Acta Mathematicae Applicatae Sinica, English Series
Vol. 39, No. 2 (2023) 354–380
https://doi.org/10.1007/s10255-023-1048-5
http://www.ApplMath.com.cn & www.SpringerLink.com

Acta Mathematicae Applicatae Sinica, English Series

A Bregman-style Partially Symmetric Alternating Direction Method of Multipliers for Nonconvex Multi-block Optimization

Peng-jie LIU[1], **Jin-bao JIAN**[2,†], **Guo-dong MA**[2]

[1]School of Mathematics, China University of Mining and Technology, Xuzhou 221116, China
[2]College of Mathematics and Physics, Center for Applied Mathematics of Guangxi, Guangxi Minzu University, Nanning 530006, China (†E-mail: jianjb@gxu.edu.cn)

Abstract The alternating direction method of multipliers (ADMM) is one of the most successful and powerful methods for separable minimization optimization. Based on the idea of symmetric ADMM in two-block optimization, we add an updating formula for the Lagrange multiplier without restricting its position for multi-block one. Then, combining with the Bregman distance, in this work, a Bregman-style partially symmetric ADMM is presented for nonconvex multi-block optimization with linear constraints, and the Lagrange multiplier is updated twice with different relaxation factors in the iteration scheme. Under the suitable conditions, the global convergence, strong convergence and convergence rate of the presented method are analyzed and obtained. Finally, some preliminary numerical results are reported to support the correctness of the theoretical assertions, and these show that the presented method is numerically effective.

Keywords nonconvex optimization; multi-block optimization; alternating direction method with multipliers; Kurdyka-Łojasiewicz property; convergence rate
2000 MR Subject Classification 90C26; 90C30

1 Introduction

In this paper, we consider the multi-block nonconvex optimization problem in the following form:

$$\begin{aligned}&\min \sum_{i=1}^{m} f_i(x_i) + g(y),\\ &\text{s.t.} \ \sum_{i=1}^{m} A_i x_i + By = b,\end{aligned} \tag{1.1}$$

where the function $f_i : \mathbb{R}^{n_i} \to \mathbb{R}\cup\{+\infty\}$ is proper and lower semi-continuous but not necessarily smooth, $g : \mathbb{R}^p \to \mathbb{R}$ is a continuously differentiable function, $A_i \in \mathbb{R}^{t\times n_i}$, $B \in \mathbb{R}^{t\times p}$ and $b \in \mathbb{R}^t$. All functions can be nonconvex. Let $\mathbf{x}_{[h:l]}$ denote $(x_h, x_{h+1}, \cdots, x_l)^\top$ for given positive integers h and l ($l \geq h$). Then $\mathbf{x}_{[1:m]} = (x_1, \cdots, x_m)^\top \in \mathbb{R}^n$ with $n = \sum_{i=1}^{m} n_i$. Denote $\mathbf{A} = (A_1, \cdots, A_m) \in \mathbb{R}^{t\times n}$. Then $\mathbf{A}\mathbf{x}_{[1:m]} = \sum_{i=1}^{m} A_i x_i \in \mathbb{R}^t$. With these notations, the problem (1.1)

Manuscript received June 07, 2021. Accepted on January 27, 2022.
This paper is supported by the National Natural Science Foundation of China (No. 12171106) and the Natural Science Foundation of Guangxi Province (No. 2020GXNSFDA238017).
†Corresponding author.

can be written as

$$\begin{aligned} &\min \sum_{i=1}^{m} f_i(x_i) + g(y), \\ &\text{s.t. } \mathbf{A}\mathbf{x}_{[1:m]} + By = b. \end{aligned} \tag{1.2}$$

Let $\beta(>0)$ be a penalty parameter and

$$\mathcal{L}_\beta(\mathbf{x}_{[1:m]}, y, \lambda) = \sum_{i=1}^{m} f_i(x_i) + g(y) - \langle \lambda, \mathbf{A}\mathbf{x}_{[1:m]} + By - b\rangle + \frac{\beta}{2}\|\mathbf{A}\mathbf{x}_{[1:m]} + By - b\|^2 \tag{1.3}$$

is the augmented Lagrangian function (ALF) of (1.2) with the Lagrangian multiplier (dual variable) $\lambda \in \mathbb{R}^t$, and the ALF has the following property:

$$\mathcal{L}_\beta(\mathbf{x}_{[1:m]}, y, \lambda - \theta(\mathbf{A}\mathbf{x}_{[1:m]} + By - b)) = \mathcal{L}_\beta(\mathbf{x}_{[1:m]}, y, \lambda) + \theta\|\mathbf{A}\mathbf{x}_{[1:m]} + By - b\|^2, \tag{1.4}$$

where $\theta \in \mathbb{R}$ and $(\mathbf{x}_{[1:m]}, y, \lambda)^\top \in \mathbb{R}^{n+p+t}$. Even though the problem (1.2) has a simple form, its efficient solution is often very challenging in practice due to the nonconvexness and nonsmoothness of objective functions and the curse of high dimensionality arising from many real-world applications, such as signal and image processing[29, 30], machine learning[22, 33], etc. Thus, the development of efficient structure-utilizing algorithms for (1.2), which can overcome the difficulties caused by the aforementioned characteristics, becomes increasingly important recently. Given λ^k, the augmented Lagrangian method (ALM)[13, 20] for solving (1) updates the primal and the dual variables via

$$(\mathbf{x}_{[1:m]}^{k+1}, y^{k+1}) \in \arg\min_{\mathbf{x},\ y} \{\mathcal{L}_\beta(\mathbf{x}_{[1:m]}, y, \lambda^k)\}, \tag{1.5}$$

$$\lambda^{k+1} = \lambda^k - \beta(\mathbf{A}\mathbf{x}_{[1:m]}^{k+1} + By^{k+1} - b). \tag{1.6}$$

In general, the ALM specified in (1.5) is sufficiently efficient for solving (1.2) with one-block. However, for the cases of $m \geq 1$, the ALM could lose its efficiency due to the joint minimization for all x_i's or/and y's in (1.5), where the variables are mixed-all-together in the quadratic term of $\mathcal{L}_\beta$. This difficulty does not disappear even if all f_i's and y's are sufficiently simple. Therefore, an important motivating question is that whether one can improve the performance of the ALM by fully exploiting the valuable structures hidden in f_i's and g's. In fact, a series of ALM-based splitting methods have received considerable attention in recent years, and this paper contributes further in this line.

Restricted to the special case of $m = 1$, the alternating direction method of multipliers (ADMM)[7, 15], which decomposes the subproblem (1.5) into two easier ones by minimizing one variable and fixing the other, followed by an immediate update of the dual variable after each sweep of alternating minimization, is perhaps the most popular and influential approach for solving (1.2) with $m = 1$. Due to the remarkable performance of ADMM in various disciplines of scientific computing, it is natural to extend ADMM to the multi-block model (1.2) with $m \geq 2$. The directly extended ADMM appears as

$$\begin{aligned} x_1^{k+1} &\in \arg\min_{x_1} \{\mathcal{L}_\beta(x_1, \mathbf{x}_{[2:m]}^k, y^k, \lambda^k)\}, \\ x_2^{k+1} &\in \arg\min_{x_2} \{\mathcal{L}_\beta(x_1^{k+1}, x_2, \mathbf{x}_{[3:m]}^k, y^k, \lambda^k)\}, \\ &\vdots \\ x_m^{k+1} &\in \arg\min_{x_m} \{\mathcal{L}_\beta(\mathbf{x}_{[1:m-1]}^{k+1}, x_m, y^k, \lambda^k)\}, \end{aligned}$$

$$y^{k+1} \in \arg\min_{y}\{\mathcal{L}_\beta(\mathbf{x}^{k+1}_{[1:m]}, y, \lambda^k)\},$$
$$\lambda^{k+1} = \lambda^k - \beta(\mathbf{A}\mathbf{x}^{k+1}_{[1:m]} + By^{k+1} - b).$$

However, the iteration above may not converge[5, 6]. Han and Yuan[10] first proved that when all the separable objective functions are strongly convex and when the penalty parameter is less than a certain threshold, the multi-block version directly generalized by the classical two-block ADMM has global convergence. On the other hand, He, Tao and Yuan obtained multi-block ADMM by modifying and extending the ADMM, such as the ADMM with Gaussian back substitution[11, 12].

At present, most of the researches about splitting algorithm of multi-block optimization focus on convex problem. There are few studies on ADMM or other splitting algorithms for nonconvex multi-block optimization. Guo et al.[9] directly generalized the classical two-block nonconvex ADMM to the three-block case (i.e., $m = 2$ and $B = I$), and analyzed the global convergence of the algorithm when A_i is column full rank and the penalty parameter is limited to a certain range. In addition, if the potential function satisfies the Kurdyka-Łojasiewicz (KL) property, the strong convergence of the algorithm is also obtained. Similarly, for the same three-block case with $m = 2$ and $B = I$, Zhang et al.[32] proposed a linearized ADMM in combination with the proximal linearization technique and A_i no longer need column full rank in the convergence analysis. Combining with the Bregman distance, Wang et al.[24] obtained the convergence of Bregman ADMM (BADMM) in the nonconvex multi-block case for solving (1.2) with $b = 0$. Meanwhile, it should be noted that the convergence results of these algorithms in the nonconvex case often depend on the KL property. For more discussion on multi-block ADMM in nonconvex case, see Refs. [17, 28, 31].

On the other hand, for the two-block optimization problem, a Lagrange multiplier update intermediate step is added to become the symmetric ADMM, also called Peaceman-Rachford splitting method (PRSM)[19]. In the convex case, Gabay[8] pointed out that PRSM converges faster than ADMM in the case of convergence, and PRSM is less "robust" in the sense that it converges under more restrictive conditions than ADMM. With regard to this, for the ADM-M iterative framework of multi-block optimization problems, we also want to add a Lagrange multiplier update intermediate term without restricting its position. In addition, similar to the Bregman-style ADMM iterative pattern[14, 16, 23, 24, 26], the Bregman distance is added to the subproblem. Based on the above analysis, the proposed iterative format of the Bregman-style partially symmetric ADMM (in short, BPSADMM), for problem (1.2) with $B = I$, is as follows.

Step 0. According to the sparse and dense conditions of coefficient matrix A_i, then, the j matrices with relatively complex and dense structures are denoted as $A_1, A_2, \cdots, A_j$, and other matrices with simple structure and good sparsity are denoted as $A_{j+1}, \cdots, A_m$. Choose an initial point $(\mathbf{x}^0_{[1:m]}, y^0, \lambda^0)$, the relaxation factors satisfy $r + s > 0$, the penalty parameter $\beta > 0$, and set $k := 0$.

