{i=1}^{q} \log\left(1 + \|x_{\mathcal{G}_i}\|/\epsilon\right)$ is equivalent to the following form:

$$\sum_{i=1}^{q} \psi^{\mathrm{CapLog}}((x_\mathcal{G})_i) \tag{42}$$

where ψ^{CapLog} is the capped function defined as:

$$\psi^{\mathrm{CapLog}}(t) = \begin{cases} \log\left(1 + \frac{|t|}{\epsilon}\right), & \text{if } |t| \in [0, \nu), \\ C, & \text{if } |t| \in [\nu, \infty), \end{cases} \qquad (43)$$

with $\nu = \epsilon\left(e^C - 1\right)$, $C = \sum_{i=1}^{q} \log\left(1 + \frac{(\tilde{x}_{\mathcal{G}})_i}{\epsilon}\right)$ and $\tilde{x}_{\mathcal{G}} = G(A^\dagger b)$.

Our next goal is to derive the closed-form expression of the proximal mapping for the objective function in the reformulated problem (42). We first present the following lemma:

Lemma 2 *Let C be defined as in* (43). *Define $f : \mathbb{R}^n \to \mathbb{R}$ to be*

$$f(x) = \begin{cases} \log\left(1 + \frac{\|x\|}{\epsilon}\right), & \text{if } \|x\| \in [0, \nu), \\ C, & \text{if } \|x\| \in [\nu, \infty), \end{cases}$$

where $\epsilon > 0$ is a given parameter. The proximal mapping of f with scaling parameter $\lambda > 0$ at a point $x \in \mathbb{R}^n$ is:

$$\mathrm{prox}_{\lambda f}(x) = \begin{cases} 0, & \text{if } \|x\| = 0, \\ \frac{x}{\|x\|} \mathrm{prox}_{\lambda\psi^{\mathrm{CapLog}}}(\|x\|), & \text{otherwise,} \end{cases} \qquad (44)$$

where ψ^{CapLog} is defined in (43).

Proof If $x = 0$, the definition of the proximal mapping (given in (4)) directly implies $\mathrm{prox}_{\lambda f}(x) = 0$.

For $x \neq 0$, we use the definition of the proximal mapping in (4) to obtain:

$$\begin{aligned} \mathrm{prox}_{\lambda f}(x) &= \argmin_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{\|x - y\|^2}{2\lambda} \right\} \\ &= \argmin_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{\|x\|^2 + \|y\|^2 - 2\langle x, y\rangle}{2\lambda} \right\} \\ &= \frac{x}{\|x\|} \argmin_{\|y\| \in \mathbb{R}_+} \left\{ \psi^{\mathrm{CapLog}}(\|y\|) + \frac{(\|x\| - \|y\|)^2}{2\lambda} \right\} \\ &= \frac{x}{\|x\|} \argmin_{\omega \in \mathbb{R}_+} \left\{ \psi^{\mathrm{CapLog}}(\omega) + \frac{(\|x\| - \omega)^2}{2\lambda} \right\} \\ &= \frac{x}{\|x\|} \argmin_{\tilde{\omega} \in \mathbb{R}} \left\{ \psi^{\mathrm{CapLog}}(\tilde{\omega}) + \frac{(\|x\| - \tilde{\omega})^2}{2\lambda} \right\} = \frac{x}{\|x\|} \mathrm{prox}_{\lambda\psi^{\mathrm{CapLog}}}(\|x\|), \end{aligned}$$

where the third equality holds because the minimizer of the optimization problem in the RHS of the second equation is achieved only if the vectors x and y are in the same direction, the fifth equality follows from the same reason as that in the third one. This completes the proof.

Next, we present the closed form of $\text{prox}_{\lambda f}(x)$ (from (44)) for the case $\|x\| \neq 0$, which is established in the following lemma:

Lemma 3 *Consider the function f in Lemma 2 with $x \neq 0$, and ψ in (43). Define $f^{\mathrm{C-log}} : \mathbb{R}_+ \to \mathbb{R}$ as*

$$
f^{\mathrm{C-log}}(u) = \begin{cases} \frac{1}{2}(u - \|x\|)^2 + \lambda \log\left(1 + \frac{u}{\epsilon}\right), & \text{if } 0 \leq u < \nu, \\ \frac{1}{2}(u - \|x\|)^2 + \lambda C, & \text{if } u \geq \nu. \end{cases}
$$

Then the proximal mapping of f at x with the scaling parameter $\lambda > 0$ is

$$
\text{prox}_{\lambda f}(x) = \begin{cases} u_1^*(\|x\|)\frac{x}{\|x\|}, & \text{if } f^{\mathrm{C-log}}\left(u_1^*(\|x\|)\right) \leq f^{\mathrm{C-log}}\left(u_2^*(\|x\|)\right), \\ u_2^*(\|x\|)\frac{x}{\|x\|}, & \text{otherwise,} \end{cases} \tag{45}
$$

where $u_1^*(t) = \min\left\{\left(\text{prox}_{\lambda g}(t)\right)_+, \nu\right\}$, $u_2^*(t) = \max\{t, \nu\}$ for all $t \in \mathbb{R}$, and $g(t) = \log\left(1 + \frac{|t|}{\epsilon}\right)$ for all $t \in \mathbb{R}$.

