MATHEMATICS OF COMPUTATION
https://doi.org/10.1090/mcom/3506
Article electronically published on January 30, 2020

A LINEARLY CONVERGENT MAJORIZED ADMM WITH INDEFINITE PROXIMAL TERMS FOR CONVEX COMPOSITE PROGRAMMING AND ITS APPLICATIONS

NING ZHANG, JIA WU, AND LIWEI ZHANG

Abstract. This paper aims to study a majorized alternating direction method of multipliers with indefinite proximal terms (iPADMM) for convex composite optimization problems. We show that the majorized iPADMM for 2-block convex optimization problems converges globally under weaker conditions than those used in the literature and exhibits a linear convergence rate under a local error bound condition. Based on these, we establish the linear rate convergence results for a symmetric Gauss-Seidel based majorized iPADMM, which is designed for multiblock composite convex optimization problems. Moreover, we apply the majorized iPADMM to solve different types of regularized logistic regression problems. The numerical results on both synthetic and real datasets demonstrate the efficiency of the majorized iPADMM and also illustrate the effectiveness of the introduced indefinite proximal terms.

1. Introduction

In this paper, we consider the following convex composite optimization problem:

$$
\begin{aligned}
\min_{y,z} \quad & p(y) + f(y) + q(z) + g(z) \\
\text{s.t.} \quad & \mathcal{A}^* y + \mathcal{B}^* z = c, \\
& y \in \mathcal{Y}, z \in \mathcal{Z},
\end{aligned} \tag{1.1}
$$

where $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ are given finite dimensional Euclidean spaces each equipped with an inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$; $f : \mathcal{Y} \to (-\infty, +\infty)$ and $g : \mathcal{Z} \to (-\infty, +\infty)$ are two convex functions with Lipschitz continuous gradients on $\mathcal{Y}$ and $\mathcal{Z}$, respectively; $p : \mathcal{Y} \to (-\infty, +\infty]$ and $q : \mathcal{Z} \to (-\infty, +\infty]$ are two closed proper convex (not necessarily smooth) functions; $\mathcal{A}^* : \mathcal{Y} \to \mathcal{X}$ and $\mathcal{B}^* : \mathcal{Z} \to \mathcal{X}$ are adjoints of the linear operators $\mathcal{A} : \mathcal{X} \to \mathcal{Y}$ and $\mathcal{B} : \mathcal{X} \to \mathcal{Z}$, respectively; and $c \in \mathcal{X}$.

Received by the editor February 6, 2018, and, in revised form, February 22, 2019, and October 28, 2019.

2010 *Mathematics Subject Classification.* Primary 90C25, 90C30, 65K05, 62J12.

Key words and phrases. Alternating direction method of multiplier, linear rate convergence, indefinite proximal term, logistic regression, symmetric Gauss-Seidel decomposition.

The first author's research was supported by the National Natural Science Foundation of China (11901083).

The second author is the corresponding author.

The third author's research was supported by the National Natural Science Foundation of China (11971089, 11571059 and 11731013).

Let $\mathcal{M} : \mathcal{X} \to \mathcal{X}$ be a self-adjoint linear operator (not necessarily positive semidefinite), denote $\|x\|_{\mathcal{M}}^2 := \langle x, \mathcal{M}x\rangle$. Let $\sigma > 0$ be a given parameter. The augmented Lagrangian function of (1.1) is defined by

$$\mathcal{L}_\sigma(y,z;x) = f(y) + p(y) + g(z) + q(z) + \langle x, \mathcal{A}^*x + \mathcal{B}^*y - c\rangle + \frac{\sigma}{2}\|\mathcal{A}^*x + \mathcal{B}^*y - c\|^2.$$

Consider the following general 2-block ADMM iterative scheme

$$\left\{\begin{array}{l} y^{k+1} \in \operatorname{argmin}\left\{\mathcal{L}_\sigma(y, z^k; x^k) + \frac{1}{2}\|y - y^k\|_{\mathcal{S}}^2 \mid y \in \mathcal{Y}\right\}, \\ z^{k+1} \in \operatorname{argmin}\left\{\mathcal{L}_\sigma(y^{k+1}, z; x^k) + \frac{1}{2}\|z - z^k\|_{\mathcal{T}}^2 \mid z \in \mathcal{Z}\right\}, \\ x^{k+1} = x^k + \tau\sigma(\mathcal{A}^*y^{k+1} + \mathcal{B}^*z^{k+1} - c). \end{array}\right. \tag{1.2}$$

It is well known that if $\mathcal{S} = 0$ and $\mathcal{T} = 0$, the iterative scheme (1.2) is exactly the classic ADMM designed by Glowinski and Marrocco [23] and Gabay and Bertrand [20]; if $\mathcal{S} \succ 0$, $\mathcal{T} \succ 0$ and $\tau = 1$, iterative scheme (1.2) reduces to the method of the proximal ADMM introduced by Eckstein [15]; if both $\mathcal{S}$ and $\mathcal{T}$ are self-adjoint positive semidefinite linear operators, $\tau \in (0, (1+\sqrt{5})/2)$, iterative scheme (1.2) is known as the semiproximal ADMM (sPADMM) which is proposed by Fazel et al. [17]. To know more about the above-mentioned works and their relationships with well-known methods, such as proximal point algorithm (PPA) and Douglas-Rachford (DR) splitting method, we refer the readers to [8, 14–16, 18, 21–23, 25, 31].

One of the most important motivations behind ADMM is to fully use the separable structures in the problems. In other words, a potential assumption of using ADMM is that each subproblem can be efficiently solved. Generally speaking, if f or g is not a quadratic or linear function, the corresponding subproblem does not have closed-form solutions or cannot be solved easily. In order to continue enjoying benefits of the separable structure, Li et al. [31] extended the sPADMM to a majorized ADMM with indefinite proximal terms (iPADMM). Compared with the majorized techniques mentioned in [1, 8], the majorized iPADMM uses the positive semidefinite operators $\widehat{\Sigma}_f$ and $\widehat{\Sigma}_g$ (see (1.5) and (1.6)) instead of the Lipschitz constants of the gradient mappings ∇f and ∇g. The motivation behind this is for the better numerical performance. This will be illustrated in Table 3. Li et al. [31] established the global convergence and the iteration-complexity in the nonergodic sense of the majorized iPADMM, but not the rate of convergence. As for the rate of the convergence, most recently, based on the easy-to-use convergence theorem in Fazel et al. [17], the Q-linear rate convergence theorem for sPADMM with $\tau \in (0, (1+\sqrt{5})/2)$ was established in [25] under a calmness condition, which holds automatically for convex composite piecewise-linear programming. To know more about the convergence rate analysis of the ADMM-type methods, we refer to [25] and the references therein.

Next, we introduce a majorized iPADMM for solving the convex composite optimization problem (1.1) which will be proved to be a globally linearly convergent algorithm under some mild conditions. Since both $f(\cdot)$ and $g(\cdot)$ are smooth convex functions, there exist self-adjoint and positive semidefinite linear operators Σ_f and Σ_g such that for any $y, y' \in \mathcal{Y}$ and any $z, z' \in \mathcal{Z}$,

$$f(y) \geq f(y') + \langle \nabla f(y'), y - y'\rangle + \frac{1}{2}\|y - y'\|_{\Sigma_f}^2, \tag{1.3}$$

$$g(z) \geq g(z') + \langle \nabla g(z'), z - z'\rangle + \frac{1}{2}\|z - z'\|_{\Sigma_g}^2. \tag{1.4}$$

In addition, by the condition that the gradients $\nabla f(\cdot)$ and $\nabla g(\cdot)$ are Lipschitz continuous, we know that there exist self-adjoint and positive semidefinite linear operators $\widehat{\Sigma}_f \succeq \Sigma_f$ and $\widehat{\Sigma}_g \succeq \Sigma_g$ such that for any $y, y' \in \mathcal{Y}$ and any $z, z' \in \mathcal{Z}$,

$$f(y) \leq \hat{f}(y, y') := f(y') + \langle \nabla f(y'), y - y' \rangle + \frac{1}{2}\|y - y'\|^2_{\widehat{\Sigma}_f}, \tag{1.5}$$

$$g(z) \leq \hat{g}(z, z') := g(z') + \langle \nabla g(z'), z - z' \rangle + \frac{1}{2}\|z - z'\|^2_{\widehat{\Sigma}_g}. \tag{1.6}$$

For any given $(y', z') \in \mathcal{Y} \times \mathcal{Z}$, $\sigma \in (0, +\infty)$ and any $(y, z, x) \in \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$, the majorized augmented Lagrangian function is defined as

$$\begin{aligned} \widehat{\mathcal{L}}_\sigma(y, z, x; y', z') := \; & \hat{f}(y, y') + p(y) + \hat{g}(z, z') + q(z) + \langle x, \mathcal{A}^* y + \mathcal{B}^* z - c \rangle \\ & + \frac{\sigma}{2}\|\mathcal{A}^* y + \mathcal{B}^* z - c\|^2, \end{aligned} \tag{1.7}$$

where $\hat{f}$ and $\hat{g}$ are defined by (1.5) and (1.6), respectively. Then the majorized iPADMM for solving (1.1) proposed in our paper can be described as follows.

Algorithm 1 A majorized iPADMM for solving (1.1)

Let $\sigma > 0$ and $\tau \in (0, (1+\sqrt{5})/2)$ be given parameters. Let $\mathcal{S}$ and $\mathcal{T}$ be given self-adjoint, linear operators and satisfy

$$\tfrac{1}{2}\widehat{\Sigma}_f + \mathcal{S} \succeq 0, \; \tfrac{1}{2}\widehat{\Sigma}_g + \mathcal{T} \succeq 0, \; \tfrac{1}{2}\widehat{\Sigma}_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* \succ 0, \; \tfrac{1}{2}\widehat{\Sigma}_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* \succ 0. \tag{1.8}$$

Input $(y^0, z^0, x^0) \in \operatorname{dom} p \times \operatorname{dom} q \times \mathcal{X}$. Set $k := 0$.

Step 1.: Compute

$$\left\{ \begin{aligned} & y^{k+1} = \arg\min_{y \in \mathcal{Y}} \; \widehat{\mathcal{L}}_\sigma(y, z^k, x^k; y^k, z^k) + \frac{1}{2}\|y - y^k\|^2_{\mathcal{S}} \\ & \qquad = \arg\min_{y \in \mathcal{Y}} \; p(y) + \langle y, \nabla f(y^k) + \mathcal{A}x^k \rangle + \frac{\sigma}{2}\|\mathcal{A}^* y + \mathcal{B}^* z^k - c\|^2 \\ & \qquad\qquad + \frac{1}{2}\|y - y^k\|^2_{\widehat{\Sigma}_f + \mathcal{S}}, & (1.9\text{a}) \\ & z^{k+1} = \arg\min_{z \in \mathcal{Z}} \; \widehat{\mathcal{L}}_\sigma(y^{k+1}, z, x^k; y^k, z^k) + \frac{1}{2}\|z - z^k\|^2_{\mathcal{T}} \\ & \qquad = \arg\min_{z \in \mathcal{Z}} \; q(z) + \langle z, \nabla g(z^k) + \mathcal{B}x^k \rangle + \frac{\sigma}{2}\|\mathcal{A}^* y^{k+1} + \mathcal{B}^* z - c\|^2 \\ & \qquad\qquad + \frac{1}{2}\|z - z^k\|^2_{\widehat{\Sigma}_g + \mathcal{T}}, & (1.9\text{b}) \\ & x^{k+1} := x^k + \tau\sigma(\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^{k+1} - c). & (1.9\text{c}) \end{aligned} \right.$$

Step 2.: If a termination criterion is not met, set $k := k + 1$ and go to Step 1.

Note that Algorithm 1 and its convergence results are greatly motivated by [31] and [25], respectively. However, our algorithm could substantially improve both the numerical and the theoretical results. Details are listed below.

(i) We refine the conditions in [31, Theorem 10 (b)] with

$$\tfrac{1}{2}\widehat{\Sigma}_f + \mathcal{S} \succeq 0, \; \tfrac{1}{2}\widehat{\Sigma}_g + \mathcal{T} \succeq 0, \; \tfrac{1}{2}\Sigma_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* \succ 0, \; \tfrac{1}{2}\Sigma_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* \succ 0 \tag{1.10}$$

being replaced by (1.8). Note that condition (1.10) coincides with (1.8) when f and g are linear/quadratic. However, there are many widely used loss functions in statistical inference and machine learning that are not linear/quadratic, such as the logistic loss function, the multinomial logistic loss function (see, e.g., [19]), and the cox loss function (see, e.g., [46]). In these cases, the positive semidefinite linear operator Σ_f or Σ_g is not easy to estimate and accordingly has to be chosen as a zero matrix. Consequently, the majorized iPADMM proposed in [31] with condition (1.10) for solving the regularized (multinomial) logistic/cox regression problem is reduced to the majorized ADMM with semidefinite proximal terms (majorized sPADMM). As mentioned earlier, the indefinite proximal terms can improve the convergence speed of the algorithm in practice. Therefore, the improvement from (1.8) could be significant when f or g is not linear/quadratic.

(ii) Note that the global convergence theorem and the linear rate convergence results on sPADMM [17, 25] are no longer applicable due to the indefiniteness of the proximal terms as well as the introduction of the majorization technique. Moreover, Algorithm 1 is also beyond the framework of the majorized iPADMM with condition (1.10) in general. Therefore, to establish the global convergence theorem with condition (1.8), we refine most of the important results in [31] by developing a key result (Lemma 3.2). Consequently, we can prove that the majorized iPADMM with condition (1.8) is globally linearly convergent under a metric subregularity condition (see (2.2)) which holds automatically for a class of widely used problems.

Inspired by the global convergence and excellent numerical performance of symmetric Gauss-Seidel based multiblock sPADMM (sGS-sPADMM) for multiblock linearly constrained convex programming, we present a symmetric Gauss-Seidel based multiblock majorized iPADMM (majorized sGS-iPADMM). The linear rate convergence results for majorized sGS-iPADMM are established through converting it into an equivalent 2-block majorized iPADMM. This is one of the most important motivations that we consider the objective function in (1.1) in the form of $f(y) + p(y)$ and $g(z) + q(z)$. It is worth mentioning that the sGS-sPADMM was initially presented by Li et al. [32]. For more discussions about the symmetric Gauss-Seidel techniques, we refer to [4, 32–34] and the references therein.

The rest of the paper is organized as follows. In section 2, we give some preliminaries that will be frequently used in other sections. In section 3, we refine the convergence result [31, Theorem 10 (b)] with condition (1.10) and then establish a general Q-linear rate convergence theorem under a metric subregularity condition. In section 4, we propose a majorized sGS-iPADMM for the multiblock composite optimization problem. Moreover, we show its convergence rate by establishing the relationship between the majorized sGS-iPADMM and the 2-block majorized iPADMM. In section 5, we apply the majorized iPADMM to three types of regularized logistic regression and then present the numerical results. Finally, we give some concluding remarks and future works in section 6.