Step 1. Solve the $\mathbf{x}_{[1:j]}-$subproblems, and then update the Lagrange multiplier:

$$x_1^{k+1} \in \arg\min_{x_1}\{\mathcal{L}_\beta(x_1, \mathbf{x}^k_{[2:m]}, y^k, \lambda^k) + \triangle_{\phi_1}(x_1, x_1^k)\}, \tag{1.7}$$
$$x_2^{k+1} \in \arg\min_{x_2}\{\mathcal{L}_\beta(x_1^{k+1}, x_2, \mathbf{x}^k_{[3:m]}, y^k, \lambda^k) + \triangle_{\phi_2}(x_2, x_2^k)\}, \tag{1.8}$$
$$\vdots$$
$$x_j^{k+1} \in \arg\min_{x_j}\{\mathcal{L}_\beta(\mathbf{x}^{k+1}_{[1:j-1]}, x_j, \mathbf{x}^k_{[j+1:m]}, y^k, \lambda^k) + \triangle_{\phi_j}(x_j, x_j^k)\}, \tag{1.9}$$

$$\lambda^{k+\frac{1}{2}} = \lambda^k - r\beta\Big(\sum_{i=1}^{j} A_i x_i^{k+1} + \sum_{i=j+1}^{m} A_i x_i^k + y^k - b\Big). \tag{1.10}$$

Step 2. Solve the $\mathbf{x}_{[j+1:m]}-$subproblems:

$$x_{j+1}^{k+1} \in \arg\min_{x_{j+1}} \{\mathcal{L}_\beta(\mathbf{x}_{[1:j]}^{k+1}, x_{j+1}, \mathbf{x}_{[j+2:m]}^k, y^k, \lambda^{k+\frac{1}{2}}) + \triangle_{\phi_{j+1}}(x_{j+1}, x_{j+1}^k)\}, \tag{1.11}$$

$$\vdots$$

$$x_m^{k+1} \in \arg\min_{x_m} \{\mathcal{L}_\beta(\mathbf{x}_{[1:m-1]}^{k+1}, x_m, y^k, \lambda^{k+\frac{1}{2}}) + \triangle_{\phi_m}(x_m, x_m^k)\}. \tag{1.12}$$

Step 3. Solve the $y-$subproblem, and then update the Lagrange multiplier:

$$y^{k+1} \in \arg\min_{y} \{\mathcal{L}_\beta(\mathbf{x}_{[1:m]}^{k+1}, y, \lambda^{k+\frac{1}{2}}) + \triangle_{\hat{\phi}}(y, y^k)\}, \tag{1.13}$$

$$\lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta\Big(\sum_{i=1}^{m} A_i x_i^{k+1} + y^{k+1} - b\Big). \tag{1.14}$$

Step 4. If a termination criterion is met, stop; otherwise, replace k by $k+1$ and turn to Step 1.

In the iterative algorithm above, $\triangle_{\phi_i}$ $(i = 1, \cdots, m)$ and $\triangle_{\hat{\phi}}$ are Bregman distance with respect to convex differentiable functions ϕ_i $(i = 1, \cdots, m)$ and $\hat{\phi}$, respectively. The BPSADMM has the following three characteristics:

(i) The introduction of (r, s) makes the BPSADMM widely representative, including many variants of ADMM-type and PRSM-type splitting methods.

(ii) Choosing an appropriate Bregman distance can simplify the subproblem of BPSADMM.

• For the x_i-subproblem, a suitable Bregman distance is added to linearize the differentiable terms (e.g., the quadratic penalty term). Further, the x_i-subproblem can be simplified or its explicit solution can be obtained. For example, if the Bregman distance generating function is chosen to be

$$\phi_1 = \frac{\mu_1}{2}\|x_1\|^2 - \frac{\beta}{2}\Big\|A_1 x_1 + \sum_{q=2}^{m} A_q x_q^k + y^k - b\Big\|^2,$$

the x_1-update can be simplified as $\min\limits_{x_1}\{f_1(x_1) + \frac{\mu_1}{2}\|x_1 - b_1^k\|\}$ with a certain known b_1^k.

• For the y-subproblem, due to $g(y)$ being a continuously differentiable function, an appropriate Bregman distance choice is of particular interest since it simplifies the y-update to a convex quadratic programming problem. Indeed, if $\hat{\phi}$ is chosen to be $\frac{L_g}{2}\|y\|^2 - g(y)$, where L_g is at least as large as the Lipschitz continuity modulus of $\nabla g(y)$, the $y-$subproblem becomes

$$\min_y \Big\{\frac{L_g}{2}\|y - y^k\|^2 + \langle \nabla g(y^k) - \lambda^{k+\frac{1}{2}},\ y\rangle + \frac{\beta}{2}\|\mathbf{A}\mathbf{x}_{[1:m]}^{k+1} + y - b\|^2\Big\}.$$

(iii) It can be seen from the multiplier λ correction system that the higher the density of coefficient matrix A_i, the greater its influence on multiplier λ correction, theoretically, and vice versa. For this reason, we try to determine the correction position of the multiplier update intermediate step $\lambda^{k+\frac{1}{2}}$ according to the sparsity of the coefficient matrix A_i in the equality constraint of problem (1.2), so as to further improve the computational performance of the proposed algorithm.

1.1 Main Contribution

In this paper, we propose a BPSADMM for nonconvex multi-block optimization by combining the Bregman distance. It should be noted that we do not limit the position of multiplier update intermediate step $\lambda^{k+\frac{1}{2}}$ in the iterative algorithm. Moreover, the introduction of different relaxation factors r and s gives the BPSADMM a wide range of characteristics. For the proposed method, $\mathcal{L}_\beta(\cdot)$ might not be monotomically decreasing. Then, a modified version $\hat{\mathcal{L}}_\beta(\cdot)$, defined in (2.11) below, is monotonically decreasing with a right parameter. If $\{\hat{\mathcal{L}}_\beta(\cdot)\}$ satisfies KL property, we prove that the whole iteration sequence generated by the BPSADMM converges to a critical point of the problem (1.2). In addition, the convergence rates of the merit function and iteration sequence are obtained under Łojasiewicz property.

1.2 Notation and Elementary Results

In this work, let $\mathbb{R}^n$ denote the n-dimensional vector space, and we use $\mathbb{R}^{m\times n}$ to denote the set of all $m\times n$ matrices. For a vector $a\in\mathbb{R}^n$, the associated norm is denoted by $\|a\|$. For two vectors a and b of the same size, we denote their inner product by $\langle a,b\rangle = a^\top b$. For a matrix $A\in\mathbb{R}^{m\times n}$, let a_{ij} denote its (i,j)-th entry, and the largest (resp., smallest) eigenvalue of the symmetric matrix A is denoted by $\lambda_{\max}(A)$ (resp., $\lambda_{\min}(A)$). If A is a symmetric and positive semidefinite matrix, the seminorm is defined by $\|x\|_A = \langle Ax, x\rangle$ for all $x\in\mathbb{R}^n$. Moreover, we also use $d(x,\mathbb{S})$ to denote the distance from point $x\in\mathbb{R}^n$ to subset $\mathbb{S}\subseteq\mathbb{R}^n$, i.e., $d(x,\mathbb{S}) = \inf_{y\in\mathbb{S}}\|y-x\|$. When $\mathbb{S}=\emptyset$, we set $d(x,\mathbb{S}) = +\infty$ for all $x\in\mathbb{R}^n$.

For function $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$, the effective domain and epigraph of f are defined as follows:

$$\mathrm{dom}f = \{x\in\mathbb{R}^n : f(x) < +\infty\}, \qquad \mathrm{epi}f = \{(x,\alpha)\in\mathbb{R}^n\times\mathbb{R} : f(x)\le\alpha\},$$

and f is proper iff $\mathrm{dom}f\neq\emptyset$ and $f > -\infty$.

Definition 1.1[21]**.** *The function $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ is lower semicontinuous at x_0 if $\liminf\limits_{x\to x_0} f(x)\ge f(x_0)$. If f is lower semicontinuous at any point $x\in$ dom f, then it is called the lower semicontinuous function.*

Definition 1.2[21]**.** *Let $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ be a proper lower semicontinuous function.*

(i) The Fréchet subdifferential of f at $x\in$ dom f, denoted by $\hat\partial f(x)$, is the set of vectors $x^*\in\mathbb{R}^n$ that satisfies

$$\lim_{y\neq x}\inf_{y\to x}\frac{f(y)-f(x)-\langle x^*, y-x\rangle}{\|y-x\|}\ge 0,$$

and when $x\notin \mathrm{dom}f$, we set $\hat\partial f(x)=\emptyset$.

(ii) The (limiting-)subdifferential of f at $x\in$ dom f, written as $\partial f(x)$, is

$$\partial f(x) = \{x^*\in\mathbb{R}^n : \exists\ x_k \underset{f}{\to} x,\ x_k^*\in\hat\partial f(x_k), \text{with } x_k^*\to x^*\}.$$

Remark. From Definition 1.2, if $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ is a proper lower semicontinuous function, the following conclusions hold (see Ref. [21]).

(i) $\hat\partial f(x)\subseteq\partial f(x)$ for each $x\in\mathbb{R}^n$, where the first set is closed convex while the second one is only closed.

(ii) Let $x_k^*\in\partial f(x_k)$ and $\lim\limits_{k\to\infty}(x_k,x_k^*) = (x,x^*)$, then $x^*\in\partial f(x)$, i.e., $\partial f(x)$ is closed.

(iii) A necessary (but not sufficient) condition for $x \in \mathbb{R}^n$ to be a minimizer of f is $0 \in \partial f(x)$. And a point that satisfies $0 \in \partial f(x)$ is called a critical point.

(iv) If $g : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function, then $\partial(f+g)(x) = \partial f(x) + \nabla g(x)$ for any $x \in \mathrm{dom} f$.

For a convex differentiable function ϕ, the associated Bregman distance is defined as

$$\triangle_\phi(x,y) = \phi(x) - \phi(y) - \langle \nabla\phi(y), x-y\rangle.$$

If $\phi(x) = \|x\|^2$, $\triangle_\phi(x,y) = \|x-y\|^2$; if $\phi(x) = x^\top Ax$, where A is a positive semidefinite matrix, $\triangle_\phi(x,y) = (x-y)^\top A(x-y)$. Now we give some useful properties (see Lemma 1.1) about the Bregman distance.

Lemma 1.1[25]**.** *Let ϕ be a convex differentiable function and $\triangle_\phi(x,y)$ be the associated Bregman distance. Then,*

(i) $\triangle_\phi(x,y) \geq 0$, $\triangle_\phi(x,x) = 0$ for any $x,y \in \mathbb{R}^n$;

(ii) $\triangle_\phi(x,y)$ is convex at x, but not necessarily at y;

(iii) If ϕ is $\sigma-$strongly convex, $\triangle_\phi(x,y) \geq \frac{\sigma}{2}\|x-y\|^2$ for any $x,y \in \mathbb{R}^n$.

Lemma 1.2[18]**.** *Let $h : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function with gradient ∇h being Lipschitz continuous with constant $L > 0$. Then for any $x, y \in \mathbb{R}^n$, we have*

$$|h(y) - h(x) - \langle \nabla h(x), y-x\rangle| \leq \frac{L}{2}\|y-x\|^2.$$

Lemma 1.3. *$(\mathbf{x}^*_{[1:m]}, y^*, \lambda^*)$ is a critical point of the ALF $\mathcal{L}_\beta(\cdot)$ defined in (1.3), i.e., $0 \in \partial\mathcal{L}_\beta(\mathbf{x}^*_{[1:m]}, y^*, \lambda^*)$, if and only if the following relations hold*

$$A_i^\top\lambda^* \in \partial f_i(x_i^*), \quad i = 1,2,\cdots,m, \qquad \nabla g(y^*) = \lambda^*, \qquad \sum_{i=1}^m A_i x_i^* + y^* - b = 0. \tag{1.15}$$

The set of critical points of $\mathcal{L}_\beta(\cdot)$ is denoted by $\mathrm{crit}\mathcal{L}_\beta$.