Proof In view of the definition of proximal mapping, one has

$$
\text{prox}_{\lambda \psi^{\mathrm{CapLog}}}(\|x\|) = \underset{u \in \mathbb{R}_+}{\arg\min} f^{\mathrm{C-log}}(u).
$$

For $0 \leq u < \nu$, we have $\text{prox}_{\lambda \psi^{\mathrm{CapLog}}}(\|x\|) = \left(\text{prox}_{\lambda g}(x)\right)_+$; for $u \geq \nu$, $\text{prox}_{\lambda \psi^{\mathrm{CapLog}}}(\|x\|) = \|x\|$. Combining this with the analysis in [21, Section A.2], the minimizer of $f^{\mathrm{C-log}}$ over $[0, \nu]$ is $u_1^*(x) = \min\left\{\left(\text{prox}_{\lambda g}(x)\right)_+, \nu\right\}$, and the minimizer over $[\nu, \infty)$ is $u_2^*(x) = \max\{x, \nu\}$. This directly implies that (45) holds, completing the proof.

Remark 6 Consider problem (42). Let

$$
\Psi^{\mathrm{CapLog}}(x) := \sum_{i=1}^q \psi^{\mathrm{CapLog}}(x_{\mathcal{G}_i}) = \sum_{i=1}^q f(x_{\mathcal{G}_i}).
$$

Then it follows that: $\Psi^{\mathrm{CapLog}}(x) = \sum_{i=1}^q f(x_{\mathcal{G}_i})$, with f defined as in Lemma 2. In view of the separability of Ψ^{CapLog} with respect to i, we follow the similar argument to both the proof of [22, Theorem 3] and the discussion in [23, Section 5.1] to derive:

$$
\text{prox}_{\lambda \Psi^{\mathrm{CapLog}}}(x) = \text{prox}_{\lambda f}(x_{\mathcal{G}_1}) \times \text{prox}_{\lambda f}(x_{\mathcal{G}_2}) \times \cdots \times \text{prox}_{\lambda f}(x_{\mathcal{G}_q}),
$$

where $\text{prox}_{\lambda f}$ is established in (44) and (45).

Declarations

- Conflict of interest/Competing interests: Not Applicable.