2. Preliminaries

In this section, we summarize and study some preliminaries that will be used in the subsequent analysis. Let $(\bar{y}, \bar{z})$ be the optimal solution of problem (1.1). If there exists $\bar{x} \in \mathcal{X}$ such that $(\bar{y}, \bar{z}, \bar{x})$ satisfies the following Karush-Kuhn-Tucker (KKT) system:

$$0 \in \partial p(y) + \nabla f(y) + \mathcal{A}x, \ 0 \in \partial q(z) + \nabla g(z) + \mathcal{B}x, \ \mathcal{A}^*y + \mathcal{B}^*z - c = 0, \tag{2.1}$$

then, $(\bar{y}, \bar{z}, \bar{x})$ is called a KKT point of problem (1.1). Denote the set of KKT points by $\overline{\Omega}$.

For any convex function $\theta : \mathcal{Y} \to \Re \cup \{+\infty\}$, the Moreau-Yosida proximal mapping $\mathrm{Pr}_\theta(\cdot)$ associated with θ is defined by

$$\mathrm{Pr}_\theta(y) := \arg\min_{y' \in \mathcal{Y}} \left\{ \theta(y') + \frac{1}{2}\|y' - y\|^2 \right\} \ \forall y \in \mathcal{Y}.$$

It is well known that the Moreau-Yosida proximal mapping $\mathrm{Pr}_\theta(\cdot)$ is globally Lipschitz continuous with modulus one; see, e.g., [26,30].

Denote $u := (y, z, x) \in \mathcal{U}$ with $\mathcal{U} := \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$. Define the KKT mapping $\mathcal{R} : \mathcal{U} \to \mathcal{U}$ as

$$\mathcal{R}(u) := \begin{pmatrix} y - \mathrm{Pr}_p[y - (\nabla f(y) + \mathcal{A}x)] \\ z - \mathrm{Pr}_q[z - (\nabla g(z) + \mathcal{B}x)] \\ c - \mathcal{A}^* y - \mathcal{B}^* z \end{pmatrix}.$$

From [38], we know that $u \in \overline{\Omega}$ if and only if $\mathcal{R}(u) = 0$. Let $\mathcal{F} : \mathcal{X} \rightrightarrows \mathcal{Y}$ be a multivalued mapping. Denote its inverse by $\mathcal{F}^{-1}$. Define the graph of multivalued function $\mathcal{F}$ as follows:

$$\mathrm{gph}\,\mathcal{F} := \{(x, y) \in \mathcal{X} \times \mathcal{Y} \,|\, y \in \mathcal{F}(x)\}.$$

Definition 2.1. A multivalued mapping $\mathcal{F} : \mathcal{X} \rightrightarrows \mathcal{Y}$ is said to be metrically subregular at $(\bar{x}, \bar{y}) \in \mathrm{gph}\,\mathcal{F}$ with modulus $\eta > 0$ if there exists a neighborhood $\mathcal{U}$ of $\bar{x}$ such that

$$\mathrm{dist}(x, \mathcal{F}^{-1}(\bar{y})) \le \eta \mathrm{dist}(\bar{y}, \mathcal{F}(x)) \ \forall x \in \mathcal{U}.$$

The definition of metric subregularity is directly from [11, Definition 3.1]. It is well known that $\mathcal{F}$ is metrically subregular at $(\bar{x}, \bar{y}) \in \mathrm{gph}\,\mathcal{F}$ if and only if its inverse multivalued mapping $\mathcal{F}^{-1}$ is calm (cf. [48, Definition 2.6], [12, 3.8(3H)]) at $(\bar{y}, \bar{x}) \in \mathrm{gph}\,\mathcal{F}^{-1}$.

From [41, Proposition 1] and [44], we know that if $\mathcal{F}$ is piecewise polyhedral or $\mathcal{F}$ is the subdifferential mapping of a convex piecewise linear-quadratic function, then $\mathcal{F}$ is metrically subregular at $(\bar{x}, \bar{y}) \in \mathrm{gph}\,\mathcal{F}$. Till now, numerous works have been done to study the sufficient conditions of calmness of KKT solution mappings, we refer to [9,13,34,49] and the references therein. In order to establish the linear rate convergence of the majorized iPADMM, we need the metric subregularity of the KKT mapping $\mathcal{R}$. From Definition 2.1, the metric subregularity of $\mathcal{R}$ at $(\bar{u}, 0) \in \mathrm{gph}\,\mathcal{R}$ with modulus $\eta > 0$ can be described as: there exists a scalar $\rho > 0$ such that

$$\mathrm{dist}(u, \overline{\Omega}) \le \eta \|\mathcal{R}(u)\| \ \forall u \in \{u \in \mathcal{U} : \|u - \bar{u}\| \le \rho\}. \tag{2.2}$$

The above condition is also referred to as the existence of a local error bound. From [49, Theorem 4.3] and [10, Theorem 3.1], we know that the KKT mappings corresponding to the Lasso, Fused Lasso regularized logistic regression models are metrically subregular at $(\bar{u}, 0) \in \mathrm{gph}\,\mathcal{R}$.

Since for any proper closed convex function $\vartheta : \mathcal{Y} \to \Re \cup \{+\infty\}$, the subdifferential $\partial\vartheta(\cdot)$ is a monotone multivalued mapping (see [42]), i.e., for any $y^1, y^2 \in \mathrm{dom}\vartheta$, it holds that

$$\langle \zeta^1 - \zeta^2, y^1 - y^2 \rangle \ge 0 \ \forall \zeta^1 \in \partial\vartheta(y^1) \, \forall \zeta^2 \in \partial\vartheta(y^2). \tag{2.3}$$

Proposition 2.1 ([6, Proposition 2.6.5]). *Let multivalued mapping $\mathcal{F}$ be Lipschitz on an open convex set $\mathcal{V}$ in $\Re^n$, and let x and y be points in $\mathcal{V}$. Then one has*

$$\mathcal{F}(y) - \mathcal{F}(x) \in \operatorname{conv} \partial\mathcal{F}([x, y])(y - x),$$

where $\operatorname{conv} \partial\mathcal{F}([x, y])(y-x)$ *denotes the convex hull of all points of the form* $\zeta(y-x)$*, where* $\zeta \in \partial\mathcal{F}(u)$ *for some point* u *in* $[x, y]$.

Throughout the subsequent analysis, we always assume that the following two assumptions hold.

Assumption 2.1. The KKT system (2.1) has at least one solution, i.e., $\overline{\Omega} \neq \emptyset$.

Assumption 2.2. The two self-adjoint linear operators $\mathcal{S} : \mathcal{Y} \to \mathcal{Y}$ and $\mathcal{T} : \mathcal{Z} \to \mathcal{Z}$ in majorized iPADMM satisfy

$$\mathcal{S} \succeq -\frac{1}{2}\widehat{\Sigma}_f \quad \text{and} \quad \mathcal{T} \succeq -\frac{1}{2}\widehat{\Sigma}_g. \tag{2.4}$$

Remark 2.1. Assumption 2.2 means that the proximal terms $\mathcal{S}$ and $\mathcal{T}$ cannot be too indefinite as long as $\widehat{\Sigma}_f$ and $\widehat{\Sigma}_g$ are not very big. Note that $\widehat{\Sigma}_f$ and $\widehat{\Sigma}_g$ should be chosen as small as possible provided (1.5) and (1.6) are satisfied. For example, when f is a convex quadratic function, we choose $\widehat{\Sigma}_f = \Sigma_f = \nabla^2 f$, where $\nabla^2 f$ is the Hessian matrix of f.

3. Q-linear rate of convergence of the majorized iPADMM

This section aims to analyze the convergence rate of the majorized iPADMM for solving (1.1). We show that the algorithm achieves a Q-linear rate of convergence under some mild conditions. Before formally stating our main results, we first give some technical results.

3.1. Technical lemmas.

For notational convenience, for any $\tau \in (0, +\infty)$, define

$$s_\tau := \frac{5 - \tau - 3\min(\tau, \tau^{-1})}{4}, \;\; t_\tau := \frac{1 - \tau + \min(\tau, \tau^{-1})}{2},$$

and two self-adjoint linear operators:

$$\mathcal{M} := \operatorname{Diag}\left(\widehat{\Sigma}_f + \mathcal{S}, \, \widehat{\Sigma}_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^*, (\tau\sigma)^{-1}\mathcal{I}\right) + s_\tau\sigma\mathcal{E}\mathcal{E}^*, \tag{3.1}$$

$$\mathcal{H} := \operatorname{Diag}\left(\mathcal{H}_f, \mathcal{H}_g, t_\tau(\tau^2\sigma)^{-1}\mathcal{I}\right) + \frac{1}{4}t_\tau\sigma\mathcal{E}\mathcal{E}^*, \tag{3.2}$$

where

$$\mathcal{H}_f := \frac{1}{2}\,\widehat{\Sigma}_f + \mathcal{S} \succeq 0, \;\; \mathcal{H}_g := \frac{1}{2}\,\widehat{\Sigma}_g + \mathcal{T} + 2t_\tau\tau\sigma\mathcal{B}\mathcal{B}^* \succeq 0,$$

and $\mathcal{E} : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z} \times \mathcal{X}$ is the linear operator such that its adjoint $\mathcal{E}^*$ satisfies $\mathcal{E}^*(y, z, x) = \mathcal{A}^*y + \mathcal{B}^*z$. For a given self-adjoint linear operator $\mathcal{G} : \mathcal{X} \to \mathcal{X}$, we denote the largest eigenvalue by $\lambda_{\max}(\mathcal{G})$ and for any $k \geq 0$,

$$r^k := \mathcal{A}^*y^k + \mathcal{B}^*z^k - c.$$

The proofs of the lemmas in this subsection are all presented in Appendix A for readability.

Lemma 3.1. *Let $\{u^k := (y^k, z^k, x^k)\}$ be the infinite sequence generated by the majorized iPADMM. Then for any $k \geq 0$,*

$$\|\mathcal{R}(u^{k+1})\|^2 \leq \|u^{k+1} - u^k\|^2_{\mathcal{H}_0}, \tag{3.3}$$

where

$$\mathcal{H}_0 := \max\{\kappa_1, \kappa_2, \kappa_3\}\mathrm{Diag}\left(\mathcal{I},\, \mathcal{I} + \sigma\mathcal{B}\mathcal{B}^*,\, (\tau^2\sigma)^{-1}\mathcal{I}\right)$$

with

$$\begin{aligned}
\kappa_1 &:= 3(\lambda_{\max}(\mathcal{S} + \tfrac{1}{2}\widehat{\Sigma}_f) + \tfrac{1}{2}\lambda_{\max}(\widehat{\Sigma}_f))^2,\\
\kappa_2 &:= \max\{2(\lambda_{\max}(\mathcal{T} + \tfrac{1}{2}\widehat{\Sigma}_g) + \tfrac{1}{2}\lambda_{\max}(\widehat{\Sigma}_g))^2,\, 3\sigma\lambda_{\max}(\mathcal{A}^*\mathcal{A})\},\\
\kappa_3 &:= \sigma^{-1} + (1-\tau)^2\sigma(3\lambda_{\max}(\mathcal{A}^*\mathcal{A}) + 2\lambda_{\max}(\mathcal{B}^*\mathcal{B})).
\end{aligned}$$

The above lemma is inspired by [25, Lemma 1], but its proof is more complicated due to the majorization techniques. As mentioned in the introduction, the following lemma will play an essential role in refining the global convergence results of the majorized iPADMM presented in [31, Theorem 10(b)].

Lemma 3.2. *Let $h : \mathcal{X} \to \Re$ be a smooth convex function and there is a self-adjoint positive semidefinite linear operator $\mathcal{P}$ such that, for any given $\bar{x} \in \mathcal{X}$,*

$$h(x) \leq h(\bar{x}) + \langle \nabla h(\bar{x}),\, x - \bar{x}\rangle + \frac{1}{2}\|x - \bar{x}\|^2_{\mathcal{P}} \ \ \forall x \in \mathcal{X}.$$

Then it holds that

$$\langle \nabla h(x) - \nabla h(\bar{x}), y - \bar{x}\rangle \geq -\frac{1}{4}\|x - y\|^2_{\mathcal{P}} \ \ \forall x, y \in \mathcal{X}.$$

Greatly based on the inequality obtained from the above lemma, we can derive the following results, which will be used to improve the conclusions in [31, Proposition 8]. It is worth noting that the following lemma will be a critical tool for establishing the convergence theorem for majorized iPADMM (Algorithm 1).

Lemma 3.3. *Let $\{(y^k, z^k, x^k)\}$ be the infinite sequence generated by the majorized iPADMM. Then, for any $\bar{u} := (\bar{y}, \bar{z}, \bar{x}) \in \overline{\Omega}$, $\tau > 0$, and $k \geq 0$, we have*

$$\begin{aligned}
\phi_k - \phi_{k+1} \geq \ & \|y^{k+1} - y^k\|^2_{\frac{1}{2}\widehat{\Sigma}_f + \mathcal{S}} + \|z^{k+1} - z^k\|^2_{\frac{1}{2}\widehat{\Sigma}_g + \mathcal{T}}\\
& + (1-\tau)\sigma\|r^{k+1}\|^2 + \sigma\|\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^k - c\|^2,
\end{aligned} \tag{3.4}$$

where for any $k \geq 0$,

$$\phi_k := (\tau\sigma)^{-1}\|x^k - \bar{x}\|^2 + \|y^k - \bar{y}\|^2_{\widehat{\Sigma}_f + \mathcal{S}} + \|z^k - \bar{z}\|^2_{\widehat{\Sigma}_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^*}. \tag{3.5}$$

Remark 3.1. Recall that the inequality proposed in [31, Proposition 8] for the same purpose as (3.4) is given as

$$\begin{aligned}
\phi_k - \phi_{k+1} \geq \ & \|y^{k+1} - y^k\|^2_{\frac{1}{2}\Sigma_f + \mathcal{S}} + \|z^{k+1} - z^k\|^2_{\frac{1}{2}\Sigma_g + \mathcal{T}}\\
& + (1-\tau)\sigma\|r^{k+1}\|^2 + \sigma\|\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^k - c\|^2,
\end{aligned} \tag{3.6}$$

where Σ_f and Σ_g are given by (1.3) and (1.4), respectively. Since $\Sigma_f \preceq \widehat{\Sigma}_f$ and $\Sigma_g \preceq \widehat{\Sigma}_g$, we know that the inequality (3.4) is much tighter than (3.6) when f or g is not a linear/quadratic function. Consequently, the global convergence theorem can be established under the weaker condition (1.8).

Since the proof of the following lemma is not much different from the one in [31, Theorem 10, Inequality (55)] except for replacing Inequality (33) in [31] by (3.4) in Lemma 3.3, we include an outline in Appendix A.4.