Next, we recall the Łojasiewicz property and Kurdyka-Łojasiewicz (KL) property, which play an important role in our convergence and convergence rate analysis.

Definition 1.3[1] (Łojasiewicz property)**.** *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semi-continuous function. Let x^* be a critical point of f, that is, $0 \in \partial f(x^*)$. The function f has Łojasiewicz property at x^* if there exist $c_l > 0$, $\theta \in [0,1)$ and a neighborhood U of x^* such that*

$$|f(x) - f(x^*)|^\theta \leq c_l \cdot d(0, \partial f(x)), \qquad \forall\, x \in U.$$

For notational simplicity, we use Ξ_η to denote the set of all concave functions $\varphi : [0,\eta) \to [0,+\infty)$ satisfying: (i) $\varphi(0) = 0$; (ii) φ is continuously differentiable on $(0,\eta)$ and continuous at 0; (iii) $\varphi'(t) > 0$ for any $t \in (0,\eta)$. Then, the KL property can be described as follows.

Definition 1.4[2] (KL property)**.** *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. For $-\infty < \eta_1 < \eta_2 \leq +\infty$, set $[\eta_1 < f < \eta_2] = \{x \in \mathbb{R}^n : \eta_1 < f(x) < \eta_2\}$. We say that function f has the Kurdyka-Łojasiewicz (KL) property at $x^* \in \mathrm{dom}\partial f := \{x \in \mathbb{R}^n : \partial f(x) \neq \emptyset\}$ if there exist $\eta \in (0,+\infty]$, a neighborhood U of x^*, and a continuous concave function $\varphi \in \Xi_\eta$, such that, for all $x \in U \cap [f(x^*) < f < f(x^*) + \eta]$, the Kurdyka-Łojasiewicz inequality holds:*

$$\varphi'(f(x) - f(x^*))d(0, \partial f(x)) \geq 1.$$

If f satisfies the KL property at each point dom ∂f, then f is called a KL function. Moreover, φ is called the associated KL property function.

Lemma 1.4[3] (Uniformized KL property). *Let $\Omega \subset \mathbb{R}^n$ be a compact set and $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. Assume that f is constant f^* on Ω and satisfies the KL property at each point of Ω. Then, there exist $\epsilon > 0, \eta > 0$, and $\varphi \in \Xi_\eta$ such that for all $x \in \{x \in \mathbb{R} : d(x, \Omega) < \epsilon\} \cap [f^* < f < f^* + \eta]$, one has,*

$$\varphi'(f(x) - f^*)d(0, \partial f(x)) \geq 1.$$

1.3 Organization

The rest of the paper is organized as follows. In Sect. 2 and Sect. 3, we establish the convergence and convergence rate results of the BPSADMM for problem (1.2) under the suitable assumptions, respectively. In Sect. 4, some preliminary numerical results are reported to support the efficiency of the proposed algorithm. In Sect. 5, we draw some conclusions and perspectives.

2 Convergence Analysis

In this section, we discuss the convergence of BPSADMM, and the convergence analysis relies on the following basic assumptions for the problem (1.2) and the BPSADMM iteration:

Assumption H. (i) The function $f_i(x_i) : \mathbb{R}^{n_i} \to \mathbb{R} \cup \{+\infty\}$ $(i = 1, \cdots, m)$ is proper and lower semi-continuous, and $g(y) : \mathbb{R}^p \to \mathbb{R}$ is continuously differentiable such that ∇g is Lipschitz continuous with constant $L_g > 0$;

(ii) the relaxation factors r and s satisfy $r + s > 0$;

(iii) ϕ_i $(i = 1, \cdots, m)$ is a σ_i-strongly convex continuously differentiable function, where

$$\sigma_i > \frac{2(m+2)\beta(1-s)^2\lambda_{\max}(A_i^\top A_i)}{r+s}, \qquad i = 1, \cdots, j, \tag{2.1}$$

and

$$\sigma_i > \frac{2[(m+2)\beta(1-s)^2 + (m+1-j)\beta|rs|]\lambda_{\max}(A_i^\top A_i)}{r+s}, \qquad i = j+1, \cdots, m, \tag{2.2}$$

as well as $\nabla\phi_i$ $(i = 1, \cdots, m)$ is L_{ϕ_i}-Lipschitz continuous;

(iv) $\hat{\phi}$ is a $\hat{\sigma}-$strongly convex continuously differentiable function, and $\nabla\hat{\phi}$ is $L_{\hat{\phi}}$-Lipschitz continuous, where $\hat{\sigma}$ satisfies

$$\hat{\sigma} > \frac{2(m+2)[(L_g + \beta|1-s| + L_{\hat{\phi}})^2 - L_{\hat{\phi}}^2]}{(r+s)\beta} + \frac{2(m+1-j)\beta|rs|}{r+s}. \tag{2.3}$$

For convenience, the following notations are used uniformly throughout the paper:

$$w = (\mathbf{x}_{[1:m]}, y, \lambda), \qquad w^k = (\mathbf{x}^k_{[1:m]}, y^k, \lambda^k), \qquad w^* = (\mathbf{x}^*_{[1:m]}, y^*, \lambda^*);$$

$$\hat{w} = (\mathbf{x}_{[1:m]}, y, \lambda, \hat{y}), \qquad \hat{w}^k = (\mathbf{x}^k_{[1:m]}, y^k, \lambda^k, y^{k-1}), \qquad \hat{w}^* = (\mathbf{x}^*_{[1:m]}, y^*, \lambda^*, y^*);$$

$$\delta_i := \delta_i(r, s, \beta) = \frac{\sigma_i}{2} - \frac{(m+2)\beta(1-s)^2\lambda_{\max}(A_i^\top A_i)}{r+s}, \qquad i = 1, \cdots, j, \tag{2.4}$$

$$\delta_i := \delta_i(r, s, \beta) = \frac{\sigma_i}{2} - \frac{[(m+2)\beta(1-s)^2 + (m+1-j)\beta|rs|]\lambda_{\max}(A_i^\top A_i)}{r+s}, \quad i = j+1, \cdots, m, \tag{2.5}$$

$$\hat{\delta} := \hat{\delta}(r, s, \beta) = \frac{\hat{\sigma}}{2} - \frac{(m+2)[(L_g + \beta|1-s| + L_{\hat{\phi}})^2 - L_{\hat{\phi}}^2]}{(r+s)\beta} - \frac{(m+1-j)\beta|rs|}{r+s}, \tag{2.6}$$

$$\delta := \delta(r, s, \beta) = \min\{\delta_1, \cdots, \delta_m, \hat{\delta}\}. \tag{2.7}$$

In this work, we assume that the sequence $\{w^k\}$ is generated by the BPSADMM. Next, we present the first-order optimality conditions for the subproblems in BPSADMM, which will be used in our convergence analysis below.

$$\begin{aligned} 0 \in& \partial f_i(x_i^{k+1}) - A_i^\top \lambda^k + \beta A_i^\top \Big(\sum_{q=1}^{i} A_q x_q^{k+1} + \sum_{q=i+1}^{m} A_q x_q^k + y^k - b \Big) \\ &+ \nabla \phi_i(x_i^{k+1}) - \nabla \phi_i(x_i^k), \qquad i = 1, \cdots, j, \end{aligned} \tag{2.8}$$

$$\begin{aligned} 0 \in& \partial f_i(x_i^{k+1}) - A_i^\top \lambda^{k+\frac{1}{2}} + \beta A_i^\top \Big(\sum_{q=1}^{i} A_q x_q^{k+1} + \sum_{q=i+1}^{m} A_q x_q^k + y^k - b \Big) \\ &+ \nabla \phi_i(x_i^{k+1}) - \nabla \phi_i(x_i^k), \qquad i = j+1, \cdots, m, \end{aligned} \tag{2.9}$$

$$0 = \nabla g(y^{k+1}) - \lambda^{k+\frac{1}{2}} + \beta \Big(\sum_{i=1}^{m} A_i x_i^{k+1} + y^{k+1} - b \Big) + \nabla \hat{\phi}(y^{k+1}) - \nabla \hat{\phi}(y^k). \tag{2.10}$$

In addition, our convergence analysis is largely based on the following potential function:

$$\hat{\mathcal{L}}_\beta(\hat{w}) = \hat{\mathcal{L}}_\beta(\mathbf{x}_{[1:m]}, y, \lambda, \hat{y}) = \mathcal{L}_\beta(\mathbf{x}_{[1:m]}, y, \lambda) + \delta_0 \|y - \hat{y}\|^2, \qquad \delta_0 = \frac{(m+2)L_{\hat{\phi}}^2}{(r+s)\beta}, \tag{2.11}$$

and assume that $\mathcal{L}_\beta(w^0) = \hat{\mathcal{L}}_\beta(\hat{w}^0) < +\infty$ holds.

Lemma 2.1. *If Assumption H holds, then*

$$\hat{\mathcal{L}}_\beta(\hat{w}^{k+1}) - \hat{\mathcal{L}}_\beta(\hat{w}^k) \le -\delta(\|\mathbf{x}_{[1:m]}^{k+1} - \mathbf{x}_{[1:m]}^k\|^2 + \|y^{k+1} - y^k\|^2), \tag{2.12}$$

where δ is shown in (2.7). Moreover, the sequence $\{\hat{\mathcal{L}}_\beta(\hat{w}^k)\}$ decreases monotonically.

Proof. First, using the fact that $x_1^{k+1}, \cdots, x_m^{k+1}, y^{k+1}$ are minimizers in (1.7)–(1.9) and (1.11)–(1.13), respectively, we obtain

$$\begin{aligned} &\mathcal{L}_\beta(x_1^{k+1}, \mathbf{x}_{[2:m]}^k, y^k, \lambda^k) - \mathcal{L}_\beta(w^k) \le -\triangle_{\phi_1}(x_1^{k+1}, x_1^k), \\ &\qquad \vdots \\ &\mathcal{L}_\beta(\mathbf{x}_{[1:j]}^{k+1}, \mathbf{x}_{[j+1:m]}^k, y^k, \lambda^k) - \mathcal{L}_\beta(\mathbf{x}_{[1:j-1]}^{k+1}, \mathbf{x}_{[j:m]}^k, y^k, \lambda^k) \le -\triangle_{\phi_j}(x_j^{k+1}, x_j^k), \\ &\mathcal{L}_\beta(\mathbf{x}_{[1:j+1]}^{k+1}, \mathbf{x}_{[j+2:m]}^k, y^k, \lambda^{k+\frac{1}{2}}) - \mathcal{L}_\beta(\mathbf{x}_{[1:j]}^{k+1}, \mathbf{x}_{[j+1:m]}^k, y^k, \lambda^{k+\frac{1}{2}}) \le -\triangle_{\phi_{j+1}}(x_{j+1}^{k+1}, x_{j+1}^k), \\ &\qquad \vdots \\ &\mathcal{L}_\beta(\mathbf{x}_{[1:m]}^{k+1}, y^k, \lambda^{k+\frac{1}{2}}) - \mathcal{L}_\beta(\mathbf{x}_{[1:m-1]}^{k+1}, x_m^k, y^k, \lambda^{k+\frac{1}{2}}) \le -\triangle_{\phi_m}(x_m^{k+1}, x_m^k), \\ &\mathcal{L}_\beta(\mathbf{x}_{[1:m]}^{k+1}, y^{k+1}, \lambda^{k+\frac{1}{2}}) - \mathcal{L}_\beta(\mathbf{x}_{[1:m]}^{k+1}, y^k, \lambda^{k+\frac{1}{2}}) \le -\triangle_{\hat{\phi}}(y^{k+1}, y^k). \end{aligned}$$