References

1. Abernethy, J., Bach, F., Evgeniou, T., Vert, J.P.: A new approach to collaborative filtering: operator estimation with spectral regularization. *Journal of Machine Learning Research,* 10(3), 803-826 (2009)
2. Banerjee, A., Merugu, S., Dhillon, I.S., Ghosh, J.: Clustering with Bregman divergences. Journal of machine learning research. *Journal of Machine Learning Research,* 6, 1705-1749 (2005)
3. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory,* 52(2), 489-509 (2006)
4. Chen, X., Lu, Z., Pong, T.K.: Penalty methods for a class of non-Lipschitz optimization problems. *SIAM Journal on Optimization,* 26(3), 1465-1492 (2016)
5. Chen, P.Y., Selesnick, I.W.: Group-sparse signal denoising: non-convex regularization, convex optimization. *IEEE Transactions on Signal Processing,* 62(13), 3464-3478 (2014)
6. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing,* 15(12), 3736-3745 (2006)
7. Fan, J., Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association,* 96(456), 1348-1360 (2001)
8. Fazel, M., Pong, T.K., Sun, D., Tseng, P.: Hankel matrix rank minimization with applications to system identification and realization. *SIAM Journal on Matrix Analysis and Applications,* 34(3), 946-977 (2013)
9. Hu, Y., Li, C., Meng, K., Qin, J., Yang, X.: Group sparse optimization via $\ell_{p,q}$ regularization. *Journal of Machine Learning Research,* 18(30), 1-52 (2017)
10. Huber, P.J.: Robust estimation of a location parameter. *The Annals of Mathematical Statistics,* 35(1), 73-101 (1964)
11. Huang, J., Zhang, T.: The benefit of group sparsity. *Institute of Mathematical Statistics,* 38(4), 1978-2004 (2010)
12. Huang, J., Breheny, P., Ma, S.: A selective review of group selection in high dimensional models. *Statistical Science,* 27(4), 481-499 (2012)
13. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. *Springer, New York* (2009)
14. Jiang, L., Huang, Z., Chen, Y., Zhu, W.: Iterative-weighted thresholding method for group-sparsity-constrained optimization with applications. *IEEE Transactions on Neural Networks and Learning Systems,* 36(6), 11602-11616 (2025)
15. Jiang, D.: Concave selection in generalized linear models. *Doctoral Dissertation, University of Iowa* (2012)
16. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. *Computer,* 42(8), 30-37 (2009)
17. Li, J., So, A.M.-C., Ma, W.K.: Understanding notions of stationarity in nonsmooth optimization. *IEEE Signal Processing Magazine,* 37(5), 18-31 (2020)
18. Li, X., Lu, Q., Dong, Y., Tao, D.: Robust subspace clustering by Cauchy loss function. *IEEE Transactions on Neural Networks and Learning Systems,* 30(7), 2067-2078 (2018)
19. Mosteller, F., Tukey, J.W.: Data Analysis and Regression: A Second Course in Statistics. *Addison Wesley, Sydney* (1977)
20. Peng, D., Chen, X.: Computation of second-order directional stationary points for group sparse optimization. *Optimization Methods and Software,* 35(2), 348-376 (2019)
21. Pan, L., Chen, X.: Group sparse optimization for images recovery using capped folded concave functions. *SIAM Journal on Imaging Sciences,* 14(1), 1-25 (2021)
22. Prater-Bennette, A., Shen, L., Tripp, E.E.: The proximity operator of the log-sum penalty. *Journal of Scientific Computing,* 93(3), 67 (2022)
23. Parikh, N., Boyd, S.: Proximal algorithms. *Foundations and Trends in Optimization,* 1(3), 127-239 (2014)
24. Qin, J., Hu, Y., Yao, J.C., Leung, R.W.T., Zhou, Y., Qin, Y., Wang, J.: Cell fate conversion prediction by group sparse optimization method utilizing single-cell and bulk OMICs data. *Briefings in Bioinformatics,* 22(6), bbab311 (2021)
25. Rockafellar, R.T., Wets, R.J.: Variational Analysis. *Springer Berlin, Heidelberg* (1998)

26. Sun, S., Pong, T.K.: Doubly iteratively reweighted algorithm for constrained compressed sensing models. *Computational Optimization and Applications,* 85(2), 583-619 (2023)
27. Schapire, R.E.: The strength of weak learnability. *Machine learning,* 5(2), 197-227 (1990)
28. Solodov, M.V., Svaiter, B.F.: An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. *Mathematics of Operations Research,* 25(2), 214-230 (2000)
29. Schuster, K., Trettner, P., Kobbelt, L.: High-performance image filters via sparse approximations. *Association for Computing Machinery,* 3(2), 1-19 (2020)
30. Simon, N., Friedman, J., Hastie, T., Tibshirani, R.: A sparse-group lasso. *Journal of Computational and Graphical Statistics,* 22(2), 231-245 (2013)
31. Tu, K., Zhang, H., Gao, H., Feng, J.: A hybrid Bregman alternating direction method of multipliers for the linearly constrained difference-of-convex problems. *Journal of Global Optimization,* 76(4), 665-693 (2020)
32. Tang, G., Nehorai, A.: Computable performance bounds on sparse recovery. *IEEE Transactions on Signal Processing,* 63(1), 132-141 (2015)
33. Tan, Z., Eldar, Y.C., Nehorai, A.: Direction of arrival estimation using co-prime arrays: a super resolution viewpoint. *IEEE Transactions on Signal Processing,* 62(21), 5565-5576 (2014)
34. Wang, L., Chen, G., Li, H.: Group SCAD regression analysis for microarray time course gene expression data. *Bioinformatics,* 23(12), 1486-1494 (2007)
35. Yuan, M., Lin, Y.: Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology,* 68(1), 49-67 (2005)
36. Zhang, C.H.: Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics,* 38(2), 894-942 (2010)
37. Zhang, Y., Li, G., Pong, T.K., Xu, S.: Retraction-based first-order feasible methods for difference-of-convex programs with smooth inequality and simple geometric constraints. *Advances in Computational Mathematics,* 49(1), 8 (2023)
38. Zha, Z., Wen, B., Yuan, X., Zhou, J., Zhu, C.: Image restoration via reconciliation of group sparsity and low-rank models. *IEEE Transactions on Image Processing,* 30, 5223-5238 (2021)
39. Zhang, F., Yang, Z., Chen, Y., Yang, J., Yang, G.: Matrix completion via capped nuclear norm. *IET Image Processing,* 12(6), 959-966 (2018)
40. Zhang, X., Peng, D.: Solving constrained nonsmooth group sparse optimization via group Capped-ℓ_1 relaxation and group smoothing proximal gradient algorithm. *Computational Optimization and Applications,* 83(3), 801-844 (2022)