Lemma 3.4. *Let $\{u^k := (y^k, z^k, x^k)\}$ be the infinite sequence generated by the majorized iPADMM. For each k and any KKT point $\bar{u} := (\bar{y}, \bar{z}, \bar{x})$, let ϕ_k be defined in* (3.5). *Then, for any $k \geq 1$, one has*

$$\begin{aligned}
&\left[\phi_k + (1-\min(\tau,\tau^{-1}))\sigma\|r^k\|^2 + \|z^k - z^{k-1}\|^2_{\widehat{\Sigma}_g+\mathcal{T}}\right] \\
&- \left[\phi_{k+1} + (1-\min(\tau,\tau^{-1}))\sigma\|r^{k+1}\|^2 + \|z^{k+1} - z^k\|^2_{\widehat{\Sigma}_g+\mathcal{T}}\right] \\
&\geq t_{k+1} + (-\tau + \min(1+\tau, 1+\tau^{-1}))\sigma\|r^{k+1}\|^2,
\end{aligned} \tag{3.7}$$

where

$$t_{k+1} := \|y^{k+1} - y^k\|^2_{\mathcal{H}_f} + \|z^{k+1} - z^k\|^2_{\mathcal{H}_g}.$$

The following lemma will be used to derive the global convergence theorem for the majorized iPADMM.

Lemma 3.5. *Let $\tau \in (0, (1+\sqrt{5})/2)$, $\mathcal{M}$ and $\mathcal{H}$ be defined by* (3.1) *and* (3.2), *respectively. Let*

$$\frac{1}{2}\widehat{\Sigma}_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* \succ 0 \quad \text{and} \quad \frac{1}{2}\widehat{\Sigma}_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* \succ 0. \tag{3.8}$$

Then, one has that

(a) *condition* (3.8) *holds if and only if $\mathcal{M} \succ 0$;*
(b) *if condition* (3.8) *holds, then $\mathcal{H} \succ 0$.*

3.2. **Convergence analysis.** In this subsection, we investigate the rate of convergence of majorized iPADMM for solving (1.1). Inspired by [25, Proposition 4], we first develop a key inequality needed for proving the linear rate convergence for the majorized iPADMM.

Proposition 3.1. *Let $\tau \in (0, (1+\sqrt{5})/2)$ and $\{u^k := (y^k, z^k, x^k)\}$ be the infinite sequence generated by the majorized iPADMM. Then for any KKT point $\bar{u} := (\bar{y}, \bar{z}, \bar{x})$ and any $k \geq 1$,*

$$\begin{aligned}
&\|u^{k+1} - \bar{u}\|^2_{\mathcal{M}} + \|z^{k+1} - z^k\|^2_{\widehat{\Sigma}_g+\mathcal{T}} \\
&\leq \|u^k - \bar{u}\|^2_{\mathcal{M}} + \|z^k - z^{k-1}\|^2_{\widehat{\Sigma}_g+\mathcal{T}} - \|u^{k+1} - u^k\|^2_{\mathcal{H}}.
\end{aligned} \tag{3.9}$$

Consequently, we have for all $k \geq 1$,

$$\begin{aligned}
&\mathrm{dist}^2_{\mathcal{M}}(u^{k+1}, \overline{\Omega}) + \|z^{k+1} - z^k\|^2_{\widehat{\Sigma}_g+\mathcal{T}} \\
&\leq \mathrm{dist}^2_{\mathcal{M}}(u^k, \overline{\Omega}) + \|z^k - z^{k-1}\|^2_{\widehat{\Sigma}_g+\mathcal{T}} - \|u^{k+1} - u^k\|^2_{\mathcal{H}}.
\end{aligned} \tag{3.10}$$

Proof. By reorganizing the inequality in (3.7), one has

$$\begin{aligned}
&\left[\phi_{k+1} + s_\tau\sigma\|r^{k+1}\|^2 + \|z^{k+1} - z^k\|^2_{\widehat{\Sigma}_g+\mathcal{T}}\right] - \left[\phi_k + s_\tau\sigma\|r^k\|^2 + \|z^k - z^{k-1}\|^2_{\widehat{\Sigma}_g+\mathcal{T}}\right] \\
&\leq -\left\{t_\tau\sigma\|r^{k+1}\|^2 + \tfrac{1}{2}\sigma t_\tau(\|r^{k+1}\|^2 + \|r^k\|^2) + t_{k+1}\right\}.
\end{aligned} \tag{3.11}$$

From the definitions of x^{k+1} and $(\bar{y}, \bar{z}, \bar{x})$, we have

$$\begin{aligned}
&r^{k+1} = (\tau\sigma)^{-1}(x^{k+1} - x^k) = \mathcal{A}^*(y^{k+1} - \bar{y}) + \mathcal{B}^*(z^{k+1} - \bar{z}), \\
&r^k = \mathcal{A}^*(y^k - \bar{y}) + \mathcal{B}^*(z^k - \bar{z}), \\
&\|r^{k+1}\|^2 + \|r^k\|^2 \geq \tfrac{1}{2}\|\mathcal{A}^*(y^{k+1} - y^k) + \mathcal{B}^*(z^{k+1} - z^k)\|^2.
\end{aligned} \tag{3.12}$$

Then we can get (3.9) by substituting (3.12) into (3.11). Since (3.9) holds for any $\bar{u} \in \overline{\Omega}$, we can get (3.10) from the fact that $\overline{\Omega}$ is a nonempty closed convex set. The proof is completed. $\square$

Now we are ready to establish the global convergence and the linear rate of convergence for the majorized iPADMM under a metric subregularity condition of $\mathcal{R}$ at some $(\bar{u}, 0) \in \mathrm{gph}\mathcal{R}$.

Theorem 3.1. *Let $\tau \in (0, (1+\sqrt{5})/2)$ and $\{u^k := (y^k, z^k, x^k)\}$ be the infinite sequence generated by the majorized iPADMM. Then, one has the following results:*

(a) *The sequence $\{(y^k, z^k)\}$ converges to an optimal solution of problem* (1.1) *and $\{x^k\}$ converges to an optimal solution of the dual of problem* (1.1).

(b) *Suppose that the sequence $\{(y^k, z^k, x^k)\}$ converges to a KKT point $\bar{u} := (\bar{y}, \bar{z}, \bar{x})$ and the KKT mapping $\mathcal{R}$ is metrically subregular at $(\bar{u}, 0) \in \mathrm{gph}\mathcal{R}$ with modulus $\eta > 0$. Then there exist a positive number $\mu \in (0, 1)$ and an integer $k_0 \geq 1$ such that for all $k \geq k_0$,*

$$\mathrm{dist}^2_{\mathcal{M}}(u^{k+1}, \overline{\Omega}) + \|z^{k+1} - z^k\|^2_{\widehat{\Sigma}_g + \mathcal{T}} \leq \mu \left[\mathrm{dist}^2_{\mathcal{M}}(u^k, \overline{\Omega}) + \|z^k - z^{k-1}\|^2_{\widehat{\Sigma}_g + \mathcal{T}}\right]. \tag{3.13}$$

Moreover, there exists a positive number $\hat{\mu} \in [\mu, 1)$ such that for all $k \geq 1$,

$$\mathrm{dist}^2_{\mathcal{M}}(u^{k+1}, \overline{\Omega}) + \|z^{k+1} - z^k\|^2_{\widehat{\Sigma}_g + \mathcal{T}} \leq \hat{\mu} \left[\mathrm{dist}^2_{\mathcal{M}}(u^k, \overline{\Omega}) + \|z^k - z^{k-1}\|^2_{\widehat{\Sigma}_g + \mathcal{T}}\right]. \tag{3.14}$$

Proof. We first prove the convergence on the sequences $\{(y^k, z^k)\}$ and $\{x^k\}$. From Lemma 3.5 and (1.8), we know that $\mathcal{H} \succ 0$, $\mathcal{M} \succ 0$, and $\widehat{\Sigma}_g + \mathcal{T} \succeq 0$. Then it holds from Proposition 3.1 that $\{u^{k+1}\}$ is bounded and

$$\lim_{k \to \infty} \|u^{k+1} - u^k\| = 0.$$

Consequently, there is a subsequence $\{u^{k_i}\}$ which converges to a cluster point u^∞. From Lemma 3.1, we know that

$$\|\mathcal{R}(u^{k_i})\|^2 \leq \|u^{k_i} - u^{k_i - 1}\|^2_{\mathcal{H}_0},$$

where $\mathcal{H}_0 \succ 0$. Taking limits on both sides of the above inequality, we obtain $\|\mathcal{R}(u^\infty)\| = 0$. Thus, the subsequence $\{u^{k_i}\}$ converges to $u^\infty \in \overline{\Omega}$. Therefore, the sequence $\{\|u^{k_i+1} - u^\infty\|^2_{\mathcal{M}} + \|z^{k_i+1} - z^{k_i}\|^2_{\widehat{\Sigma}_g + \mathcal{T}}\}$ converges to 0 as $k_i \to \infty$. Since the sequence $\{\|u^{k+1} - \bar{u}\|^2_{\mathcal{M}} + \|z^{k+1} - z^k\|^2_{\widehat{\Sigma}_g + \mathcal{T}}\}$ is nonincreasing and $\|u^{k+1} - u^k\| \to 0$, we have

$$\lim_{k \to \infty} \|u^k - u^\infty\| = 0.$$

Therefore, the whole sequence $\{u^k\}$ converges to u^∞. This completes the proof of the result (a). Next, we prove (b). From (a), we know that the sequence $\{(y^k, z^k, x^k)\}$ generated by the majorized iPADMM converges to a KKT point $\bar{u} = (\bar{y}, \bar{z}, \bar{x})$. Then there exist $\rho > 0$ and an integer $k_0 \geq 1$ such that for all $k \geq k_0$,

$$\|u^{k+1} - \bar{u}\| \leq \rho.$$

Therefore, by using Lemma 3.1 and (2.2), we know that for all $k \geq k_0$,

$$\mathrm{dist}^2(u^{k+1}, \overline{\Omega}) \leq \eta^2 \|\mathcal{R}(u^{k+1})\|^2 \leq \eta^2 \|u^{k+1} - u^k\|^2_{\mathcal{H}_0}. \tag{3.15}$$

The definition of $\mathcal{H}$ and the fact that $\mathcal{H} \succ 0$ imply that $\mathcal{H}_g \succ 0$. Then there exists a finite real number $\varrho_1 > 0$ such that $\widehat{\Sigma}_g + \mathcal{T} \preceq \varrho_1 \mathcal{H}_g$ and consequently, it holds that

$$\|z^{k+1} - z^k\|^2_{\widehat{\Sigma}_g + \mathcal{T}} \leq \varrho_1 \|u^{k+1} - u^k\|^2_{\mathcal{H}}.$$

Similarly, there exists a finite real number $\varrho_2 > 0$ such that $\mathcal{H}_0 \preceq \varrho_2 \mathcal{H}$. It follows from (3.15) that for all $k \geq k_0$,

$$\begin{aligned} \|u^{k+1} - u^k\|^2_{\mathcal{H}} &\geq \varrho_2^{-1} \|u^{k+1} - u^k\|^2_{\mathcal{H}_0} \\ &\geq \varrho_2^{-1} \eta^{-2} \mathrm{dist}^2(u^{k+1}, \overline{\Omega}) \geq \varrho_2^{-1} \eta^{-2} \lambda^{-1}_{\max}(\mathcal{M}) \mathrm{dist}^2_{\mathcal{M}}(u^{k+1}, \overline{\Omega}). \end{aligned}$$

Let $\kappa := \frac{1}{1+\varrho_1 \beta}$ with $\beta := \varrho_2^{-1} \eta^{-2} \lambda^{-1}_{\max}(\mathcal{M})$. From (3.10), we have that, for all $k \geq k_0$,

$$\begin{aligned} &\left[\mathrm{dist}^2_{\mathcal{M}}(u^{k+1}, \overline{\Omega}) + \|z^{k+1} - z^k\|^2_{\widehat{\Sigma}_g + \mathcal{T}}\right] - \left[\mathrm{dist}^2_{\mathcal{M}}(u^k, \overline{\Omega}) + \|z^k - z^{k-1}\|^2_{\widehat{\Sigma}_g + \mathcal{T}}\right] \\ &\leq -\big((1-\kappa)\|u^{k+1} - u^k\|^2_{\mathcal{H}} + \kappa \|u^{k+1} - u^k\|^2_{\mathcal{H}}\big) \\ &\leq -\left((1-\kappa)\varrho_1^{-1} \|z^{k+1} - z^k\|^2_{\widehat{\Sigma}_g + \mathcal{T}} + \kappa \varrho_2^{-1} \eta^{-2} \lambda^{-1}_{\max}(\mathcal{M}) \mathrm{dist}^2_{\mathcal{M}}(u^{k+1}, \overline{\Omega})\right). \end{aligned}$$

Then by reorganizing the above inequality, we know that for all $k \geq k_0$,

$$\begin{aligned} &\big(1 + \kappa \varrho_2^{-1} \eta^{-2} \lambda^{-1}_{\max}(\mathcal{M})\big) \mathrm{dist}^2_{\mathcal{M}}(u^{k+1}, \overline{\Omega}) + \big(1 + (1-\kappa)\varrho_1^{-1}\big) \|z^{k+1} - z^k\|^2_{\widehat{\Sigma}_g + \mathcal{T}} \\ &\leq \mathrm{dist}^2_{\mathcal{M}}(u^k, \overline{\Omega}) + \|z^k - z^{k-1}\|^2_{\widehat{\Sigma}_g + \mathcal{T}}. \end{aligned}$$

It is easy to check that

$$1 + \kappa \varrho_2^{-1} \eta^{-2} \lambda^{-1}_{\max}(\mathcal{M}) = 1/\mu,$$

where

$$\mu := \frac{\varrho_1 \beta + 1}{1 + \beta + \varrho_1 \beta} < 1.$$

Then we know that inequality (3.13) holds.

By combining (3.13) with Lemma 3.1, (3.10) in Proposition 3.1, we can obtain directly that there exists a positive number $\hat{\mu} \in [\mu, 1)$ such that (3.14) holds for all $k \geq 1$. This completes the proof. □

Remark 3.2. We make the following comments:

(i) For $\tau \in (0, (1+\sqrt{5})/2)$, $\alpha \in (\tau/\min(1+\tau, 1+\tau^{-1}), 1]$, based on the lemmas presented in subsection 3.1 and Proposition 3.1, by mimicking the proof of [31, Theorem 10(b)], we know that the global convergence result (a) also holds if condition (1.8) being replaced by the following condition:

$$\begin{aligned} &\widehat{\Sigma}_f + \mathcal{S} \succeq 0,\ \tfrac{1}{2}\widehat{\Sigma}_g + \mathcal{T} \succeq 0,\ \tfrac{1}{2}\widehat{\Sigma}_f + \mathcal{S} + \tfrac{1}{2}(1-\alpha)\sigma \mathcal{A}\mathcal{A}^* \succeq 0, \\ &\tfrac{1}{2}\widehat{\Sigma}_f + \mathcal{S} + \sigma \mathcal{A}\mathcal{A}^* \succ 0,\ \tfrac{1}{2}\widehat{\Sigma}_g + \mathcal{T} + \min\{\tau, 1+\tau-\tau^2\}\alpha\sigma \mathcal{B}\mathcal{B}^* \succ 0. \end{aligned} \tag{3.16}$$

This is also much weaker than the conditions used in [31, Theorem 10(b)], i.e.,

$$\begin{aligned} &\widehat{\Sigma}_f + \mathcal{S} \succeq 0,\ \tfrac{1}{2}\widehat{\Sigma}_g + \mathcal{T} \succeq 0,\ \tfrac{1}{2}\Sigma_f + \mathcal{S} + \tfrac{1}{2}(1-\alpha)\sigma \mathcal{A}\mathcal{A}^* \succeq 0, \\ &\tfrac{1}{2}\Sigma_f + \mathcal{S} + \sigma \mathcal{A}\mathcal{A}^* \succ 0,\ \tfrac{1}{2}\Sigma_g + \mathcal{T} + \min\{\tau, 1+\tau-\tau^2\}\alpha\sigma \mathcal{B}\mathcal{B}^* \succ 0. \end{aligned} \tag{3.17}$$

Analogous to condition (1.8), the improvement from condition (3.16) could be significant when f or g is not linear/quadratic. From the discussion in [31, Remark 11], we know that $\alpha \in [0.99998, 1]$ when $\tau = 1.618$. Therefore, for the purpose of readability, we focus on the case that $\alpha = 1$ in this paper.