Again, summing up the inequalities above, and combining the the strong convexities of $\triangle_{\phi_i}$ ($i = 1, \cdots, m$) and $\triangle_{\hat{\phi}}$, we have

$$\mathcal{L}_\beta(\mathbf{x}_{[1:m]}^{k+1}, y^{k+1}, \lambda^{k+\frac{1}{2}}) - \mathcal{L}_\beta(\mathbf{x}_{[1:j]}^{k+1}, \mathbf{x}_{[j+1:m]}^k, y^k, \lambda^{k+\frac{1}{2}})$$

$$
\begin{aligned}
&+\mathcal{L}_\beta(\mathbf{x}_{[1:j]}^{k+1}, \mathbf{x}_{[j+1:m]}^{k}, y^k, \lambda^k) - \mathcal{L}_\beta(w^k) \\
\leq & -\sum_{i=1}^{m} \triangle_{\phi_i}(x_i^{k+1}, x_i^k) - \triangle_{\hat{\phi}}(y^{k+1}, y^k) \\
\leq & -\sum_{i=1}^{m} \frac{\sigma_i}{2}\|x_i^{k+1} - x_i^k\|^2 - \frac{\hat{\sigma}}{2}\|y^{k+1} - y^k\|^2.
\end{aligned} \tag{2.13}
$$

Second, from relations (1.4), (1.14) and (1.10), it follows that

$$
\mathcal{L}_\beta(w^{k+1}) = \mathcal{L}_\beta(\mathbf{x}_{[1:m]}^{k+1}, y^{k+1}, \lambda^{k+\frac{1}{2}}) + s\beta\|A\mathbf{x}_{[1:m]}^{k+1} + y^{k+1} - b\|^2 \tag{2.14}
$$

and

$$
\begin{aligned}
&\mathcal{L}_\beta(\mathbf{x}_{[1:j]}^{k+1}, \mathbf{x}_{[j+1:m]}^{k}, y^k, \lambda^{k+\frac{1}{2}}) \\
= \ & \mathcal{L}_\beta(\mathbf{x}_{[1:j]}^{k+1}, \mathbf{x}_{[j+1:m]}^{k}, y^k, \lambda^k) + r\beta\Big\|\sum_{i=1}^{j} A_i x_i^{k+1} + \sum_{i=j+1}^{m} A_i x_i^k + y^k - b\Big\|^2.
\end{aligned} \tag{2.15}
$$

Third, summing up (2.13)–(2.15), we have

$$
\begin{aligned}
\mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^k) \leq & -\sum_{i=1}^{m} \frac{\sigma_i}{2}\|x_i^{k+1} - x_i^k\|^2 - \frac{\hat{\sigma}}{2}\|y^{k+1} - y^k\|^2 \\
& + s\beta\|A\mathbf{x}_{[1:m]}^{k+1} + y^{k+1} - b\|^2 \\
& + r\beta\Big\|\sum_{i=1}^{j} A_i x_i^{k+1} + \sum_{i=j+1}^{m} A_i x_i^k + y^k - b\Big\|^2.
\end{aligned} \tag{2.16}
$$

To proceed, from (1.10) and (1.14), one has

$$
\begin{aligned}
& A\mathbf{x}_{[1:m]}^{k+1} + y^{k+1} - b \\
= & \frac{1}{(r+s)\beta}(\lambda^k - \lambda^{k+1}) + \frac{r}{r+s}\Big(\sum_{i=j+1}^{m} A_i(x_i^{k+1} - x_i^k) + (y^{k+1} - y^k)\Big)
\end{aligned} \tag{2.17}
$$

and

$$
\begin{aligned}
& \sum_{i=1}^{j} A_i x_i^{k+1} + \sum_{i=j+1}^{m} A_i x_i^k + y^k - b \\
= \ & \tfrac{1}{(r+s)\beta}(\lambda^k - \lambda^{k+1}) - \tfrac{s}{r+s}\Big(\sum_{i=j+1}^{m} A_i(x_i^{k+1} - x_i^k) + (y^{k+1} - y^k)\Big).
\end{aligned} \tag{2.18}
$$

Moreover, from the relations above (2.17) and (2.18), we obtain

$$
\begin{aligned}
& s\beta\|A\mathbf{x}_{[1:m]}^{k+1} + y^{k+1} - b\|^2 + r\beta\Big\|\sum_{i=1}^{j} A_i x_i^{k+1} + \sum_{i=j+1}^{m} A_i x_i^k + y^k - b\Big\|^2 \\
= & \frac{1}{(r+s)\beta}\|\lambda^k - \lambda^{k+1}\|^2 + \frac{rs\beta}{r+s}\Big\|\sum_{i=j+1}^{m} A_i(x_i^{k+1} - x_i^k) + (y^{k+1} - y^k)\Big\|^2 \\
\leq & \frac{1}{(r+s)\beta}\|\lambda^k - \lambda^{k+1}\|^2 \\
& + \frac{(m+1-j)\beta|rs|}{r+s}\Big(\sum_{i=j+1}^{m} \|A_i(x_i^{k+1} - x_i^k)\|^2 + \|y^{k+1} - y^k\|^2\Big).
\end{aligned} \tag{2.19}
$$

On the other hand, recalling (1.14) and the optimality condition (2.10) of the y-subproblem, we have

$$\lambda^{k+1} = \nabla g(y^{k+1}) + \beta(1-s)(A\mathbf{x}_{[1:m]}^{k+1} + y^{k+1} - b) + \nabla\hat{\phi}(y^{k+1}) - \nabla\hat{\phi}(y^k), \tag{2.20}$$

and using Assumption H (i) and (iv), then

$$\begin{aligned}
\|\lambda^{k+1} - \lambda^k\| =& \Big\| [\nabla g(y^{k+1}) - \nabla g(y^k)] + \beta(1-s)\Big(\sum_{i=1}^m A_i(x_i^{k+1} - x_i^k) + (y^{k+1} - y^k)\Big) \\
&+ [\nabla\hat{\phi}(y^{k+1}) - \nabla\hat{\phi}(y^k)] - [\nabla\hat{\phi}(y^k) - \nabla\hat{\phi}(y^{k-1})] \Big\| \\
\le& L_g\|y^{k+1} - y^k\| + \beta|1-s|\Big(\sum_{i=1}^m \|A_i(x_i^{k+1} - x_i^k)\| + \|y^{k+1} - y^k\|\Big) \\
&+ L_{\hat{\phi}}\|y^{k+1} - y^k\| + L_{\hat{\phi}}\|y^k - y^{k-1}\| \\
\le& (L_g + \beta|1-s| + L_{\hat{\phi}})\|y^{k+1} - y^k\| + \beta|1-s|\sum_{i=1}^m \|A_i(x_i^{k+1} - x_i^k)\| \\
&+ L_{\hat{\phi}}\|y^k - y^{k-1}\|.
\end{aligned} \tag{2.21}$$

Furthermore, we have

$$\begin{aligned}
\frac{\|\lambda^{k+1} - \lambda^k\|^2}{m+2} \le& (L_g + \beta|1-s| + L_{\hat{\phi}})^2\|y^{k+1} - y^k\|^2 + L_{\hat{\phi}}^2\|y^k - y^{k-1}\|^2 \\
&+ \beta^2(1-s)^2\sum_{i=1}^m \|A_i(x_i^{k+1} - x_i^k)\|^2.
\end{aligned} \tag{2.22}$$

Next, substituting (2.22) into (2.19), we get

$$\begin{aligned}
&s\beta\|A\mathbf{x}_{[1:m]}^{k+1} + y^{k+1} - b\|^2 + r\beta\|\sum_{i=1}^j A_i x_i^{k+1} + \sum_{i=j+1}^m A_i x_i^k + y^k - b\|^2 \\
\le& \frac{(m+2)\beta(1-s)^2}{r+s}\sum_{i=1}^j \|A_i(x_i^{k+1} - x_i^k)\|^2 \\
&+ \frac{(m+2)\beta(1-s)^2 + (m+1-j)\beta|rs|}{r+s}\sum_{i=j+1}^m \|A_i(x_i^{k+1} - x_i^k)\|^2 \\
&+ \Big(\frac{(m+2)(L_g + \beta|1-s| + L_{\hat{\phi}})^2}{(r+s)\beta} + \frac{(m+1-j)\beta|rs|}{r+s}\Big)\|y^{k+1} - y^k\|^2 \\
&+ \frac{(m+2)L_{\hat{\phi}}^2}{(r+s)\beta}\|y^k - y^{k-1}\|^2,
\end{aligned} \tag{2.23}$$

and taking into account (2.16), it follows that

$$\begin{aligned}
&\mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^k) \\
\le& -\sum_{i=1}^m \frac{\sigma_i}{2}\|x_i^{k+1} - x_i^k\|^2 - \frac{\hat{\sigma}}{2}\|y^{k+1} - y^k\|^2 + \frac{(m+2)\beta(1-s)^2}{r+s}\sum_{i=1}^j \|A_i(x_i^{k+1} - x_i^k)\|^2 \\
&+ \frac{(m+2)\beta(1-s)^2 + (m+1-j)\beta|rs|}{r+s}\sum_{i=j+1}^m \|A_i(x_i^{k+1} - x_i^k)\|^2
\end{aligned}$$

$$
\begin{aligned}
&+\Big(\frac{(m+2)(L_g+\beta|1-s|+L_{\hat{\phi}})^2}{(r+s)\beta}+\frac{(m+1-j)\beta|rs|}{r+s}\Big)\|y^{k+1}-y^k\|^2\\
&+\frac{(m+2)L_{\hat{\phi}}^2}{(r+s)\beta}\|y^k-y^{k-1}\|^2,
\end{aligned}\tag{2.24}
$$

which implies

$$
\begin{aligned}
&\Big[\mathcal{L}_\beta(w^{k+1})+\frac{(m+2)L_{\hat{\phi}}^2}{(r+s)\beta}\|y^{k+1}-y^k\|^2\Big]-\Big[\mathcal{L}_\beta(w^k)+\frac{(m+2)L_{\hat{\phi}}^2}{(r+s)\beta}\|y^k-y^{k-1}\|^2\Big]\\
\le&-\sum_{i=1}^{j}\Big(\frac{\sigma_i}{2}-\frac{(m+2)\beta(1-s)^2\lambda_{\max}(A_i^TA_i)}{r+s}\Big)\|x_i^{k+1}-x_i^k\|^2\\
&-\sum_{i=j+1}^{m}\Big(\frac{\sigma_i}{2}-\frac{[(m+2)\beta(1-s)^2+(m+1-j)\beta|rs|]\lambda_{\max}(A_i^TA_i)}{r+s}\Big)\|x_i^{k+1}-x_i^k\|^2\\
&-\Big(\frac{\hat{\sigma}}{2}-\frac{(m+2)[(L_g+\beta|1-s|+L_{\hat{\phi}})^2-L_{\hat{\phi}}^2]}{(r+s)\beta}-\frac{(m+1-j)\beta|rs|}{r+s}\Big)\|y^{k+1}-y^k\|^2.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\hat{\mathcal{L}}_\beta(\hat{w}^{k+1})-\hat{\mathcal{L}}_\beta(\hat{w}^k)\le&-\sum_{i=1}^{m}\delta_i\|x^{k+1}-x^k\|^2-\hat{\delta}\|y^{k+1}-y^k\|^2\\
\le&-\delta(\|\mathbf{x}_{[1:m]}^{k+1}-\mathbf{x}_{[1:m]}^k\|^2+\|y^{k+1}-y^k\|^2),
\end{aligned}
$$

and relation (2.12) is immediately at hand.

Finally, from the definition of δ in (2.7), as well as the restrictions of strong convex coefficients σ_i $(i=1,\cdots,m)$ and $\hat{\sigma}$ in (2.1)–(2.3) (see Assumption H (iii) and (iv)), then it is easily known that $\delta>0$. Hence, the sequence $\{\hat{\mathcal{L}}_\beta(\hat{w}^k)\}$ is monotonically nonincreasing. This completes the proof. □

Lemma 2.2. *Suppose that the sequence $\{w^k\}$ is bounded. If Assumption H holds, then*

$$
\lim_{k\to+\infty}\|\hat{w}^{k+1}-\hat{w}^k\|=\lim_{k\to+\infty}(\|\mathbf{x}_{[1:m]}^{k+1}-\mathbf{x}_{[1:m]}^k\|+\|y^{k+1}-y^k\|+\|\lambda^{k+1}-\lambda^k\|)=0. \tag{2.25}
$$

Proof. Since $\{w^k\}$ is bounded, we obtain that the $\{\hat{w}^k\}$ is bounded, and thus a cluster point exists. Suppose that $\hat{w}^*$ is a cluster point of $\{\hat{w}^k\}$ and let $\{\hat{w}^{k_j}\}$ be a convergent subsequence such that $\lim\limits_{k_j\to+\infty}\hat{w}^{k_j}=\hat{w}^*$. In view of the fact that f_i $(i=1,2,\cdots,m)$ are lower semicontinuous and g is continuous, the function $\hat{\mathcal{L}}_\beta(\cdot)$ is lower semicontinuous, and hence

$$
\hat{\mathcal{L}}_\beta(\hat{w}^*)\le\liminf_{k_j\to+\infty}\hat{\mathcal{L}}_\beta(\hat{w}^{k_j+1}). \tag{2.26}
$$

Together with the fact that $\{\hat{\mathcal{L}}_\beta(\hat{w}^k)\}$ is nonincreasing, we know that $\{\hat{\mathcal{L}}_\beta(\hat{w}^k)\}$ is convergent and $\hat{\mathcal{L}}_\beta(\hat{w}^k)\ge\hat{\mathcal{L}}_\beta(\hat{w}^*)$ for any $k\ge0$. Rearranging terms in relation (2.12) and summing up it from $k=0$ to $k=l$, it follows

$$
\delta\sum_{k=0}^{l}(\|\mathbf{x}_{[1:m]}^{k+1}-\mathbf{x}_{[1:m]}^k\|^2+\|y^{k+1}-y^k\|^2)\le\hat{\mathcal{L}}_\beta(\hat{w}^0)-\hat{\mathcal{L}}_\beta(\hat{w}^{l+1})\le\hat{\mathcal{L}}_\beta(\hat{w}^0)-\hat{\mathcal{L}}_\beta(\hat{w}^*)<+\infty.
$$

This, together with $\delta > 0$, implies that

$$\sum_{k=0}^{+\infty} \|\mathbf{x}_{[1:m]}^{k+1} - \mathbf{x}_{[1:m]}^{k}\|^2 < +\infty, \qquad \sum_{k=0}^{+\infty} \|y^{k+1} - y^k\|^2 < +\infty. \tag{2.27}$$

Hence,

$$\mathbf{x}_{[1:m]}^{k+1} - \mathbf{x}_{[1:m]}^{k} \to 0, \qquad y^{k+1} - y^k \to 0. \tag{2.28}$$

Next, it follows from (2.21) and (2.27) that $\sum\limits_{k=0}^{+\infty} \|\lambda^{k+1} - \lambda^k\|^2 < +\infty$. Therefore, we conclude that

$$\lambda^{k+1} - \lambda^k \to 0. \tag{2.29}$$

Further, from (2.28) and (2.29), we obtain (2.25), and this completes the proof. □

Theorem 2.3 (Global convergence)**.** *Let Ω denote the cluster point set of the sequence $\{w^k\}$, and $\hat{\Omega}$ denote the cluster point set of the sequence $\{\hat{w}^k\}$. If Assumption H holds, and $\{w^k\}$ is bounded, then*

(i) Ω and $\hat{\Omega}$ are nonempty compact sets, and $d(w^k, \Omega) \to 0, \quad d(\hat{w}^k, \hat{\Omega}) \to 0, \quad k \to +\infty$;

(ii) $\Omega \subseteq \mathrm{crit}\mathcal{L}_\beta$;

(iii) $\hat{w}^* \in \hat{\Omega}$ if and only if $\hat{y}^* = y^*$ and $w^* \in \Omega$;

(iv) the whole potential function sequence $\{\hat{\mathcal{L}}_\beta(\hat{w}^k)\}$ is convergent, and $\hat{\mathcal{L}}_\beta(\hat{w}^*) = \lim\limits_{k\to+\infty} \hat{\mathcal{L}}_\beta(\hat{w}^k)$ $= \inf\limits_k \hat{\mathcal{L}}_\beta(\hat{w}^k)$ for all $\hat{w}^* \in \hat{\Omega}$. Moreover, $\hat{\mathcal{L}}_\beta(\cdot)$ is finite and constant on $\hat{\Omega}$.

Proof. (i) By the definitions of Ω and $\hat{\Omega}$, they are trivial.

(ii) Let $w^* \in \Omega$, then there exists a subsequence $\{w^{k_j}\}$ of $\{w^k\}$ converging to w^*. It follows from (2.25) that $\lim\limits_{k\to+\infty} \|w^{k+1} - w^k\| = 0$, and thus $\lim\limits_{j\to+\infty} w^{k_j+1} = \lim\limits_{j\to+\infty} w^{k_j} = w^*$. Let $\lim\limits_{j\to+\infty} \lambda^{k_j+\frac{1}{2}} = \lambda^{**}$. Letting $k := k_j$ and taking the limit $j \to +\infty$ in (1.10) and (1.14), we have

$$\lambda^{**} = \lambda^* - r\beta\Big(\sum_{i=1}^m A_i x_i^* + y^* - b\Big), \qquad \lambda^* = \lambda^{**} - s\beta\Big(\sum_{i=1}^m A_i x_i^* + y^* - b\Big).$$

This, together with $r + s > 0$, yields

$$\sum_{i=1}^m A_i x_i^* + y^* - b = 0, \qquad \lambda^{**} = \lambda^*. \tag{2.30}$$

Therefore, $(\mathbf{x}_{[1:m]}^*, y^*)$ is a feasible point of problem (1.1). Since $x_i^{k_j+1}$ $(i = 1, 2, \cdots, j)$ is a minimizer in x_i-subproblem, it holds that

$$\begin{aligned}
&f_i(x_i^{k_j+1}) - \langle \lambda^{k_j}, A_i x_i^{k_j+1} \rangle + \frac{\beta}{2}\Big\| \sum_{q=1}^{i} A_q x_q^{k_j+1} + \sum_{q=i+1}^{m} A_q x_q^{k_j} - b \Big\|^2 + \triangle_{\phi_i}(x^{k_j+1}, x^{k_j}) \\
\le & f_i(x_i^*) - \langle \lambda^{k_j}, A_i x_i^* \rangle + \frac{\beta}{2}\Big\| \sum_{q=1}^{i-1} A_q x_q^{k_j+1} + A_i x_i^* + \sum_{q=i+1}^{m} A_q x_q^{k_j} - b \Big\|^2 + \triangle_{\phi_i}(x^*, x^{k_j}).
\end{aligned}$$

This, together with $\lim_{j\to+\infty} w^{k_j} = \lim_{j\to+\infty} w^{k_j+1} = w^*$, implies $\limsup_{j\to+\infty} f_i(x_i^{k_j+1}) \le f_i(x_i^*)$ from inequality above. Noticing that the lower semicontinuity of f_i $(i = 1, 2, \cdots, j)$, one has, $f_i(x_i^*) \le \liminf_{j\to+\infty} f_i(x_i^{k_j+1})$. Hence,

$$\lim_{j\to+\infty} f_i(x_i^{k_j+1}) = f_i(x_i^*), \qquad i = 1, 2, \cdots, j. \tag{2.31}$$

Similarly, we also get

$$\lim_{j\to+\infty} f_i(x_i^{k_j+1}) = f_i(x_i^*), \qquad i = j+1, 2, \cdots, m. \tag{2.32}$$

On the other hand, it follows from $\lim_{j\to+\infty} \|x_i^{k_j+1} - x_i^{k_j}\| = 0$ and Assumption H (iii), that $\lim_{j\to+\infty} \|\nabla\phi_i(x_i^{k_j+1}) - \nabla\phi_i(x_i^{k_j})\| = 0$ for $i = 1, 2, \cdots, m$. Similarly, $\lim_{j\to+\infty} \|\nabla\hat{\phi}(y^{k_j+1}) - \nabla\hat{\phi}(y^{k_j})\| = 0$. Thus, letting $k := k_j$ and taking the limit $j \to +\infty$ in (2.8)–(2.10), and invoking the closeness of ∂f_i $(i = 1, 2, \cdots, m)$, the continuity of ∇g, as well as (2.30), we have

$$A_i^\top \lambda^* \in \partial f_i(x_i^*), \quad i = 1, 2, \cdots, m, \qquad \nabla g(y^*) = \lambda^*, \qquad \sum_{i=1}^m A_i x_i^* + y^* - b = 0.$$

That is $w^* \in \mathrm{crit}\mathcal{L}_\beta$, and this proves statement (ii).

(iii) From Lemma 2.2 and the definitions of w^k and $\hat{w}^k$, we get statement (iii).

(iv) Let $\hat{w}^* \in \hat{\Omega}$. Then there exists at least one subsequence $\{\hat{w}^{k_j}\}$ of $\{\hat{w}^k\}$ converging to $\hat{w}^*$. And from (1.3), (2.31) and (2.32), we have $\lim_{j\to+\infty} \hat{\mathcal{L}}_\beta(\hat{w}^{k_j+1}) = \hat{\mathcal{L}}_\beta(\hat{w}^*)$. This, together with the monotonity of $\{\hat{\mathcal{L}}_\beta(\hat{w}^k)\}$, further shows that the whole potential function sequence $\{\hat{\mathcal{L}}_\beta(\hat{w}^k)\}$ is convergent. Thus, since $\hat{\mathcal{L}}_\beta(\hat{w}^k) \le \hat{\mathcal{L}}_\beta(\hat{w}^0) < +\infty$, the relations $+\infty > \hat{\mathcal{L}}_\beta(\hat{w}^0) \ge \lim_{k\to+\infty} \hat{\mathcal{L}}_\beta(\hat{w}^k) = \inf_k \hat{\mathcal{L}}_\beta(\hat{w}^k) = \hat{\mathcal{L}}_\beta(\hat{w}^*)$ hold. Then, $\hat{\mathcal{L}}_\beta(\hat{w}^*) \equiv \lim_{k\to+\infty} \hat{\mathcal{L}}_\beta(\hat{w}^k) < +\infty$ for all $\hat{w}^* \in \hat{\Omega}$. □

Theorem 2.4 (Strong convergence)**.** *Suppose that* $\{w^k\}$ *is bounded. If Assumption H holds, and* $\hat{\mathcal{L}}_\beta(\hat{w})$ *is a KL function, then* $\sum_{k=0}^{+\infty} \|w^{k+1} - w^k\| < +\infty$. *Furthermore,* $\{w^k\}$ *converges to a critical point of* $\mathcal{L}_\beta(\cdot)$.