(ii) By observing the expression of parameter μ in the proof of result (b) in Theorem 3.1, in order to increase the convergence speed, under the premise of satisfying (1.8), the linear operators $\widehat{\Sigma}_f$ and $\widehat{\Sigma}_g$ should be chosen such that the majorized functions $\hat{f}$ and $\hat{g}$ are as close to f and g as possible, and the proximal terms $\mathcal{S}$ and $\mathcal{T}$ should be chosen as close to $-\frac{1}{2}\widehat{\Sigma}_f$ and $-\frac{1}{2}\widehat{\Sigma}_g$ as possible. For simplicity, we assume that $\mathcal{A}\mathcal{A}^* \succ 0$, then one should choose $\mathcal{S} = -\frac{1}{2}\widehat{\Sigma}_f$. This illustrates the claim mentioned in the introduction that the linear rate convergence results for sPADMM established in [25] are no longer applicable for our majorized iPADMM.

4. Application I: A majorized sGS-iPADMM

Consider the following general multiblock convex composite programming model:

$$
\begin{array}{ll}
\min\limits_{y,z} & p(y_1) + f(y_1,\dots,y_s) + q(z_1) + g(z_1,\dots,z_t) \\
\text{s.t.} & \mathcal{A}^*y + \mathcal{B}^*z = c, \\
 & y \in \mathcal{Y},\ \ z \in \mathcal{Z},
\end{array}
\tag{4.1}
$$

where s and t are given nonnegative integers, $\mathcal{Y} := \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_s$, $\mathcal{Z} := \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_t$, $f(y_1,\dots,y_s) := \sum_{i=1}^s f_i(y_i)$, and $g(z_1,\dots,z_t) := \sum_{j=1}^t g_j(z_j)$. In order to simplify the notation, for any $y = (y_1,\dots,y_s) \in \mathcal{Y}$, denote $y_{\geq i} := (y_i,\dots,y_s)$ and $y_{\leq i} := (y_1,\dots,y_i)$; for any $z = (z_1,\dots,z_t) \in \mathcal{Z}$, denote $z_{\geq j} := (z_j,\dots,z_t)$ and $z_{\leq j} := (z_1,\dots,z_j)$.

For $i \in \{1,\dots,s\}$ and $j \in \{1,\dots,t\}$, we assume that $f_i : \mathcal{Y}_i \to \Re$ and $g_j : \mathcal{Y}_j \to \Re$ are convex functions with Lipschitz continuous gradients. Then, there exist positive semidefinite operators $\widehat{\Sigma}_{f_i}$ and $\widehat{\Sigma}_{g_j}$ such that for given $y_i' \in \mathcal{Y}_i$, $z_j' \in \mathcal{Z}_j$,

$$
\begin{aligned}
f_i(y_i) \leq \hat{f}_i(y_i, y_i') &:= f_i(y_i') + \langle \nabla f_i(y_i'), y_i - y_i' \rangle + \tfrac{1}{2}\|y_i - y_i'\|^2_{\widehat{\Sigma}_{f_i}}, \\
g_j(z_j) \leq \hat{g}_j(z_j, z_j') &:= g_j(z_j') + \langle \nabla g_j(z_j'), z_j - z_j' \rangle + \tfrac{1}{2}\|z_j - z_j'\|^2_{\widehat{\Sigma}_{g_j}}.
\end{aligned}
$$

Set

$$
\widehat{\Sigma}_f := \mathrm{Diag}(\widehat{\Sigma}_{f_1},\dots,\widehat{\Sigma}_{f_s}) \text{ and } \widehat{\Sigma}_g := \mathrm{Diag}(\widehat{\Sigma}_{g_1},\dots,\widehat{\Sigma}_{g_t}).
\tag{4.2}
$$

Then it holds that

$$
\begin{aligned}
f(y) \leq \hat{f}(y, y') &:= f(y') + \langle \nabla f(y'), y - y' \rangle + \tfrac{1}{2}\|y - y'\|^2_{\widehat{\Sigma}_f}, \\
g(z) \leq \hat{g}(z, z') &:= g(z') + \langle \nabla g(z'), z - z' \rangle + \tfrac{1}{2}\|z - z'\|^2_{\widehat{\Sigma}_g}.
\end{aligned}
$$

For any given parameter $\sigma > 0$, the majorized augmented Lagrangian function $\widehat{\mathcal{L}}_\sigma(y, z, x; y', z')$ is defined as (1.7) and the sGS based multiblock majorized ADMM with indefinite proximal terms (majorized sGS-iPADMM) is presented in Algorithm 2.

Algorithm 2 A majorized sGS-iPADMM

Let $\sigma > 0$ and $\tau \in (0, (1+\sqrt{5})/2)$ be given parameters. For $i \in \{1,\dots,s\}$ and $j \in \{1,\dots,t\}$, let $\mathcal{S}_i$, $\mathcal{T}_j$ be given self-adjoint, possibly indefinite, linear operators. Input $(y^0, z^0, x^0) \in \operatorname{dom} p \times \operatorname{dom} q \times \mathcal{X}$. For $k = 0, 1, 2, \dots$, perform the following steps.

Step 1a.: (Backward GS sweep) Compute for $i = s, \dots, 2$,

$$\bar{y}_i^k = \arg\min_{y_i} \widehat{\mathcal{L}}_\sigma(y_{\le i-1}^k, y_i, \bar{y}_{\ge i+1}^k, z^k, x^k; y^k, z^k) + \frac{1}{2}\|y_i - y_i^k\|_{\mathcal{S}_i}^2,$$

$$\text{and } y_1^{k+1} = \arg\min_{y_1} \widehat{\mathcal{L}}_\sigma(y_1, \bar{y}_{\ge 2}^k, z^k, x^k; y^k, z^k) + \tfrac{1}{2}\|y_1 - y_1^k\|_{\mathcal{S}_1}^2.$$

Step 1b.: (Forward GS sweep) Compute for $i = 1, \dots, s$,

$$y_i^{k+1} = \arg\min_{y_i} \widehat{\mathcal{L}}_\sigma(y_{\le i-1}^{k+1}, y_i, \bar{y}_{\ge i+1}^k, z^k, x^k; y^k, z^k) + \frac{1}{2}\|y_i - y_i^k\|_{\mathcal{S}_i}^2.$$

Step 1c.: (Backward GS sweep) Compute for $j = t, \dots, 2$,

$$\bar{z}_j^k = \arg\min_{z_j} \widehat{\mathcal{L}}_\sigma(y^{k+1}, z_{\le j-1}^k, z_i, \bar{z}_{\ge j+1}^k, x^k; y^k, z^k) + \frac{1}{2}\|z_j - z_j^k\|_{\mathcal{T}_j}^2,$$

$$\text{and } z_1^{k+1} = \arg\min_{z_1} \widehat{\mathcal{L}}_\sigma(y^{k+1}, z_1, \bar{z}_{\ge 2}^k, x^k; y^k, z^k) + \tfrac{1}{2}\|z_1 - z_1^k\|_{\mathcal{T}_1}^2.$$

Step 1d.: (Forward GS sweep) Compute for $j = 1, \dots, t$,

$$z_i^{k+1} = \arg\min_{z_j} \widehat{\mathcal{L}}_\sigma(y^{k+1}, z_{\le j-1}^{k+1}, z_j, \bar{z}_{\ge j+1}^k, x^k; y^k, z^k) + \frac{1}{2}\|z_j - z_j^k\|_{\mathcal{T}_j}^2.$$

Step 2.: Compute $x^{k+1} = x^k + \tau\sigma(\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^{k+1} - c)$.

Denote the following two linear operators:

$$\widetilde{\mathcal{M}} := \frac{1}{2}\widehat{\Sigma}_f + \sigma\mathcal{A}\mathcal{A}^* + \operatorname{Diag}(\mathcal{S}_1, \dots, \mathcal{S}_s), \;\; \widetilde{\mathcal{N}} := \frac{1}{2}\widehat{\Sigma}_g + \sigma\mathcal{B}\mathcal{B}^* + \operatorname{Diag}(\mathcal{T}_1, \dots, \mathcal{T}_t),$$

where for $i \in \{1,\dots,s\}$ and $j \in \{1,\dots,t\}$, $\mathcal{S}_i \succeq -\frac{1}{2}\widehat{\Sigma}_{f_i}$ and $\mathcal{T}_j \succeq -\frac{1}{2}\widehat{\Sigma}_{g_j}$ are self-adjoint linear operators such that the ith diagonal block operator of $\widetilde{\mathcal{M}}$ and jth diagonal block operator of $\widetilde{\mathcal{N}}$ are positive definite, i.e.,

$$\widetilde{\mathcal{M}}_{ii} = \frac{1}{2}\widehat{\Sigma}_{f_i} + \mathcal{S}_i + \sigma\mathcal{A}_i\mathcal{A}_i^* \succ 0 \quad \text{and} \quad \widetilde{\mathcal{N}}_{jj} = \frac{1}{2}\widehat{\Sigma}_{g_j} + \mathcal{T}_j + \sigma\mathcal{B}_j\mathcal{B}_j^* \succ 0.$$

Moreover, define

(4.3)
$$\mathcal{S} := \operatorname{Diag}(\mathcal{S}_1, \dots, \mathcal{S}_s) + \operatorname{sGS}(\widetilde{\mathcal{M}}) \quad \text{and} \quad \mathcal{T} := \operatorname{Diag}(\mathcal{T}_1, \dots, \mathcal{T}_t) + \operatorname{sGS}(\widetilde{\mathcal{N}}),$$

where $\operatorname{sGS}(\widetilde{\mathcal{M}}) := \widetilde{\mathcal{M}}_u \widetilde{\mathcal{M}}_d^{-1} \widetilde{\mathcal{M}}_u^*$ and $\operatorname{sGS}(\widetilde{\mathcal{N}}) := \widetilde{\mathcal{N}}_u \widetilde{\mathcal{N}}_d^{-1} \widetilde{\mathcal{N}}_u^*$. Here the notation $\widetilde{\mathcal{M}}_u$ ($\widetilde{\mathcal{M}}_d$) and $\widetilde{\mathcal{N}}_u$ ($\widetilde{\mathcal{N}}_d$) stand for the strictly upper triangular (diagonal) block operators of $\widetilde{\mathcal{M}}$ and $\widetilde{\mathcal{N}}$, respectively.

By using the same process in Li et al. [32] (see also Chen et al. [4]), the majorized sGS-iPADMM can be equivalently converted into the following 2-block majorized iPADMM:

$$\begin{cases} y^{k+1} = \arg\min\limits_{y\in\mathcal{Y}} \ p(y) + \langle y, \nabla f(y^k) + \mathcal{A}x^k\rangle \\ \qquad\qquad\qquad\qquad + \dfrac{\sigma}{2}\|\mathcal{A}^*y + \mathcal{B}^*z^k - c\|^2 + \dfrac{1}{2}\|y - y^k\|^2_{\widehat{\Sigma}_f + \mathcal{S}}, \\ z^{k+1} = \arg\min\limits_{z\in\mathcal{Z}} \ q(z) + \langle z, \nabla g(z^k) + \mathcal{B}x^k\rangle \\ \qquad\qquad\qquad\qquad + \dfrac{\sigma}{2}\|\mathcal{A}^*y^{k+1} + \mathcal{B}^*z - c\|^2 + \dfrac{1}{2}\|z - z^k\|^2_{\widehat{\Sigma}_g + \mathcal{T}}, \\ x^{k+1} := x^k + \tau\sigma(\mathcal{A}^*y^{k+1} + \mathcal{B}^*z^{k+1} - c), \end{cases}$$

where the operators $\widehat{\Sigma}_f$ and $\widehat{\Sigma}_g$ are defined by (4.2), the proximal terms $\mathcal{S}$ and $\mathcal{T}$ are defined by (4.3). It follows from the choice of $\mathcal{S}_i$ and $\mathcal{T}_j$ that Assumption 2.2 holds and

$$\frac{1}{2}\widehat{\Sigma}_f + \mathcal{S} + \sigma\mathcal{A}\mathcal{A}^* \succ 0 \quad \text{and} \quad \frac{1}{2}\widehat{\Sigma}_g + \mathcal{T} + \sigma\mathcal{B}\mathcal{B}^* \succ 0.$$

Therefore, directly from Theorem 3.1, we can get the following convergence results for our majorized sGS-iPADMM.

Proposition 4.1. *Let $\{u^k := (y^k, z^k, x^k)\}$ be the infinite sequence generated by the majorized sGS-iPADMM and proximal terms $\mathcal{S}$ and $\mathcal{T}$ be defined by* (4.3)*. Then, we have the following results:*

(a) *The sequence $\{(y^k, z^k)\}$ converges to an optimal solution of problem* (4.1) *and $\{x^k\}$ converges to an optimal solution of the dual of problem* (4.1)*.*
(b) *Suppose that the sequence $\{(y^k, z^k, x^k)\}$ converges to a KKT point $\bar{u} := (\bar{y}, \bar{z}, \bar{x})$ and the KKT mapping $\mathcal{R}$ is metrically subregular at $(\bar{u}, 0) \in \mathrm{gph}\mathcal{R}$. Then the sequence $\{u^k\}$ is linearly convergent to $\overline{u}$.*

5. Application II: The regularized logistic regression

In this section, we apply the majorized iPADMM to general regularized logistic regression in the following form:

$$\min_{y,y_0} \ f(y, y_0) + \varphi(y), \tag{5.1}$$

where $f : \Re^{n+1} \to \Re$ is the logistic loss function and $\varphi : \Re^n \to \Re \cup \{+\infty\}$ is a general convex Lasso regularizer. Specifically, the logistic loss function f takes the following form:

$$f(y, y_0) = \frac{1}{N}\sum_{i=1}^{N} \log(1 + \exp(-b_i(B_i^T y + y_0))), \tag{5.2}$$

where $B_i \in \Re^n$ are the predictor variables and $b_i \in \{1, -1\}$ are the responses, $i = 1, \ldots, N$. For notational convenience, set $\tilde{y} := [y; y_0] \in \Re^{n+1}$, $A_i := [-b_iB_i; -b_i] \in \Re^{n+1}$ and denote the gradient of f at $\tilde{y} \in \mathrm{dom}f$ by $\nabla f(\tilde{y})$.

Since the gradient ∇f is Lipschitz continuous on $\mathrm{dom}f$, we know that there exists a positive semidefinite matrix $\widehat{\Sigma}_f$ such that for any given $\tilde{y}\,' \in \Re^{n+1}$,

$$f(\tilde{y}) \le \hat{f}(\tilde{y}\,; \tilde{y}\,') := f(\tilde{y}\,') + \langle \nabla f(\tilde{y}\,'), \tilde{y} - \tilde{y}\,'\rangle + \frac{1}{2}\|\tilde{y} - \tilde{y}\,'\|^2_{\widehat{\Sigma}_f}.$$

Elementary calculations show that

$$\nabla f(\tilde{y}) = \frac{1}{N}\sum_{i=1}^{N}\frac{A_i\exp(A_i^T\tilde{y})}{1+\exp(A_i^T\tilde{y})},$$

$$\nabla^2 f(\tilde{y}) = \frac{1}{N}\sum_{i=1}^{N}A_iA_i^T\frac{\exp(A_i^T\tilde{y})}{(1+\exp(A_i^T\tilde{y}))^2} \preceq \frac{1}{4N}\sum_{i=1}^{N}A_iA_i^T.$$

Therefore, the proximal term can be chosen as

$$\widehat{\Sigma}_f := \frac{1}{4N}AA^T,\ \ A := [A_1,\ldots,A_N] \in \Re^{(n+1)\times N}. \tag{5.3}$$

In this section, we consider the logistic regression with three types of $\varphi(\cdot)$, namely, Lasso, fused Lasso, and constrained Lasso. Next, we reformulate these three types of regularized logistic regression into the framework of (1.1) and tailor the majorized iPADMM for each of them. Since our main purpose is to test the numerical performance of our majorized iPADMM, we omit the history and development of these regularized logistic regression models. To know more about these models, we refer to [5, 27, 28, 37, 45, 47].

5.1. **Lasso and fused Lasso logistic regression.** The fused Lasso method is introduced by Tibshirani et al. [47] to study the situation that the features have a natural order. In this case, for any given $\lambda_1 \geq 0, \lambda_2 \geq 0$, the function φ takes the following form:

$$\varphi(y) = \lambda_1\|y\|_1 + \lambda_2\|Fy\|_1, \tag{5.4}$$

where $F \in \Re^{(n-1)\times n}$ is the matrix defined by

$$Fy = [y_1 - y_2, y_2 - y_3, \ldots, y_{n-1} - y_n]^T\ \forall y \in \Re^n.$$

The definition (5.4) contains the Lasso regularizer as a special case if we take $\lambda_2 = 0$.

By introducing an auxiliary variable $z \in \Re^n$, we can reformulate the fused Lasso problem into the framework of (1.1), i.e.,

$$\begin{array}{ll}\min\limits_{y,y_0,z} & f(y,y_0) + \varphi(z)\\ \text{s.t.} & y - z = 0.\end{array} \tag{5.5}$$

The KKT system can be written as follows:

$$\nabla f(y,y_0) + [x;0] = 0,\ z - \Pr_{\varphi}(x+z) = 0,\ y - z = 0.$$

The function $\widehat{\mathcal{L}}_\sigma(\cdot)$ in (1.7) can be specifically written as

$$\widehat{\mathcal{L}}_\sigma(y,y_0,z,x;\tilde{y}') = \hat{f}(\tilde{y},\tilde{y}\,') + \varphi(z) + \langle y - z,\, x\rangle + \frac{\sigma}{2}\|y - z\|^2.$$

Consequently, the majorized iPADMM scheme for solving (5.5) can be described as follows:

$$\begin{cases} (y^{k+1}, y_0^{k+1}) = \arg\min\limits_{y,y_0}\ \widehat{\mathcal{L}}_\sigma(y,y_0,z^k,x^k;\tilde{y}^k) + \frac{1}{2}\|\tilde{y} - \tilde{y}^k\|_{\mathcal{S}}^2,\\ z^{k+1} = \arg\min\limits_{z}\ \varphi(z) + \frac{\sigma}{2}\|z - (y^{k+1} + x^k/\sigma)\|^2,\\ x^{k+1} = x^k + \tau\sigma(y^{k+1} - z^{k+1}),\end{cases} \tag{5.6}$$

where $\tau \in (0, (1+\sqrt{5})/2)$ and $\mathcal{S} = -\frac{1}{2}\widehat{\Sigma}_f + \mathrm{Diag}(0, \sigma r)$ with r being a given positive number. It is obvious that

$$\frac{1}{2}\widehat{\Sigma}_f + \mathcal{S} + \sigma \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \succ 0.$$

Based on the optimality conditions, we measure the accuracy of an approximate KKT point (y, y_0, z, x) via

$$\eta_{\mathrm{FL}} = \max\{\eta_P, \, \eta_D, \, \eta_C\},$$

where

$$\eta_P = \left\{ \frac{\|y - z\|}{1 + \|y\| + \|z\|} \right\}, \; \eta_D = \frac{\|\nabla f(y, y_0) + [x; 0]\|}{1 + \|\nabla f(y, y_0)\| + \|x\|}, \; \eta_C := \frac{\|z - \mathrm{Pr}_{\varphi}(x + z)\|}{1 + \|x\| + \|z\|}.$$

It is worth mentioning that, the solution z^{k+1} can be obtained by the following result which was first shown in [45] by using the subgradient technique and an alternative proof can be found in [37].

Lemma 5.1. *For any $\lambda_1, \lambda_2 \geq 0$, the optimal solution z^* of*

$$\min_{z} \lambda_1 \|z\|_1 + \lambda_2 \|Fz\|_1 + \frac{1}{2}\|z - v\|^2,$$

can be described as

$$z^* = \mathrm{sgn}(\mathrm{z}_0) \cdot \max(|z_0| - \lambda_1, 0),$$

where $z_0 := \arg\min_{z} \lambda_2 \|Fz\|_1 + \frac{1}{2}\|z - v\|^2$.

Though there is no closed-form expression of z_0 when $\lambda_2 > 0$, the algorithm[1] presented in [7] can be used to get z_0 efficiently.

The function φ defined by (5.4) is a piecewise-linear function. Therefore, it holds from [49, Theorem 4.3] and [10, Theorem 3.1] that the KKT mapping $\mathcal{R}$ corresponding to (5.5) is metrically subregular at any KKT point for the origin when $\lambda_1 + \lambda_2 > 0$. Therefore, directly from Theorem 3.1, we can get the following results.

Proposition 5.1. *Let $\lambda_1 + \lambda_2 > 0$ and $\{u^k := (y^k, y_0^k, z^k, x^k)\}$ be the infinite sequence generated by the majorized iPADMM scheme* (5.6)*. Then the sequence $\{u^k\}$ converges linearly to a KKT point of* (5.5)*.*

5.2. **Constrained logistic regression.** Inspired by the work of James et al. [27], we consider the function φ in the following form:

$$\varphi(y) = \lambda \|y\|_1 + \delta_{\mathcal{D}}(y),$$

where $\mathcal{D} := \{y |\, Dy \geq d\}$, $D \in \Re^{m \times n}$, $d \in \Re^m$, function $\delta_{\mathcal{D}}(\cdot)$ is an indicator function of convex set $\mathcal{D}$. We can rewrite (5.1) as

$$\begin{array}{cl} \min\limits_{y, y_0, v, w} & f(y, y_0) + \lambda \|v\|_1 + \delta_{\Re^m_+}(w) \\ \text{s.t.} & Dy - w = d, \\ & y - v = 0. \end{array} \tag{5.7}$$

The KKT conditions are given by

$$\nabla f(y, y_0) + [D^T \xi + \zeta; 0] = 0, \; v - \mathrm{Pr}_{\lambda\|\cdot\|_1}(\zeta + v) = 0,$$
$$w - \Pi_{\Re^m_+}(\xi + w) = 0, \; Dy - w = d, \; y - v = 0,$$

[1]The code is available at `https://www.gipsa-lab.grenoble-inp.fr/∼laurent.condat/software.html`

and the majorized augmented Lagrangian function $\widehat{\mathcal{L}}_\sigma(\cdot)$ in (1.7) can be specifically written as

$$\widehat{\mathcal{L}}_\sigma(y,y_0,w,v,\xi,\zeta;\tilde{y}') = \hat{f}(\tilde{y},\tilde{y}\,') + \lambda\|v\|_1 + \delta_{\Re^m_+}(w) + \langle Dy-w-d,\,\xi\rangle + \langle y-v,\,\zeta\rangle + \frac{\sigma}{2}\|Dy-w-d\|^2 + \frac{\sigma}{2}\|y-v\|^2.$$

Therefore, we can solve (5.7) via the following iterative scheme:

$$(5.8)\qquad \begin{cases} (y^{k+1},y_0^{k+1}) = \arg\min\limits_{y,y_0} \widehat{\mathcal{L}}_\sigma(y,y_0,w^k,v^k,\xi^k,\zeta^k;\tilde{y}^k) + \frac{1}{2}\|\tilde{y}-\tilde{y}^k\|^2_{\mathcal{S}},\\ w^{k+1} = \max\{Dy^{k+1}-d+\xi^k/\sigma,\mathbf{0}\},\\ v^{k+1} = \arg\min\limits_{v} \lambda\|v\|_1 + \frac{\sigma}{2}\|v-(y^{k+1}+\zeta^k/\sigma)\|^2,\\ \xi^{k+1} = \xi^k + \tau\sigma(Dy^{k+1}-w^{k+1}-d),\\ \zeta^{k+1} = \zeta^k + \tau\sigma(y^{k+1}-v^{k+1}), \end{cases}$$

where $\tau\in(0,(1+\sqrt{5})/2)$ and $\mathcal{S} = -\frac{1}{2}\widehat{\Sigma}_f + \mathrm{Diag}(0,\sigma r)$ with r being a given positive number. It is obvious that

$$\frac{1}{2}\widehat{\Sigma}_f + \mathcal{S} + \sigma\begin{pmatrix} D^TD+I & 0\\ 0 & 0\end{pmatrix} \succ 0.$$

Based on the optimality conditions, we measure the accuracy of an approximate KKT point (y,y_0,u,v,ξ,ζ) via

$$\eta_{\mathrm{CL}} = \max\{\eta_P,\,\eta_D,\,\eta_C\},$$

where

$$\eta_P = \max\left\{\frac{\|Dy-w-d\|}{1+\|Dy\|+\|w\|+\|d\|},\ \frac{\|y-v\|}{1+\|y\|+\|v\|}\right\},\ \eta_D = \frac{\|\nabla f(y,y_0)+[D^T\xi+\zeta;0]\|}{1+\|\nabla f(y,y_0)\|+\|D^T\xi\|+\|\zeta\|},$$

$$\eta_{\mathrm{C}} := \max\left\{\frac{\|v-\mathrm{Pr}_{\lambda\|\cdot\|_1}(\zeta+v)\|}{1+\|\zeta\|+\|v\|},\ \frac{\|u-\Pi_{\Re^m_+}(\xi+w)\|}{1+\|\xi\|+\|w\|}\right\}.$$

It follows from [49, Theorem 4.3] and [10, Theorem 3.1] that the KKT mapping $\mathcal{R}$ corresponding to (5.7) is also metrically subregular at any KKT point for the origin when $\lambda_1>0$ and $\mathcal{D}$ is nonempty. Therefore, directly from Theorem 3.1, we can get the following results.

Proposition 5.2. *Suppose that $\lambda_1>0$ and $\mathcal{D}$ is nonempty. Let $\{u^k := (y^k,y_0^k,w^k,v^k,\xi^k,\zeta^k)\}$ be the infinite sequence generated by the majorized iPADMM scheme (5.8). Then the sequence $\{u^k\}$ converges linearly to a KKT point of (5.7).*

5.3. **Numerical experiments.** In this subsection, we evaluate the performance of majorized iPADMM for solving Lasso logistic regression, fused Lasso logistic regression(5.5), and constrained Lasso logistic regression (5.7), respectively. All computational results are obtained by running Matlab R2018b on Windows 10 (Intel Core i5-7300U @ 2.60GHz 8GB RAM).

Consider the following two self-adjoint linear operators:

$$\mathcal{S}_0 = \sigma\begin{pmatrix} 0 & 0\\ 0 & r\end{pmatrix} \quad\text{and}\quad \mathcal{S} = -\frac{1}{2}\widehat{\Sigma}_f + \sigma\begin{pmatrix} 0 & 0\\ 0 & r\end{pmatrix}.$$

Note that the proximal term $\mathcal{S}$ may not be a positive semidefinite operator. In the subsequent discussions, we call the majorized iPADMM scheme with $\mathcal{S}_0$ the "majorized sPADMM". In all tests, we set $r=10^{-6}$ and choose $\widehat{\Sigma}_f$ as in (5.3) unless otherwise specified. It is worth emphasizing here that the algorithm proposed in [31] with condition (1.10) is exactly the majorized sPADMM when f is given by (5.2) and $\Sigma_f = 0$.

5.3.1. *Lasso and fused Lasso logistic regression: LIBSVM dataset.* In this part, we apply the majorized iPADMM to Lasso/fused Lasso logistic regression and test its performance with data sets: `a8a`, `a9a`, `colon-cancer`, `duke breast-cancer` (`duke-BC`), `rcv1_train`, `news20`. These data sets are obtained from the LIBSVM datasets [3, 35].

Notice that the subproblem corresponding to (y, y_0) can be reformulated as a linear system of equations. In order to exactly and efficiently solve this subproblem, we follow the strategies proposed in [29, Section 3.3]. Specifically, for `a8a, a9a`, i.e., $n \ll N$, n is moderate, the linear system of equations can be solved by computing the Cholesky factorization of the coefficient matrix $H := \widehat{\Sigma}_f + \sigma \mathrm{Diag}(I, 0) + \mathcal{S}$; for `conlon-cancer, duke-BC`, i.e., $n \gg N$, N is moderate, the Sherman-Morrison-Woodbury formula [24] can be used to get the inverse matrix of H by inverting a much smaller invertible $N \times N$ matrix and the Cholesky factorization will be further applied to the smaller matrix; for `rcv1_train, news20`, i.e., both n and N are large, by taking $\widehat{\Sigma}_f := \sum_{i=1}^{K} \mu_i P_i P_i^T$, where $\sqrt{\mu_1}, \ldots, \sqrt{\mu_K}$ are the K largest singular values of A and $P_1, \ldots, P_K$ are the corresponding left-eigenvectors, the Sherman-Morrison-Woodbury formula can also be used. For more details and sophisticated techniques, we refer the interested reader to [29] and the references therein.