Proof. From Theorem 2.3 (iv), we have $\lim_{k\to+\infty} \hat{\mathcal{L}}_\beta(\hat{w}^k) = \hat{\mathcal{L}}_\beta(\hat{w}^*)$ for all $\hat{w}^* \in \hat{\Omega}$. In the following, we consider two cases.

Case I. Suppose that $\hat{\mathcal{L}}_\beta(\hat{w}^{k_0}) = \hat{\mathcal{L}}_\beta(\hat{w}^*)$ for some integer k_0. Associated with Lemma 2.1, for any $k > k_0$, we have

$$\delta(\|\mathbf{x}_{[1:m]}^{k+1} - \mathbf{x}_{[1:m]}^k\|^2 + \|y^{k+1} - y^k\|^2) \le \hat{\mathcal{L}}_\beta(\hat{w}^k) - \hat{\mathcal{L}}_\beta(\hat{w}^{k+1}) \le \hat{\mathcal{L}}_\beta(\hat{w}^{k_0}) - \hat{\mathcal{L}}_\beta(\hat{w}^*) = 0.$$

Thus, $\mathbf{x}_{[1:m]}^{k+1} = \mathbf{x}_{[1:m]}^k$ and $y^{k+1} = y^k$ for any $k > k_0$. Together with (2.21), for any $k > k_0 + 1$, one has, $w^{k+1} = w^k$ and the assertion holds.

Case II. Now, we consider the case where $\hat{\mathcal{L}}_\beta(\hat{w}^k) > \hat{\mathcal{L}}_\beta(\hat{w}^*)$ for all $k \ge 0$. In this case, we will divide the proof into three steps: (1) we first apply the uniformized KL property to $\hat{\mathcal{L}}_\beta(\hat{w})$; (2) we bound the distance from 0 to $\partial\hat{\mathcal{L}}_\beta(\hat{w}^k)$; (3) we show that the sequence $\{w^k\}$ is a Cauchy sequence and hence is convergent. The complete proof is presented as follows.

Step 1. In view of the fact that $\lim_{k\to+\infty} d(\hat{w}^k, \hat{\Omega}) = 0$, it follows that for $\epsilon > 0$, there exists $k_1 > 0$, such that for any $k > k_1$, $d(\hat{w}^k, \hat{\Omega}) < \epsilon$. And since $\lim_{k\to+\infty} \hat{\mathcal{L}}_\beta(\hat{w}^k) = \hat{\mathcal{L}}_\beta(\hat{w}^*)$, then for $\eta > 0$, there exists $k_2 > 0$, such that for any $k > k_2$, $\hat{\mathcal{L}}_\beta(\hat{w}^k) < \hat{\mathcal{L}}_\beta(\hat{w}^*) + \eta$. Thus, when $k > \tilde{k} = \max\{k_1, k_2\}$, we have for $\epsilon,\ \eta > 0$, $d(\hat{w}^k, \hat{\Omega}) < \epsilon$, $\hat{\mathcal{L}}_\beta(\hat{w}^*) < \hat{\mathcal{L}}_\beta(\hat{w}^k) < \hat{\mathcal{L}}_\beta(\hat{w}^*) + \eta$. From Theorem 2.3, we know that $\hat{\mathcal{L}}_\beta(\cdot)$ is constant on $\hat{\Omega}$, and $\hat{\Omega}$ is a nonempty compact set. Thus, by uniformized KL property in Lemma 1.4, we obtain

$$\varphi'(\hat{\mathcal{L}}_\beta(\hat{w}^k) - \hat{\mathcal{L}}_\beta(\hat{w}^*))d(0, \partial\hat{\mathcal{L}}_\beta(\hat{w}^k)) \geq 1, \qquad \text{for all} \ \ k > \tilde{k}. \tag{2.33}$$

Step 2. We first consider the subdifferential $\partial\hat{\mathcal{L}}_\beta(\hat{w}^k)$. From the definition of $\hat{\mathcal{L}}_\beta(\hat{w})$ in (2.11), and from the partial subdifferential with respect to x_i $(i = 1, \cdots, j)$, we have

$$\begin{aligned}\partial_{x_i}\hat{\mathcal{L}}_\beta(\hat{w}^{k+1}) =& \partial f_i(x_i^{k+1}) - A_i^\top\lambda^{k+1} + \beta A_i^\top\Big(\sum_{q=1}^m A_q x_q^{k+1} + y^{k+1} - b\Big)\\ =& A_i^\top(\lambda^k - \lambda^{k+1}) + \beta A_i^\top\Big(\sum_{q=i+1}^m A_q(x_q^{k+1} - x_q^k) + (y^{k+1} - y^k)\Big)\\ &- [\nabla\phi_i(x_i^{k+1}) - \nabla\phi_i(x_i^k)], \qquad i = 1, \cdots, j,\end{aligned} \tag{2.34}$$

where the second equality follows from the optimality condition (2.8). Similarly,

$$\begin{aligned}\partial_{x_i}\hat{\mathcal{L}}_\beta(\hat{w}^{k+1}) =& \partial f_i(x_i^{k+1}) - A_i^\top\lambda^{k+1} + \beta A_i^\top\Big(\sum_{q=1}^m A_q x_q^{k+1} + y^{k+1} - b\Big)\\ =& A_i^\top(\lambda^{k+\frac{1}{2}} - \lambda^{k+1}) - [\nabla\phi_i(x_i^{k+1}) - \nabla\phi_i(x_i^k)]\\ &+ \beta A_i^\top\Big(\sum_{q=i+1}^m A_q(x_q^{k+1} - x_q^k) + (y^{k+1} - y^k)\Big)\\ =& s\beta A_i^\top\Big(\sum_{q=1}^m A_q x_q^{k+1} + y^{k+1} - b\Big) - [\nabla\phi_i(x_i^{k+1}) - \nabla\phi_i(x_i^k)]\\ &+ \beta A_i^\top\Big(\sum_{q=i+1}^m A_q(x_q^{k+1} - x_q^k) + (y^{k+1} - y^k)\Big)\\ =& \frac{s}{r+s}A_i^\top(\lambda^k - \lambda^{k+1}) - [\nabla\phi_i(x_i^{k+1}) - \nabla\phi_i(x_i^k)]\\ &+ \frac{rs\beta}{r+s}A_i^\top\Big(\sum_{q=j+1}^m A_q(x_q^{k+1} - x_q^k) + (y^{k+1} - y^k)\Big)\\ &+ \beta A_i^\top\Big(\sum_{q=i+1}^m A_q(x_q^{k+1} - x_q^k) + (y^{k+1} - y^k)\Big), \qquad i = j+1, \cdots, m,\end{aligned} \tag{2.35}$$

where the second equality follows from the optimality condition (2.9), the third equality follows from (1.14), and the last equality follows from (2.17). Moreover, it follows from the definition of $\hat{\mathcal{L}}_\beta(\hat{w})$, (1.14), (2.10) as well as (2.17), that

$$\begin{aligned}\partial_y\hat{\mathcal{L}}_\beta(\hat{w}^{k+1}) =& \nabla g(y^{k+1}) - \lambda^{k+1} + \beta\Big(\sum_{q=1}^m A_q x_q^{k+1} + y^{k+1} - b\Big) + 2\delta_0(y^{k+1} - y^k)\\ \overset{(1.14),(2.10)}{=}& s\beta\Big(\sum_{i=1}^m A_i x_i^{k+1} + y^{k+1} - b\Big) - [\nabla\hat{\phi}(y^{k+1}) - \nabla\hat{\phi}(y^k)] + 2\delta_0(y^{k+1} - y^k)\end{aligned}$$

$$\overset{(2.17)}{=}\frac{s}{r+s}(\lambda^k-\lambda^{k+1})+\frac{rs\beta}{r+s}\Big(\sum_{q=j+1}^{m}A_q(x_q^{k+1}-x_q^k)+(y^{k+1}-y^k)\Big)$$
$$-[\nabla\hat{\phi}(y^{k+1})-\nabla\hat{\phi}(y^k)]+2\delta_0(y^{k+1}-y^k). \tag{2.36}$$

Additionally, we also get

$$\partial_\lambda\hat{\mathcal{L}}_\beta(\hat{w}^{k+1})=-\Big(\sum_{q=1}^{m}A_q x_q^{k+1}+y^{k+1}-b\Big)$$
$$\overset{(2.17)}{=}\frac{1}{(r+s)\beta}(\lambda^{k+1}-\lambda^k)-\frac{r}{r+s}\Big(\sum_{q=j+1}^{m}A_q(x_q^{k+1}-x_q^k)+(y^{k+1}-y^k)\Big) \tag{2.37}$$

and

$$\partial_{\hat{y}}\hat{\mathcal{L}}_\beta(\hat{w}^{k+1})=-2\delta_0(y^{k+1}-y^k). \tag{2.38}$$

Thus, from the relations (2.34)–(2.38), and the Lipschitz continuity of $\nabla\phi_i$ $(i=1,2,\cdots,m)$ and $\nabla\hat{\phi}$, there exists $\zeta_1>0$ such that

$$d(0,\partial\hat{\mathcal{L}}_\beta(\hat{w}^{k+1}))\le\zeta_1(\|\mathbf{x}_{[1:m]}^{k+1}-\mathbf{x}_{[1:m]}^k\|+\|y^{k+1}-y^k\|+\|\lambda^{k+1}-\lambda^k\|).$$

Again, it follows from (2.21), there exists $\zeta_2>0$ such that

$$\|\lambda^{k+1}-\lambda^k\|\le\zeta_2(\|\mathbf{x}_{[1:m]}^{k+1}-\mathbf{x}_{[1:m]}^k\|+\|y^{k+1}-y^k\|+\|y^k-y^{k-1}\|).$$

Then, by the two inequalities above, we have

$$\begin{aligned}d(0,\partial\mathcal{L}_\beta(w^{k+1}))\le&\zeta_1(1+\zeta_2)(\|\mathbf{x}_{[1:m]}^{k+1}-\mathbf{x}_{[1:m]}^k\|+\|y^{k+1}-y^k\|)+\zeta_2\|y^k-y^{k-1}\|\\ \le&\zeta(\|\mathbf{x}_{[1:m]}^{k+1}-\mathbf{x}_{[1:m]}^k\|+\|y^{k+1}-y^k\|+\|y^k-y^{k-1}\|).\end{aligned}$$

with $\zeta:=\max\{\zeta_1(1+\zeta_2),\zeta_2\}$. This further implies that

$$d(0,\partial\hat{\mathcal{L}}_\beta(\hat{w}^k))\le\zeta(\|\mathbf{x}_{[1:m]}^k-\mathbf{x}_{[1:m]}^{k-1}\|+\|y^k-y^{k-1}\|+\|y^{k-1}-y^{k-2}\|). \tag{2.39}$$

Step 3. We now prove the convergence of the sequence by (2.33) and (2.39). For simplicity, define $\triangle_{p,q}=\varphi(\hat{\mathcal{L}}_\beta(\hat{w}^p)-\hat{\mathcal{L}}_\beta(\hat{w}^*))-\varphi(\hat{\mathcal{L}}_\beta(\hat{w}^q)-\hat{\mathcal{L}}_\beta(\hat{w}^*))$. Then, we obtain, for all $k>\tilde{k}$,

$$\begin{aligned}&\zeta(\|\mathbf{x}_{[1:m]}^k-\mathbf{x}_{[1:m]}^{k-1}\|+\|y^k-y^{k-1}\|+\|y^{k-1}-y^{k-2}\|)\cdot\triangle_{k,k+1}\\ \overset{(2.39)}{\ge}&d(0,\partial\hat{\mathcal{L}}_\beta(\hat{w}^k))\cdot\triangle_{k,k+1}\\ \ge&d(0,\partial\hat{\mathcal{L}}_\beta(\hat{w}^k))\cdot\varphi'(\hat{\mathcal{L}}_\beta(\hat{w}^k)-\hat{\mathcal{L}}_\beta(\hat{w}^*))(\hat{\mathcal{L}}_\beta(\hat{w}^k)-\hat{\mathcal{L}}_\beta(\hat{w}^{k+1}))\\ \overset{(2.33)}{\ge}&\hat{\mathcal{L}}_\beta(\hat{w}^k)-\hat{\mathcal{L}}_\beta(\hat{w}^{k+1})\\ \overset{(2.12)}{\ge}&\delta(\|\mathbf{x}_{[1:m]}^{k+1}-\mathbf{x}_{[1:m]}^k\|^2+\|y^{k+1}-y^k\|^2)\\ \ge&\frac{\delta}{2}(\|\mathbf{x}_{[1:m]}^{k+1}-\mathbf{x}_{[1:m]}^k\|+\|y^{k+1}-y^k\|)^2,\end{aligned} \tag{2.40}$$

where the second inequality follows from the concavity of φ, and the last inequality follows from $(a+b)^2\le 2(a^2+b^2)$.

Dividing both sides of (2.40) by $\delta/2$, and taking the square root, we have

$$\begin{aligned}&(\|\mathbf{x}_{[1:m]}^{k}-\mathbf{x}_{[1:m]}^{k-1}\|+\|y^{k}-y^{k-1}\|+\|y^{k-1}-y^{k-2}\|)^{\frac{1}{2}}\sqrt{\frac{2\zeta}{\delta}\triangle_{k,k+1}}\\ \geq&\|\mathbf{x}_{[1:m]}^{k+1}-\mathbf{x}_{[1:m]}^{k}\|+\|y^{k+1}-y^{k}\|.\end{aligned}$$

Thus,

$$\begin{aligned}&4(\|\mathbf{x}_{[1:m]}^{k+1}-\mathbf{x}_{[1:m]}^{k}\|+\|y^{k+1}-y^{k}\|)\\ \leq&2(\|\mathbf{x}_{[1:m]}^{k}-\mathbf{x}_{[1:m]}^{k-1}\|+\|y^{k}-y^{k-1}\|+\|y^{k-1}-y^{k-2}\|)^{\frac{1}{2}}\sqrt{\frac{8\zeta}{\delta}\triangle_{k,k+1}}.\end{aligned}$$

Using the inequality $2ab \leq a^2+b^2$ to get the upper bound for the right-hand side of the inequality above, we get

$$\begin{aligned}&4(\|\mathbf{x}_{[1:m]}^{k+1}-\mathbf{x}_{[1:m]}^{k}\|+\|y^{k+1}-y^{k}\|)\\ \leq&\|\mathbf{x}_{[1:m]}^{k}-\mathbf{x}_{[1:m]}^{k-1}\|+\|y^{k}-y^{k-1}\|+\|y^{k-1}-y^{k-2}\|+\frac{8\zeta}{\delta}\triangle_{k,k+1}\,.\end{aligned}\tag{2.41}$$

Summing up (2.41) for $i=k(\geq \tilde{k}+1),\cdots,q$, we have

$$\begin{aligned}&3\sum_{i=k}^{q}\|\mathbf{x}_{[1:m]}^{i+1}-\mathbf{x}_{[1:m]}^{i}\|+2\sum_{i=k}^{q}\|y^{i+1}-y^{i}\|\\ \leq&2\|y^{k}-y^{k-1}\|-2\|y^{q+1}-y^{q}\|+\|\mathbf{x}_{[1:m]}^{k}-\mathbf{x}_{[1:m]}^{k-1}\|-\|\mathbf{x}_{[1:m]}^{q+1}-\mathbf{x}_{[1:m]}^{q}\|\\ &+\|y^{k-1}-y^{k-2}\|-\|y^{q}-y^{q-1}\|+\frac{8\zeta}{\delta}\triangle_{k,q+1}\,.\end{aligned}$$

This, along with $\varphi(\hat{\mathcal{L}}_\beta(\hat{w}^{q+1})-\hat{\mathcal{L}}_\beta(\hat{w}^*))>0$, further implies

$$\begin{aligned}&3\sum_{i=k}^{q}\|\mathbf{x}_{[1:m]}^{i+1}-\mathbf{x}_{[1:m]}^{i}\|+2\sum_{i=k}^{q}\|y^{i+1}-y^{i}\|\\ \leq&2\|y^{k}-y^{k-1}\|+\|\mathbf{x}_{[1:m]}^{k}-\mathbf{x}_{[1:m]}^{k-1}\|+\|y^{k-1}-y^{k-2}\|+\frac{8\zeta}{\delta}(\varphi(\hat{\mathcal{L}}_\beta(\hat{w}^{k})-\hat{\mathcal{L}}_\beta(\hat{w}^*))).\end{aligned}$$

Thus, it follows that

$$\begin{aligned}&3\sum_{i=k}^{+\infty}\|\mathbf{x}_{[1:m]}^{i+1}-\mathbf{x}_{[1:m]}^{i}\|+2\sum_{i=k}^{+\infty}\|y^{i+1}-y^{i}\|\\ \leq&2\|y^{k}-y^{k-1}\|+\|\mathbf{x}_{[1:m]}^{k}-\mathbf{x}_{[1:m]}^{k-1}\|+\|y^{k-1}-y^{k-2}\|+\frac{8\zeta}{\delta}(\varphi(\hat{\mathcal{L}}_\beta(\hat{w}^{k})-\hat{\mathcal{L}}_\beta(\hat{w}^*))).\end{aligned}\tag{2.42}$$

Considering $k=\tilde{k}+1$ in the relation above, one has

$$\begin{aligned}&3\sum_{i=\tilde{k}+1}^{+\infty}\|\mathbf{x}_{[1:m]}^{i+1}-\mathbf{x}_{[1:m]}^{i}\|+2\sum_{i=\tilde{k}+1}^{+\infty}\|y^{i+1}-y^{i}\|\\ \leq\;&2\|y^{\tilde{k}+1}-y^{\tilde{k}}\|+\|\mathbf{x}_{[1:m]}^{\tilde{k}+1}-\mathbf{x}_{[1:m]}^{\tilde{k}}\|+\|y^{\tilde{k}}-y^{\tilde{k}-1}\|+\tfrac{8\zeta}{\delta}(\varphi(\hat{\mathcal{L}}_\beta(\hat{w}^{\tilde{k}+1})-\hat{\mathcal{L}}_\beta(\hat{w}^*))).\end{aligned}$$

From Lemma 2.2, this immediately shows that

$$3\sum_{k=\tilde{k}+1}^{+\infty}\|\mathbf{x}_{[1:m]}^{k+1}-\mathbf{x}_{[1:m]}^{k}\|+2\sum_{k=\tilde{k}+1}^{+\infty}\|y^{k+1}-y^{k}\|<+\infty.$$

Thus

$$\sum_{k=0}^{+\infty}\|\mathbf{x}_{[1:m]}^{k+1}-\mathbf{x}_{[1:m]}^{k}\|<+\infty,\qquad \sum_{k=0}^{+\infty}\|y^{k+1}-y^{k}\|<+\infty.$$

Combining this with (2.21), we further obtain $\sum\limits_{k=0}^{+\infty}\|\lambda^{k+1}-\lambda^k\|<+\infty$. Meanwhile, noticing that

$$\begin{aligned}\|w^{k+1}-w^k\| =&(\|\mathbf{x}_{[1:m]}^{k+1}-\mathbf{x}_{[1:m]}^{k}\|^2+\|y^{k+1}-y^k\|^2+\|\lambda^{k+1}-\lambda^k\|^2)^{\frac{1}{2}}\\ \le&\|\mathbf{x}_{[1:m]}^{k+1}-\mathbf{x}_{[1:m]}^{k}\|+\|y^{k+1}-y^k\|+\|\lambda^{k+1}-\lambda^k\|,\end{aligned}$$

therefore,

$$\sum_{k=0}^{+\infty}\|w^{k+1}-w^k\|<+\infty, \tag{2.43}$$

and $\{w^k\}$ is a Cauchy sequence. Indeed, with $q>p>l$, we obtain

$$\|w^q-w^p\|=\Big\|\sum_{k=p}^{q-1}(w^{k+1}-w^k)\Big\|\le\sum_{k=p}^{q-1}\|w^{k+1}-w^k\|.$$

Since (2.43) implies that $\sum\limits_{k=l+1}^{+\infty}\|w^{k+1}-w^k\|\to 0$ as $l\to+\infty$, it follows that $\{w^k\}$ is a Cauchy sequence and hence is a convergent one. From Theorem 2.3, we know that $\{w^k\}$ converges to a critical point of $\mathcal{L}_\beta(\cdot)$. This completes the proof. □

Nextly, we give some sufficient conditions to guarantee the sequence $\{w^k\}$ is bounded.

Lemma 2.5. *Assumption H and the following conditions (i)-(iii) are true. Then, the sequence $\{w^k\}$ is bounded.*

(i) f_i $(i=1,2,\cdots,m)$ are coercive, respectively, i.e., $\liminf\limits_{\|x_i\|\to+\infty} f_i(x_i)=+\infty$;

(ii) The relaxation factors r and s satisfy $(2m+3)s-r\ge m+2$ and $s>\frac{1}{2}$;