In order to evaluate the numerical performance, we also report majorized sPADMM and a commonly used accelerated proximal gradient (APG) algorithm [2, 40] as implemented in [36, 37]. In the comparison, we terminate APG if the objective value obj_A obtained by APG satisfies

$$(\mathrm{obj}_A - \mathrm{obj}_M)/|\mathrm{obj}_M| \leq \varepsilon,$$

where obj_M is the objective value obtained by majorized iPADMM. In the numerical tests, the regularization parameters are chosen as follows:

(1) for Lasso logistic regression,

$$\lambda_1 = \frac{\gamma}{N} \|B^T b\|_\infty \quad \text{and} \quad \lambda_2 = 0,$$

(2) for Lasso logistic regression,

$$\lambda_1 = \lambda_2 = \frac{\gamma}{N} \|B^T b\|_\infty,$$

where N is the size of the sample and $0 < \gamma < 1$. The majorized iPADMM and majorized sPADMM will be terminated when $\eta_{FL} < 10^{-6}$ or the maximum iteration number 50,000 is reached.

Now we are ready to report the comparison results. Tables 1 and 2 report the comparison between the performance of majorized iPADMM, majorized sPADMM, and APG for solving the Lasso logistic regression and the fused Lasso logistic regression, respectively. From these two tables, we can observe that the numerical performance of the majorized iPADMM outperforms the other two methods for all cases. Besides, we can also see that the majorized iPADMM brings about 40%–50% reduction in the number of iterations needed as compared with the majorized sPADMM except cases that `a8a` and `a9a` with $\gamma = 10^{-2}$ in Table 2.

TABLE 1. Comparison between the performance of majorized iPADMM (MiPA), majorized sPADMM (MsPA), and APG for solving Lasso logistic regression. "nnz" denotes the number of nonzeros in the solution z generated by majorized iPADMM . "IterNum" denotes the number of iterations. $L_C := \lambda_{\max}(BB^T)/N$.

				IterNum					Time(sec)				
probName N $\mid$ n	L_C	γ	nnz	MiPA $\tau = 1.618$	MsPA $\tau = 1.618$	MiPA $\tau = 1$	MsPA $\tau = 1$	APG	MiPA $\tau = 1.618$	MsPA $\tau = 1.618$	MiPA $\tau = 1$	MsPA $\tau = 1$	APG
a8a	1.4e+5	10^{-2}	18	42	77	46	80	273	0.08	0.16	0.10	0.16	0.82
22696, 123		10^{-3}	49	104	248	105	221	514	0.22	0.50	0.22	0.46	1.54
a9a	2.0e+5	10^{-2}	18	41	77	45	79	261	0.12	0.20	0.12	0.20	1.11
32561, 123		10^{-3}	47	117	235	109	258	545	0.30	0.62	0.29	0.67	2.31
colon-cancer	1.9e+4	10^{-2}	31	485	833	551	874	5287	0.29	0.46	0.32	0.50	0.95
62, 2000		10^{-3}	36	5593	11374	5510	11066	18288	2.97	6.01	2.96	5.93	3.25
duke	1.1e+5	10^{-2}	29	1165	1830	1268	1907	17523	1.72	2.70	1.88	2.83	11.84
44, 7129		10^{-3}	31	6775	13081	6599	13540	20000	9.95	19.23	9.70	19.80	13.49
rcv1_train	4.5e+2	10^{-1}	27	140	207	173	232	1626	1.09	1.61	1.34	1.80	10.69
20242, 47236		10^{-2}	285	986	1779	1035	1808	4662	7.52	13.46	8.42	13.86	30.80
news20	1.2e+3	10^{-1}	46	224	278	274	313	1312	40.06	49.73	49.14	56.04	87.75
19996, 1355191		10^{-2}	420	1202	1951	1301	2061	7563	215.77	350.87	233.22	373.14	515.63

TABLE 2. Same as Table 1 but for fused Lasso logistic regression

				IterNum					Time(sec)				
probName N $\mid$ n	L_C	γ	nnz	MiPA $\tau = 1.618$	MsPA $\tau = 1.618$	MiPA $\tau = 1$	MsPA $\tau = 1$	APG	MiPA $\tau = 1.618$	MsPA $\tau = 1.618$	MiPA $\tau = 1$	MsPA $\tau = 1$	APG
a8a	1.4e+05	1.0e-02	13	37	48	38	50	416	0.10	0.12	0.10	0.15	1.46
22696, 123		1.0e-03	43	82	168	84	174	1023	0.20	0.38	0.20	0.40	3.52
a9a	2.0e+05	1.0e-02	12	37	46	38	47	337	0.12	0.13	0.11	0.14	1.67
32561, 123		1.0e-03	41	83	171	85	163	1054	0.25	0.50	0.25	0.46	5.11
colon-cancer	1.9e+04	1.0e-02	52	319	542	344	581	5244	0.22	0.36	0.23	0.38	1.47
62, 2000		1.0e-03	69	2010	3891	1965	3906	10639	1.27	2.36	1.22	2.36	2.94
duke	1.1e+05	1.0e-02	61	750	1271	712	1314	18874	1.35	2.22	1.27	2.34	19.20
44, 7129		1.0e-03	68	2287	6683	1890	6992	20000	3.98	11.59	3.31	12.13	19.85
rcv1_train	4.5e+02	1.0e-02	123	650	1099	694	1106	3155	6.02	10.27	6.43	10.61	27.33
20242, 47236		1.0e-03	744	2533	4835	2579	4876	4717	23.33	44.51	24.25	44.84	41.29
news20	1.2e+03	1.0e-02	264	795	1287	859	1319	5507	176.15	284.87	190.14	291.77	593.13
19996, 1355191		1.0e-03	2973	2382	4362	2485	4436	11032	525.58	965.55	556.19	981.70	1186.68

TABLE 3. Comparison between the performance of majorized iPADMM (MiPA), majorized sPADMM (MsPA), L-majorized iPADMM (LiPA), and L-majoirized sPADMM (LsPA); "IterNum" denotes the number of iterations. $L_C := \lambda_{\max}(BB^T)/N$. All results are averaged over 10 instances.

			IterNum				Time(sec)				η_{CL}			
probName N | n	L_C	γ	MiPA	MsPA	LiPA	LsPA	MiPA	MsPA	LiPA	LsPA	MiPA	MsPA	LiPA	LsPA
30,50,20	4.90	1.0e-02	283.1	318.3	1395.8	2798.5	0.0	0.0	0.1	0.2	9.83e-06	9.81e-06	9.96e-06	9.94e-06
		1.0e-03	1278.5	2247.1	13627.6	23261.8	0.1	0.2	1.0	1.7	9.97e-06	9.96e-06	9.98e-06	1.21e-05
		1.0e-04	8126.3	16111.9	48259.2	50000.0	0.6	1.2	3.4	3.5	1.00e-05	1.00e-05	2.29e-05	8.50e-05
50,100,60	5.61	1.0e-02	503.4	528.6	1339.8	2301.4	0.2	0.2	0.3	0.6	9.91e-06	9.92e-06	9.93e-06	9.94e-06
		1.0e-03	1557.9	2336.8	13401.5	26898.2	0.4	0.6	3.0	6.0	9.98e-06	9.98e-06	1.00e-05	1.00e-05
		1.0e-04	8430.0	16111.2	48854.3	50000.0	1.9	3.6	10.9	11.1	1.00e-05	1.00e-05	1.62e-05	5.80e-05
50,200,30	8.75	1.0e-02	584.9	592.5	2951.0	5209.5	0.4	0.4	1.7	3.0	9.91e-06	9.88e-06	9.95e-06	9.99e-06
		1.0e-03	1480.0	2169.1	24417.8	42413.7	0.9	1.2	13.7	23.9	9.97e-06	9.98e-06	1.00e-05	1.30e-05
		1.0e-04	6321.7	11632.8	50000.0	50000.0	3.6	6.7	28.6	28.6	1.00e-05	9.99e-06	3.23e-05	7.81e-05
50,500,10	16.80	1.0e-02	672.7	694.8	8838.6	16681.3	3.1	3.3	40.1	75.3	9.90e-06	9.88e-06	9.98e-06	1.00e-05
		1.0e-03	1407.1	2280.6	47452.0	50000.0	6.4	10.3	213.5	225.8	9.96e-06	9.96e-06	1.56e-05	4.82e-05
		1.0e-04	6761.7	13456.7	50000.0	50000.0	31.0	62.4	229.9	229.2	1.00e-05	9.97e-06	6.64e-05	1.26e-04

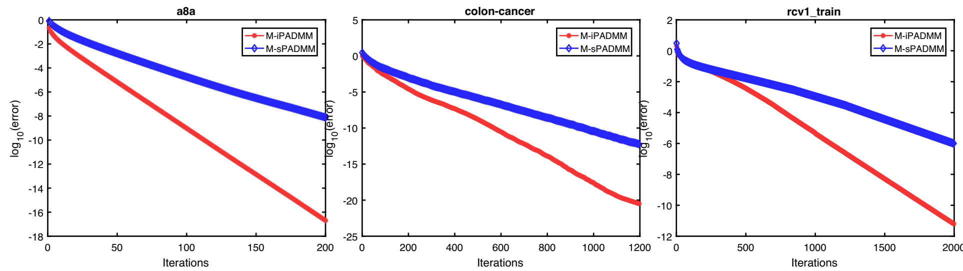

FIGURE 1. Comparison between the performance of our majorized iPADMM (M-iPADMM) and the majorized sPADMM (M-sPADMM) for solving Lasso logistic regression on datasets `a8a`($\gamma =$ 1e-3), `colon-cancer` (γ =1e-3), and `rcv1_train` (γ = 1e-2). error $:= \|u^k - \bar{u}\|_{\mathcal{M}}$, $\tau = 1.618$.

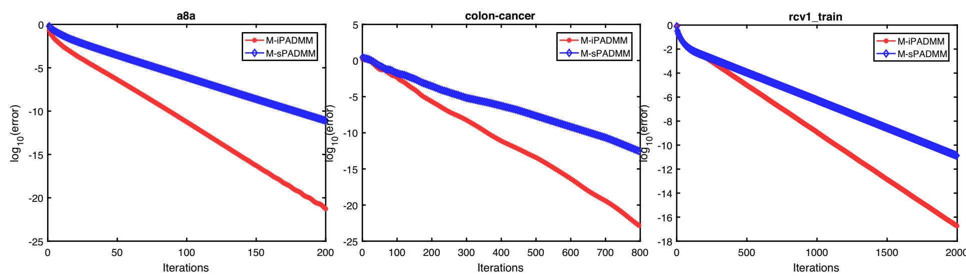

FIGURE 2. Same as Figure 1 but for fused Lasso logistic regression.

Figures 1 and 2 show that the sequence generated by majorized iPADMM converges to a KKT point approximately as a linear rate. This is consistent with Theorem 3.1 and Proposition 5.1. Moreover, we can also observe that the indefinite proximal terms can improve the numerical performance. The majorized iPADMM brings about 35%– 60% reduction in the number of iterations needed as compared with the majorized sPADMM, when one chooses $\|u^k - \bar{u}\|_{\mathcal{M}} \leq 10^{-6}$.

In summary, the linearly convergent majorized iPADMM proposed in this paper for solving Lasso/fused Lasso logistic regression is superior to both the majorized sPADMM which is exactly the method studied in [31] with condition (1.10) and the accelerated proximal gradient method.

5.3.2. *Constrained logistic regression: Synthetic data.* This part tests the performance of the majorized iPADMM for the constrained logistic regression by using synthetic data. The data $B \in \Re^{n\times N}$, $D \in \Re^{m\times n}$, and $d \in \Re^m$ are generated by the standard normal distribution. This example can also be used to illustrate the claim mentioned in the introduction that our majorized iPADMM with the positive semidefinite operator $\widehat{\Sigma}_f$ outperforms that with the Lipschitz constant $\lambda_{\max}(\widehat{\Sigma}_f)$ (denoted as L-majorized ADMM).

We compare our majorized iPADMM with majorized sPADMM, L-majorized ADMM with an indefinite proximal term $(\mathrm{Diag}(-\frac{1}{2}\lambda_{\max}(\widehat{\Sigma}_f)I, \sigma r))$, and L-majorized sPADMM. All the algorithms will be terminated when $\eta_{CL} \leq 10^{-5}$

or they reach the maximum number of iterations (5000 iterations for all the algorithms). In this test, we choose the regularization parameter $\lambda = \gamma\|B^T b\|_\infty/N$, where $0 < \gamma < 1$.

Table 3 reports the number of iterations required, runtime as well as the relative KKT residual η_{CL} of four different methods with $\tau = 1.618$. From the table, we can see that our majorized iPADMM outperforms all the other three methods. In each case, the majorized iPADMM can sometimes bring about 40% reduction in the number of iterations needed for convergence as compared with the majorized sPADMM. Note that though the size of each scenario is small, some cases still cannot be solved within the maximum number of iterations by using L-majorized iPADMM and L-majorized sPADMM. This table can also be used to illustrate the advantage of using the positive semidefinite operator $\widehat{\Sigma}_f$ over the Lipschitz constant $\lambda_{\max}(\widehat{\Sigma}_f)$. Figure 3 shows that the sequence generated by majorized iPADMM converges to a KKT point approximately as a linear rate. This coincides with the result presented in Theorem 3.1 and Proposition 5.2.

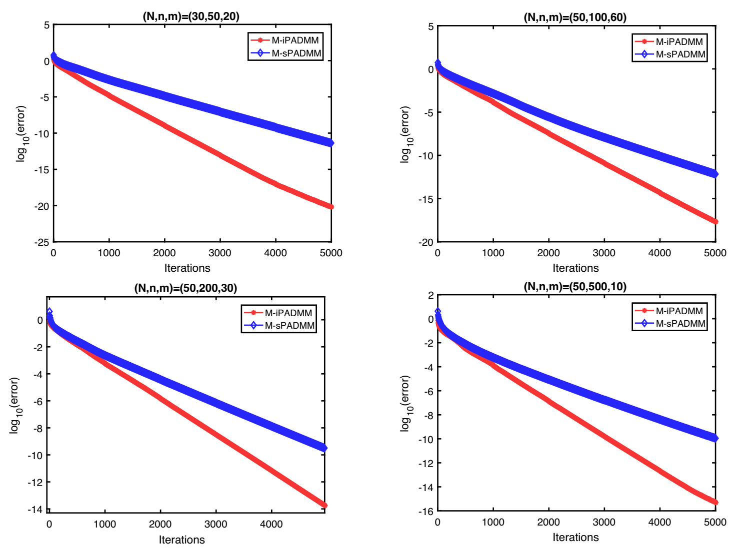

FIGURE 3. Comparison between the performance of our majorized iPADMM (M-iPADMM) and the majorized sPADMM (M-sPADMM) on synthetic data. error $:= \|u^k - \bar{u}\|_{\mathcal{M}}$, $\tau = 1.618$. All results are averaged over 10 instances.

6. CONCLUSION

In this paper, we have established the linear rate convergence of the majorized ADMM with indefinite proximal terms for solving the 2-block linearly constrained convex composite optimization problem under a metric subregularity assumption. Numerical results on three types of regularized logistic regression have been given to evaluate the effectiveness of the 2-block majorized ADMM with indefinite proximal terms. From these results, we can see that, for most cases, the majorized ADMM with indefinite proximal terms can bring about 30%–50% reduction in the number of iterations needed for convergence as compared with the majorized ADMM with semiproximal terms. Moreover, the proposed algorithm exhibits linear convergence for all the numerical examples.

Strongly motivated by the numerical performance of the symmetric Gauss-Seidel based ADMM for solving multiblock convex composite quadratic programming, we also proved the linear rate of convergence of a symmetric Gauss-Seidel based majorized ADMM with indefinite proximal terms by building its equivalence to the 2-block majorized ADMM with specially constructed proximal terms (possibly indefinite). This will greatly facilitate the future exploration of the multiblock general linear/nonlinear models. We leave this topic as our future work.

Acknowledgments

We are grateful to two anonymous referees for their helpful suggestions and comments, which have helped us improve the quality of this paper substantially. We also would like to thank Professor Defeng Sun for his creative suggestions on the introduction of Lemma 3.2.

Appendix A. Proofs of the Lemmas

A.1. **Proof of Lemma 3.1.**

Proof. The optimality condition for (1.9a) is

$$0 \in \partial p(y^{k+1}) + \nabla f(y^k) + \mathcal{A}x^k + \sigma\mathcal{A}(\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^k - c) + (\mathcal{S} + \widehat{\Sigma}_f)(y^{k+1} - y^k). \tag{A.1}$$

It follows from (1.9c) that

$$x^k + \sigma(\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^k - c) = \tau^{-1}(x^{k+1} - x^k) + x^k - \sigma\mathcal{B}^*(z^{k+1} - z^k). \tag{A.2}$$

From Proposition 2.1, we know that there exists $W_f^k \in \mathrm{conv}\partial^2 f[y^{k+1}, y^k]$ such that

$$\nabla f(y^k) - \nabla f(y^{k+1}) = W_f^k(y^k - y^{k+1}).$$

Substituting the above equation and (A.2) into (A.1), we get

$$0 \in \partial p(y^{k+1}) + \nabla f(y^{k+1}) + \mathcal{A}[x^k + \tau^{-1}(x^{k+1} - x^k) - \sigma\mathcal{B}^*(z^{k+1} - z^k)] + \mathcal{S}^k(y^{k+1} - y^k),$$

where $\mathcal{S}^k := \mathcal{S} + \widehat{\Sigma}_f - W_f^k$. Thus, it holds that

$$y^{k+1} = \mathrm{Pr}_p\{y^{k+1} - \nabla f(y^{k+1}) - \mathcal{A}[x^k + \tau^{-1}(x^{k+1} - x^k) - \sigma\mathcal{B}^*(z^{k+1} - z^k)] - \mathcal{S}^k(y^{k+1} - y^k)\}.$$

Similarly, there exists $W_g^k \in \mathrm{conv}\partial^2 g(z^{k+1}, z^k)$ such that $\nabla g(z^k) - \nabla g(z^{k+1}) = W_g^k(z^k - z^{k+1})$ and

$$z^{k+1} = \mathrm{Pr}_q\{z^{k+1} - \nabla g(z^{k+1}) - \mathcal{B}[\tau^{-1}(x^{k+1} - x^k) + x^k] - \mathcal{T}^k(z^{k+1} - z^k)\},$$

where $\mathcal{T}^k := \mathcal{T} + \widehat{\Sigma}_g - W_g^k$. Since the Moreau-Yosida proximal mappings $\mathrm{Pr}_p(\cdot)$ and $\mathrm{Pr}_q(\cdot)$ are globally Lipschitz continuous, one has that for any $k \geq 1$,

(A.3)

$$\begin{aligned}
&\|\mathcal{R}(u^{k+1})\|^2 \\
&\leq \|\mathcal{S}^k(y^{k+1} - y^k) - \sigma \mathcal{A}\mathcal{B}^*(z^{k+1} - z^k) + (\tau^{-1} - 1)\mathcal{A}(x^{k+1} - x^k)\|^2 \\
&\quad + \|\mathcal{T}^k(z^{k+1} - z^k) + (\tau^{-1} - 1)\mathcal{B}(x^{k+1} - x^k)\|^2 + \|(\tau\sigma)^{-1}(x^{k+1} - x^k)\|^2 \\
&\leq 3\lambda_{\max}^2(|\mathcal{S}^k|)\|y^{k+1} - y^k\|^2 + 3\sigma\lambda_{\max}(\mathcal{A}^*\mathcal{A})\|(z^{k+1} - z^k)\|_{\sigma\mathcal{B}\mathcal{B}^*}^2 \\
&\quad + 2\lambda_{\max}^2(|\mathcal{T}^k|)\|z^{k+1} - z^k\|^2 + \|(\tau\sigma)^{-1}(x^{k+1} - x^k)\|^2 \\
&\quad + (1 - \tau^{-1})^2[2\lambda_{\max}(\mathcal{B}^*\mathcal{B})\|(x^{k+1} - x^k)\|^2 + 3\lambda_{\max}(\mathcal{A}^*\mathcal{A})\|(x^{k+1} - x^k)\|^2].
\end{aligned}$$

Next, we estimate upper bounds of $\lambda_{\max}(\mathcal{S}^k)$ and $\lambda_{\max}(\mathcal{T}^k)$, respectively. It follows from $\Sigma_f \preceq W_f^k \preceq \widehat{\Sigma}_f$ and $\Sigma_g \preceq W_g^k \preceq \widehat{\Sigma}_g$ that

$$-\frac{1}{2}\widehat{\Sigma}_f \preceq \mathcal{S}^k \preceq \mathcal{S} + \widehat{\Sigma}_f \quad \text{and} \quad -\frac{1}{2}\widehat{\Sigma}_g \preceq \mathcal{T}^k \preceq \mathcal{T} + \widehat{\Sigma}_g, \ \forall k \geq 1.$$

Then

$$\begin{aligned}
&\lambda_{\min}(\mathcal{S}^k) \geq -\tfrac{1}{2}\lambda_{\max}(\widehat{\Sigma}_f) \ \text{ and } \ \lambda_{\max}(\mathcal{S}^k) \leq \lambda_{\max}(\mathcal{S} + \tfrac{1}{2}\widehat{\Sigma}_f) + \tfrac{1}{2}\lambda_{\max}(\widehat{\Sigma}_f), \\
&\lambda_{\min}(\mathcal{T}^k) \geq -\tfrac{1}{2}\lambda_{\max}(\widehat{\Sigma}_g) \ \text{ and } \ \lambda_{\max}(\mathcal{T}^k) \leq \lambda_{\max}(\mathcal{T} + \tfrac{1}{2}\widehat{\Sigma}_g) + \tfrac{1}{2}\lambda_{\max}(\widehat{\Sigma}_g),
\end{aligned}$$

and consequently, for any $k \geq 1$, one has

$$\|\mathcal{S}^k\|_2 \leq \lambda_{\max}(\mathcal{S} + \tfrac{1}{2}\widehat{\Sigma}_f) + \tfrac{1}{2}\lambda_{\max}(\widehat{\Sigma}_f), \ \|\mathcal{T}^k\|_2 \leq \lambda_{\max}(\mathcal{T} + \tfrac{1}{2}\widehat{\Sigma}_g) + \tfrac{1}{2}\lambda_{\max}(\widehat{\Sigma}_g).$$

By substituting the above two inequalities into (A.3), we can get (3.3). This completes the proof. □

A.2. **Proof of Lemma 3.2.**

Proof. For any $\varepsilon \in \Re$, define

$$h_\varepsilon(x) := h(x) + \frac{\varepsilon^2}{2}\|x\|^2 \ \forall\, x \in \mathcal{X}.$$

Then, similar to the proof of [39, Theorem 2.1.5], one has

$$\langle \nabla h_\varepsilon(x) - \nabla h_\varepsilon(\bar{x}),\, x - \bar{x}\rangle \geq \|\nabla h_\varepsilon(x) - \nabla h_\varepsilon(\bar{x})\|_{(\mathcal{P}+\varepsilon^2 I)^{-1}}^2 \ \forall\, \varepsilon \neq 0.$$

Consequently, for any $\varepsilon \neq 0$, it holds that

$$\begin{aligned}
&\langle \nabla h_\varepsilon(x) - \nabla h_\varepsilon(\bar{x}), y - \bar{x}\rangle \\
&\geq \|\nabla h_\varepsilon(x) - \nabla h_\varepsilon(\bar{x})\|_{(\mathcal{P}+\varepsilon^2\mathcal{I})^{-1}}^2 + \langle \nabla h_\varepsilon(x) - \nabla h_\varepsilon(\bar{x}),\, y - x\rangle \\
&= \|(\mathcal{P} + \varepsilon^2\mathcal{I})^{-1/2}(\nabla h_\varepsilon(x) - \nabla h_\varepsilon(\bar{x})) + \tfrac{1}{2}(\mathcal{P} + \varepsilon^2 I)^{1/2}(y - x)\|^2 - \tfrac{1}{4}\|x - y\|_{(\mathcal{P}+\varepsilon^2\mathcal{I})}^2 \\
&\geq -\tfrac{1}{4}\|x - y\|_{(\mathcal{P}+\varepsilon^2\mathcal{I})}^2.
\end{aligned}$$

This together with the definition of h implies that

$$\langle \nabla h(x) - \nabla h(\bar{x}), y - \bar{x}\rangle + \langle \varepsilon^2(x - \bar{x}), y - \bar{x}\rangle \geq -\frac{1}{4}\|x - y\|_{(\mathcal{P}+\varepsilon^2\mathcal{I})}^2, \ \forall \varepsilon \neq 0.$$

Therefore, by taking limits on both sides of the above inequality for $\varepsilon \to 0$, we complete the proof. □

A.3. **Proof of Lemma 3.3.**

Proof. By the first order optimality conditions of (1.9a) and (1.9b), one has

$$\begin{cases} 0 \in \partial p(y^{k+1}) + \nabla f(y^k) + (\widehat{\Sigma}_f + \mathcal{S})(y^{k+1} - y^k) + \mathcal{A}(x^k + \sigma(\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^k - c)), \\ 0 \in \partial q(z^{k+1}) + \nabla g(z^k) + (\widehat{\Sigma}_g + \mathcal{T})(z^{k+1} - z^k) + \mathcal{B}(x^k + \sigma r^{k+1}). \end{cases}$$

Since $(\bar{y}, \bar{z}, \bar{x})$ is a KKT point, it holds that

$$\begin{cases} 0 \in \partial p(\bar{y}) + \nabla f(\bar{y}) + \mathcal{A}\bar{x}, \\ 0 \in \partial q(\bar{z}) + \nabla g(\bar{z}) + \mathcal{B}\bar{x}. \end{cases}$$

It follows from the maximal monotonicity of ∂p that

$$\begin{aligned} 0 \leq & \langle -\mathcal{A}(x^k + \sigma(\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^k - c)) + \mathcal{A}\bar{x},\, y^{k+1} - \bar{y} \rangle \\ & + \langle \nabla f(\bar{y}) - \nabla f(y^k), y^{k+1} - \bar{y} \rangle \\ & - \langle (\widehat{\Sigma}_f + \mathcal{S})(y^{k+1} - y^k), y^{k+1} - \bar{y} \rangle. \end{aligned}$$

Thus, by reorganizing the above inequality and Lemma 3.2, one has
(A.4)

$$\begin{aligned} & \langle \bar{x} - (x^k + \sigma(\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^k - c)), \mathcal{A}^*(y^{k+1} - \bar{y}) \rangle - \langle (\widehat{\Sigma}_f + \mathcal{S})(y^{k+1} - y^k), y^{k+1} - \bar{y} \rangle \\ & \geq \langle \nabla f(y^k) - \nabla f(\bar{y}), y^{k+1} - \bar{y} \rangle \geq -\tfrac{1}{4}\|y^{k+1} - y^k\|^2_{\widehat{\Sigma}_f}. \end{aligned}$$

Similarly, by using the maximal monotonicity of ∂q and Lemma 3.2, it holds that

$$\begin{aligned} & \langle \bar{x} - (x^k + \sigma r^{k+1}),\, \mathcal{B}^*(z^{k+1} - \bar{z}) \rangle - \langle (\widehat{\Sigma}_g + \mathcal{T})(z^{k+1} - z^k), z^{k+1} - \bar{z} \rangle \\ & \geq -\tfrac{1}{4}\|z^{k+1} - z^k\|^2_{\widehat{\Sigma}_g}. \end{aligned} \tag{A.5}$$

By adding (A.4) and (A.5) together, we have
(A.6)

$$\begin{aligned} & \Delta_k + \langle (\widehat{\Sigma}_f + \mathcal{S})(y^{k+1} - y^k), \bar{y} - y^{k+1} \rangle + \langle (\widehat{\Sigma}_g + \mathcal{T})(z^{k+1} - z^k), \bar{z} - z^{k+1} \rangle \\ & \geq -\tfrac{1}{4}\|y^{k+1} - y^k\|^2_{\widehat{\Sigma}_f} - \tfrac{1}{4}\|z^{k+1} - z^k\|^2_{\widehat{\Sigma}_g}, \end{aligned}$$

where

$$\begin{aligned} \Delta_k := & \langle \bar{x} - (x^k + \sigma r^{k+1}) + \sigma\mathcal{B}^*(z^{k+1} - z^k), \mathcal{A}^*(y^{k+1} - \bar{y}) \rangle \\ & + \langle \bar{x} - (x^k + \sigma r^{k+1}), \mathcal{B}^*(z^{k+1} - \bar{z}) \rangle. \end{aligned}$$

Directly from [31, Equation (47)], it holds that
(A.7)

$$\begin{aligned} \Delta_k = \; & (2\tau\sigma)^{-1}\left(\|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2\right) + (2\sigma)^{-1}(\tau - 1)\|r^{k+1}\|^2 \\ & - \tfrac{\sigma}{2}\|\mathcal{A}^* y^{k+1} + \mathcal{B}^* z^k - c\|^2 + \tfrac{\sigma}{2}\left(\|\mathcal{B}^* z^k - \mathcal{B}^*\bar{z}\|^2 - \|\mathcal{B}^* z^{k+1} - \mathcal{B}^*\bar{z}\|^2\right). \end{aligned}$$

Since for any self-adjoint linear operator $\mathcal{G}$, it holds that $\langle u, \mathcal{G}v \rangle = \frac{1}{2}(\|u + v\|^2_{\mathcal{G}} - \|u\|^2_{\mathcal{G}} - \|v\|^2_{\mathcal{G}})$, we have

$$\begin{aligned} & \langle (\widehat{\Sigma}_f + \mathcal{S})(y^{k+1} - y^k), \bar{y} - y^{k+1} \rangle + \langle (\widehat{\Sigma}_g + \mathcal{T})(z^{k+1} - z^k), \bar{z} - z^{k+1} \rangle \\ & = \tfrac{1}{2}\big(\|y^k - \bar{y}\|^2_{\widehat{\Sigma}_f + \mathcal{S}} - \|y^{k+1} - \bar{y}\|^2_{\widehat{\Sigma}_f + \mathcal{S}}\big) - \tfrac{1}{2}\|y^{k+1} - y^k\|^2_{\widehat{\Sigma}_f + \mathcal{S}} \\ & \quad + \tfrac{1}{2}\big(\|z^k - \bar{z}\|^2_{\widehat{\Sigma}_g + \mathcal{T}} - \|z^{k+1} - \bar{z}\|^2_{\widehat{\Sigma}_g + \mathcal{T}}\big) - \tfrac{1}{2}\|z^{k+1} - z^k\|^2_{\widehat{\Sigma}_g + \mathcal{T}}. \end{aligned}$$

This, together with (A.7) and (A.6) implies that the conclusion holds. The proof is completed. □

A.4. **Proof of Lemma 3.4.**

Proof. From [31, Lemma 7], it holds that

$$\begin{aligned}&\|y^{k+1}-y^k\|^2_{\frac{1}{2}\widehat{\Sigma}_f+\mathcal{S}}+\|z^{k+1}-z^k\|^2_{\frac{1}{2}\widehat{\Sigma}_g+\mathcal{T}}+\sigma\|\mathcal{A}^*y^{k+1}+\mathcal{B}^*z^k-c\|^2+(1-\tau)\sigma\|r^{k+1}\|^2\\ &\geq \|y^{k+1}-y^k\|^2_{\frac{1}{2}\widehat{\Sigma}_f+\mathcal{S}}+\|z^{k+1}-z^k\|^2_{\frac{1}{2}\widehat{\Sigma}_g+\mathcal{T}}+\|z^{k+1}-z^k\|^2_{\widehat{\Sigma}_g+\mathcal{T}}-\|z^k-z^{k-1}\|^2_{\widehat{\Sigma}_g+\mathcal{T}}\\ &\quad+\min(\tau,1+\tau-\tau^2)\sigma(\tau^{-1}\|r^{k+1}\|^2+\|\mathcal{B}^*(z^{k+1}-z^k)\|^2)\\ &\qquad\qquad+(1-\min(\tau,\tau^{-1}))\sigma(\|r^{k+1}\|^2-\|r^k\|).\end{aligned}$$

This, together with the definition of ϕ_k and Lemma 3.3 implies the conclusion. □

A.5. **Proof of Lemma 3.5.**

Proof. First we show that $\mathcal{H}\succ 0$, $\mathcal{M}\succ 0$ when condition (3.8) holds, i.e.,

$$\frac{1}{2}\widehat{\Sigma}_f+\mathcal{S}+\sigma\mathcal{A}\mathcal{A}^*\succ 0 \quad \text{and} \quad \frac{1}{2}\widehat{\Sigma}_g+\mathcal{T}+\sigma\mathcal{B}\mathcal{B}^*\succ 0 \Rightarrow \mathcal{H}\succ 0,\ \ \mathcal{M}\succ 0.$$

Suppose that $\frac{1}{2}\widehat{\Sigma}_f+\mathcal{S}+\sigma\mathcal{A}\mathcal{A}^*\succ 0$ and $\frac{1}{2}\widehat{\Sigma}_g+\mathcal{T}+\sigma\mathcal{B}\mathcal{B}^*\succ 0$, but there exists a vector $0\neq d:=(d_y,d_z,d_x)\in\mathcal{Y}\times\mathcal{Z}\times\mathcal{X}$ such that $\langle d,\mathcal{H}d\rangle=0$, by using the definition of $\mathcal{H}$, we have

$$\langle d_y,\,\mathcal{H}_f d_y\rangle=0,\ \langle d_z,\,\mathcal{H}_g d_z\rangle=0,\ \ d_x=0,\ \ \mathcal{E}^*(d_y,d_z,d_x)=0.$$

Since $t_\tau>0$ and $\frac{1}{2}\widehat{\Sigma}_g+\mathcal{T}+\sigma\mathcal{B}\mathcal{B}^*\succ 0$, we know $d_z=0$. Consequently, $\mathcal{A}^*d_y=0$. This together with the assumption that $\frac{1}{2}\widehat{\Sigma}_f+\mathcal{S}+\sigma\mathcal{A}\mathcal{A}^*\succ 0$ implies $d_y=0$. This contradiction shows that $\mathcal{H}\succ 0$. We can get $\mathcal{M}\succ 0$ by using the same techniques. For brevity, we omit the details.

Next, we show that $\mathcal{H}\succ 0\Rightarrow\frac{1}{2}\widehat{\Sigma}_f+\mathcal{S}+\sigma\mathcal{A}\mathcal{A}^*\succ 0$ and $\frac{1}{2}\widehat{\Sigma}_g+\mathcal{T}+\sigma\mathcal{B}\mathcal{B}^*\succ 0$. Since $t_\tau\in(0,(1+\sqrt{5})/2)$ and for any $d=(d_y,0,0)\in\mathcal{Y}\times\mathcal{Z}\times\mathcal{X}$, we have $0<t_\tau\leq\frac{1}{2}$ and $\langle d,\,\mathcal{H}d\rangle=\langle d_y,(\mathcal{H}_f+\frac{1}{4}t_\tau\sigma\mathcal{A}\mathcal{A}^*)d_y\rangle$. Therefore, it holds that

$$\frac{1}{2}\widehat{\Sigma}_f+\mathcal{S}+\sigma\mathcal{A}\mathcal{A}^*=(\mathcal{H}_f+\frac{1}{4}t_\tau\sigma\mathcal{A}\mathcal{A}^*)+(1-\frac{1}{4}t_\tau)\sigma\mathcal{A}\mathcal{A}^*\succ 0.$$

Similarly, for any $d=(0,d_z,0)\in\mathcal{Y}\times\mathcal{Z}\times\mathcal{X}$, $\langle d,\,\mathcal{H}d\rangle=\langle d_z,(\mathcal{H}_g+\frac{1}{4}t_\tau\sigma\mathcal{B}\mathcal{B}^*)d_z\rangle$, then we can obtain that $\frac{1}{2}\widehat{\Sigma}_g+\mathcal{T}+\sigma\mathcal{B}\mathcal{B}^*\succ 0$. The proof is completed. □

References

[1] S. Banert, R. I. Bot, and E. R. Csetnek, *Fixing and extending some recent results on the admm algorithm*, arXiv preprint arXiv:1612.05057, (2016).

[2] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci. **2** (2009), no. 1, 183–202, DOI 10.1137/080716542. MR2486527

[3] C.-C. Chang and C.-J. Lin, *LIBSVM: A library for support vector machines*, ACM Transactions on Intelligent Systems and Technology, 2(3):Article 27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[4] L. Chen, D. Sun, and K.-C. Toh, *An efficient inexact symmetric Gauss-Seidel based majorized ADMM for high-dimensional convex composite conic programming*, Math. Program. **161** (2017), no. 1-2, Ser. A, 237–270, DOI 10.1007/s10107-016-1007-5. MR3592778

[5] H.-Y. Chuang, E. Lee, Y.-T. Liu, D. Lee, and T. Ideker, *Network-based classification of breast cancer metastasis*, Molecular Systems Biology, 3(1):140, 2007.

[6] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Canadian Mathematical Society Series of Monographs and Advanced Texts, John Wiley & Sons, Inc., New York, 1983. A Wiley-Interscience Publication. MR709590

[7] L. Condat. A direct algorithm for 1-D total variation denoising. *IEEE Signal Processing Letters*, 20(11):1054–1057, 2013.

[8] L. Condat, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, J. Optim. Theory Appl. **158** (2013), no. 2, 460–479, DOI 10.1007/s10957-012-0245-9. MR3084386

[9] C. Ding, D. Sun, and L. Zhang, *Characterization of the robust isolated calmness for a class of conic programming problems*, SIAM J. Optim. **27** (2017), no. 1, 67–90, DOI 10.1137/16M1058753. MR3594330

[10] Y. Dong, *An extension of Luque's growth condition*, Appl. Math. Lett. **22** (2009), no. 9, 1390–1393, DOI 10.1016/j.aml.2007.07.037. MR2536820

[11] A. L. Dontchev and R. T. Rockafellar, *Regularity and conditioning of solution mappings in variational analysis*, Set-Valued Anal. **12** (2004), no. 1-2, 79–109, DOI 10.1023/B:SVAN.0000023394.19482.30. MR2069353

[12] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings*, Springer Monographs in Mathematics, Springer, Dordrecht, 2009. A view from variational analysis. MR2515104

[13] D. Drusvyatskiy and A. S. Lewis, *Error bounds, quadratic growth, and linear convergence of proximal methods*, Math. Oper. Res. **43** (2018), no. 3, 919–948, DOI 10.1287/moor.2017.0889. MR3846078

[14] J. Eckstein, *Splitting methods for monotone operators with applications to parallel optimization*, Ph.D. thesis, Massachusetts Institute of Technology, 1989.

[15] J. Eckstein, *Some saddle-function splitting methods for convex programming*, Optimization Methods and Software, 4(1):75–83, 1994.

[16] J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Programming **55** (1992), no. 3, Ser. A, 293–318, DOI 10.1007/BF01581204. MR1168183

[17] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, *Hankel matrix rank minimization with applications to system identification and realization*, SIAM J. Matrix Anal. Appl. **34** (2013), no. 3, 946–977, DOI 10.1137/110853996. MR3073649

[18] M. Fortin and R. Glowinski, *Augmented Lagrangian methods*, Studies in Mathematics and its Applications, vol. 15, North-Holland Publishing Co., Amsterdam, 1983. Applications to the numerical solution of boundary value problems; Translated from the French by B. Hunt and D. C. Spicer. MR724072

[19] J. Friedman, T. Hastie, and R. Tibshirani, *Regularization paths for generalized linear models via coordinate descent*, Journal of Statistical Software, 33(1):1, 2010.

[20] D. Gabay and M. Bertrand, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Computers & Mathematics with Applications, 2(1):17–40, 1976.

[21] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Scientific Computation, Springer-Verlag, Berlin, 2008. Reprint of the 1984 original. MR2423313

[22] R. Glowinski, *On alternating direction methods of multipliers: a historical perspective*, Modeling, simulation and optimization for science and technology, Comput. Methods Appl. Sci., vol. 34, Springer, Dordrecht, 2014, pp. 59–82, DOI 10.1007/978-94-017-9054-3_4. MR3330832

[23] R. Glowinski and A. Marrocco, *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires* (French, with English summary), Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge Anal. Numér. **9** (1975), no. no. , no. R-2, 41–76. MR388811

[24] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 2013. MR3024913

[25] D. Han, D. Sun, and L. Zhang, *Linear rate convergence of the alternating direction method of multipliers for convex composite programming*, Math. Oper. Res. **43** (2018), no. 2, 622–637, DOI 10.1287/moor.2017.0875. MR3801109

[26] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I: Fundamentals*, volume 305. Springer Science & Business Media, 2013.

[27] G. M. James, C. Paulson, and P. Rusmevichientong, *Penalized and constrained optimization: An application to high-dimensional website advertising*, Journal of the American Statistical Association, 2019. Available at `https://doi.org/10.1080/01621459.2019.1609970`.

[28] R. Jenatton, J.-Y. Audibert, and F. Bach, *Structured variable selection with sparsity-inducing norms*, J. Mach. Learn. Res. **12** (2011), 2777–2824. MR2854347

[29] X. Y. Lam, J. S. Marron, D. Sun, and K.-C. Toh, *Fast algorithms for large-scale generalized distance weighted discrimination*, J. Comput. Graph. Statist. **27** (2018), no. 2, 368–379, DOI 10.1080/10618600.2017.1366915. MR3816272

[30] C. Lemaréchal and C. Sagastizábal, *Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries*, SIAM J. Optim. **7** (1997), no. 2, 367–385, DOI 10.1137/S1052623494267127. MR1443624

[31] M. Li, D. Sun, and K.-C. Toh, *A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization*, SIAM J. Optim. **26** (2016), no. 2, 922–950, DOI 10.1137/140999025. MR3484404

[32] X. Li, D. Sun, and K.-C. Toh, *A Schur complement based semi-proximal ADMM for convex quadratic conic programming and extensions*, Math. Program. **155** (2016), no. 1-2, Ser. A, 333–373, DOI 10.1007/s10107-014-0850-5. MR3439805

[33] X. Li, D. Sun, and K.-C. Toh, *A block symmetric Gauss-Seidel decomposition theorem for convex composite quadratic programming and its applications*, Math. Program. **175** (2019), no. 1-2, Ser. A, 395–418, DOI 10.1007/s10107-018-1247-7. MR3942895

[34] X. Li, D. Sun, and K.-C. Toh, *A highly efficient semismooth Newton augmented Lagrangian method for solving lasso problems*, SIAM J. Optim. **28** (2018), no. 1, 433–458, DOI 10.1137/16M1097572. MR3763098

[35] M. Lichman, *UCI machine learning repository*, 2013. Available at `http://archive.ics.uci.edu/ml`.

[36] J. Liu, J. Chen, and J. Ye, *Large-scale sparse logistic regression*, Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 547–556. ACM, 2009.

[37] J. Liu, L. Yuan, and J. Ye, *An efficient algorithm for a class of fused lasso problems*, Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 323–332. ACM, 2010.

[38] J.-J. Moreau, *Proximité et dualité dans un espace hilbertien* (French), Bull. Soc. Math. France **93** (1965), 273–299. MR201952

[39] Y. Nesterov, *Introductory Lectures on Convex Optimization*, A basic course. Applied Optimization, vol. 87, Kluwer Academic Publishers, Boston, MA, 2004. MR2142598

[40] Yu. E. Nesterov, *A method for solving the convex programming problem with convergence rate* $O(1/k^2)$ (Russian), Dokl. Akad. Nauk SSSR **269** (1983), no. 3, 543–547. MR701288

[41] S. M. Robinson, *Some continuity properties of polyhedral multifunctions*, Math. Programming Stud. **14** (1981), 206–214, DOI 10.1007/bfb0120929. MR600130

[42] R. T. Rockafellar, *Convex Analysis*, Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, N.J., 1970. MR0274683

[43] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, volume 317. Springer Science & Business Media, 2009.

[44] J. Sun, *On monotropic piecewise quadratic programming* (*Network, algorithm, convex programming, decomposition method*), ProQuest LLC, Ann Arbor, MI, 1986. Thesis (Ph.D.)–University of Washington. MR2635493

[45] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B **58** (1996), no. 1, 267–288. MR1379242

[46] R. Tibshirani, *The lasso method for variable selection in the Cox model*, Statistics in Medicine, 16(4):385–395, 1997.

[47] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, *Sparsity and smoothness via the fused lasso*, J. R. Stat. Soc. Ser. B Stat. Methodol. **67** (2005), no. 1, 91–108, DOI 10.1111/j.1467-9868.2005.00490.x. MR2136641

[48] J. J. Ye, *Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints*, SIAM J. Optim. **10** (2000), no. 4, 943–962, DOI 10.1137/S105262349834847X. MR1777074

[49] X. Yuan, S. Zeng, and J. Zhang. Discerning the linear convergence of admm for structured convex optimization through the lens of variational analysis. 2018. Available at http://www.optimization-online.org/DBHTML/2018/10/6882.html.

School of Computer Science and Technology, Dongguan University of Technology, Dongguan 523808, People's Republic of China
Email address: ningzhang_2008@yeah.net

School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, People's Republic of China
Email address: wujia@dlut.edu.cn

School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, People's Republic of China
Email address: lwzhang@dlut.edu.cn