(iii) $\bar g(y):=g(y)-\frac{1}{(2s-1)\beta}\|\nabla g(y)\|^2$ has a lower bound and it is coercive, i.e.,

$$\inf_y \bar g(y)>-\infty,\qquad \liminf_{\|y\|\to+\infty}\bar g(y)=+\infty.$$

Proof. It follows from $\hat{\mathcal{L}}_\beta(\hat w^k)$ is monotonically decreasing that $\hat{\mathcal{L}}_\beta(\hat w^k)\le\hat{\mathcal{L}}_\beta(\hat w^0)=\mathcal{L}_\beta(w^0)<+\infty$. Combining this with (2.11), (2.20) and the condition (ii), we have

$$\begin{aligned}&\mathcal{L}_\beta(w^k)+\delta_0\|y^k-y^{k-1}\|^2\\ =&\sum_{i=1}^m f_i(x_i^k)+g(y^k)-\langle\lambda^k,\sum_{i=1}^m A_ix_i^k+y^k-b\rangle+\frac{\beta}{2}\Big\|\sum_{i=1}^m A_ix_i^k+y^k-b\Big\|^2+\delta_0\|y^k-y^{k-1}\|^2\\ \overset{(2.20)}{=}&\sum_{i=1}^m f_i(x_i^k)+g(y^k)-\langle\nabla g(y^k)+\nabla\hat\phi(y^k)-\nabla\hat\phi(y^{k-1}),\sum_{i=1}^m A_ix_i^k+y^k-b\rangle\\ &+\frac{(2s-1)\beta}{2}\Big\|\sum_{i=1}^m A_ix_i^k+y^k-b\Big\|^2+\delta_0\|y^k-y^{k-1}\|^2\\ =&\sum_{i=1}^m f_i(x_i^k)+g(y^k)-\frac{1}{2(2s-1)\beta}\|\nabla g(y^k)+\nabla\hat\phi(y^k)-\nabla\hat\phi(y^{k-1})\|^2+\delta_0\|y^k-y^{k-1}\|^2\end{aligned}$$

$$
\begin{aligned}
&+\frac{(2s-1)\beta}{2}\Big\|\sum_{i=1}^m A_ix_i^k+y^k-b-\frac{1}{(2s-1)\beta}[\nabla g(y^k)+\nabla\hat{\phi}(y^k)-\nabla\hat{\phi}(y^{k-1})]\Big\|^2\\
\geq&\sum_{i=1}^m f_i(x_i^k)+g(y^k)-\frac{1}{(2s-1)\beta}\|\nabla g(y^k)\|^2\\
&-\frac{1}{(2s-1)\beta}\|\nabla\hat{\phi}(y^k)-\nabla\hat{\phi}(y^{k-1})\|^2+\delta_0\|y^k-y^{k-1}\|^2\\
&+\frac{(2s-1)\beta}{2}\Big\|\sum_{i=1}^m A_ix_i^k+y^k-b-\frac{1}{(2s-1)\beta}[\nabla g(y^k)+\nabla\hat{\phi}(y^k)-\nabla\hat{\phi}(y^{k-1})]\Big\|^2\\
\geq&\sum_{i=1}^m f_i(x_i^k)+g(y^k)-\frac{1}{(2s-1)\beta}\|\nabla g(y^k)\|^2-\frac{L_{\hat{\phi}}^2}{(2s-1)\beta}\|y^k-y^{k-1}\|^2+\delta_0\|y^k-y^{k-1}\|^2\\
&+\frac{(2s-1)\beta}{2}\Big\|\sum_{i=1}^m A_ix_i^k+y^k-b-\frac{1}{(2s-1)\beta}[\nabla g(y^k)+\nabla\hat{\phi}(y^k)-\nabla\hat{\phi}(y^{k-1})]\Big\|^2\\
=&\sum_{i=1}^m f_i(x_i^k)+\bar{g}(y^k)+\Big(\delta_0-\frac{L_{\hat{\phi}}^2}{(2s-1)\beta}\Big)\|y^k-y^{k-1}\|^2\\
&+\frac{(2s-1)\beta}{2}\Big\|\sum_{i=1}^m A_ix_i^k+y^k-b-\frac{1}{(2s-1)\beta}[\nabla g(y^k)+\nabla\hat{\phi}(y^k)-\nabla\hat{\phi}(y^{k-1})]\Big\|^2\\
\geq&\sum_{i=1}^m f_i(x_i^k)+\bar{g}(y^k)\\
&+\frac{(2s-1)\beta}{2}\Big\|\sum_{i=1}^m A_ix_i^k+y^k-b-\frac{1}{(2s-1)\beta}[\nabla g(y^k)+\nabla\hat{\phi}(y^k)-\nabla\hat{\phi}(y^{k-1})]\Big\|^2\\
=:&Q(x_1^k,\cdots,x_m^k,y^k,y^{k-1}).
\end{aligned}
$$

where the final inequality follows from $\delta_0-(L_{\hat{\phi}}^2)/[(2s-1)\beta]\geq 0$ since $\delta_0=[(m+2)L_{\hat{\phi}}^2]/[(r+s)\beta]$, $(2m+3)s-r\geq m+2$ and $s>\frac{1}{2}$. Noticing that the lower semicontinuity functions f_i are coercive, one has, $\inf\limits_{x_i} f_i(x_i)>-\infty$ $(i=1,,\cdots,m)$. Hence,

$$
\begin{cases}
Q(x_1^k,\cdots,x_m^k,y^k,y^{k-1})\leq\hat{\mathcal{L}}_\beta(\hat{w}^0)=\mathcal{L}_\beta(w^0)<+\infty,\\
\inf\limits_{x_i} f_i(x_i)>-\infty,\ i=1,\cdots,m,\ \inf\limits_{y}\bar{g}(y)>-\infty,\ s>\dfrac{1}{2},
\end{cases}
$$

which implies $\{x_i^k\}$ $(i=1,\cdots,m)$, $\{y^k\}$ and $\{\sum\limits_{i=1}^m A_ix_i^k+y^k-b-\frac{1}{(2s-1)\beta}[\nabla g(y^k)+\nabla\hat{\phi}(y^k)-\nabla\hat{\phi}(y^{k-1})]\}$ are bounded, further, $\{\nabla g(y^k)+\nabla\hat{\phi}(y^k)-\nabla\hat{\phi}(y^{k-1})\}$ is bounded. Together with (2.20), it holds that $\{\lambda^k\}$ is also bounded. Then, $\{w^k\}$ is bounded, and this completes the proof. □

3 Convergence Rate

Theorem 3.1 (Convergence rate of $\hat{\mathcal{L}}_\beta(\hat{w}^k)$)**.** *Suppose that Assumption H holds, and the sequence* $\{w^k\}$ *is bounded. Let* $\hat{\mathcal{L}}_\beta(\hat{w})$ *satisfies the Łojasiewicz property at* $\hat{w}^*$*. Define* $\Gamma_k=\hat{\mathcal{L}}_\beta(\hat{w}^k)-\hat{\mathcal{L}}_\beta(\hat{w}^*)$ *with* $\hat{\mathcal{L}}_\beta(\hat{w}^*)=\lim\limits_{k\to+\infty}\hat{\mathcal{L}}_\beta(\hat{w}^k)$*. There exist* $\theta\in[0,1)$ *and* $\alpha>0$*, such that,* *for any* $\check{k} \geq 2$,

$$\alpha\Gamma_k^{2\theta} \leq \Gamma_{k-2} - \Gamma_k, \qquad \forall\, k \geq \check{k}. \tag{3.1}$$

In addition, the following claims are true:

(i) if $\theta = 0$, *then* $\Gamma_k \to 0$ *in a finite number of iterations;*

(ii) if $\theta \in (0, \frac{1}{2}]$, *there exists* $\tau \in [0,1)$ *such that* $\Gamma_k = O(\tau^k)$ *for all* $k \geq \check{k}$;

(iii) if $\theta \in (\frac{1}{2}, 1)$, *then* $\Gamma_k = O(k^{1/(1-2\theta)})$ *for all* $k \geq \check{k}$.

Proof. From (2.39), we have, for any $k \geq 2$,

$$\frac{1}{3\zeta^2}\|\varepsilon^k\|^2 \leq \|\mathbf{x}_{[1:m]}^k - \mathbf{x}_{[1:m]}^{k-1}\|^2 + \|y^k - y^{k-1}\|^2 + \|y^{k-1} - y^{k-2}\|^2, \tag{3.2}$$

where $\varepsilon^k \in \partial\hat{\mathcal{L}}_\beta(\hat{w}^k)$. By (2.12), there exists $k_0 \geq 2$ such that

$$\|\mathbf{x}_{[1:m]}^k - \mathbf{x}_{[1:m]}^{k-1}\|^2 + \|y^k - y^{k-1}\|^2 + \|y^{k-1} - y^{k-2}\|^2 \leq \frac{1}{\delta}(\Gamma_{k-2} - \Gamma_k), \qquad \forall\, k \geq k_0. \tag{3.3}$$

Together with (3.2), we have

$$\frac{1}{3\zeta^2}\|\varepsilon^k\|^2 \leq \frac{1}{\delta}(\Gamma_{k-2} - \Gamma_k). \tag{3.4}$$

Since $\hat{\mathcal{L}}_\beta(\hat{w})$ satisfies the Łojasiewicz property at $\hat{w}^*$, $\hat{w}^k \to \hat{w}^*$, $\hat{\mathcal{L}}_\beta(\hat{w}^k)$ is monotonically decreasing and $\hat{\mathcal{L}}_\beta(\hat{w}^k) \to \hat{\mathcal{L}}_\beta(\hat{w}^*)$ as $k \to +\infty$, then there exist $\check{k} \geq k_0$, $\epsilon > 0$, $\theta \in [0,1)$ and $c_l > 0$ such that $d(\hat{w}^k, \hat{w}^*) < \epsilon$ and $|\hat{\mathcal{L}}_\beta(\hat{w}^k) - \hat{\mathcal{L}}_\beta(\hat{w}^*)|^\theta \leq c_l \cdot d(0, \partial\hat{\mathcal{L}}_\beta(\hat{w}^k))$ for any $k \geq \check{k}$. It follows that

$$\Gamma_k^{2\theta} \leq c_l^2\|\varepsilon^k\|^2 \;\text{ with }\; \varepsilon^k \in \partial\hat{\mathcal{L}}_\beta(\hat{w}^k), \qquad \forall\, k \geq \check{k}.$$

Combining this with (3.4), one further yields

$$\frac{\delta}{3c_l^2\zeta^2}\Gamma_k^{2\theta} \leq \Gamma_{k-2} - \Gamma_k.$$

Denote $\alpha = \frac{\delta}{3c_l^2\zeta^2}$, we get (3.1).

(i) Case $\theta = 0$. We assume that $\Gamma_k > 0$ for any $k \geq \check{k}$. From (3.1), we have $0 < \alpha \leq \Gamma_{k-2} - \Gamma_k$, which actually contradicts the fact that the right hand side converges to 0 as $k \to +\infty$. Hence, there exists $\tilde{k} \geq \check{k}$ such that $\Gamma_k = 0$ for any $k \geq \tilde{k}$.

(ii) Case $\theta \in (0, \frac{1}{2}]$. Then $2\theta - 1 < 0$. Let $k \geq \check{k} + 1$ be fixed. Since $\{\Gamma_i\}_{i \geq \check{k}}$ is a monotonically decreasing sequence, we have $\Gamma_i \leq \Gamma_{\check{k}}$ for $i = \check{k}+1, \check{k}+2, \cdots, k$. This, together with (3.1), yields

$$\alpha\Gamma_{\check{k}}^{2\theta-1}\Gamma_k \leq \alpha\Gamma_k^{2\theta-1}\Gamma_k \leq \Gamma_{k-2} - \Gamma_k, \qquad \text{i.e., } \Gamma_k \leq \frac{1}{1 + \alpha\Gamma_{\check{k}}^{2\theta-1}}\Gamma_{k-2}.$$

We rearrange this to obtain two situations:

- If $k - \check{k}$ is odd, then

$$\Gamma_k \leq \frac{\Gamma_{k-2}}{1 + \alpha\Gamma_{\check{k}}^{2\theta-1}} \leq \frac{\Gamma_{k-4}}{(1 + \alpha\Gamma_{\check{k}}^{2\theta-1})^2} \leq \cdots \leq \frac{\Gamma_{\check{k}-1}}{(1 + \alpha\Gamma_{\check{k}}^{2\theta-1})^{\frac{k-\check{k}+1}{2}}};$$

• If $k-K$ is even, then

$$\Gamma_k \le \frac{\Gamma_{k-2}}{1+\alpha e_{\check{k}}^{2\theta-1}} \le \frac{\Gamma_{k-4}}{(1+\alpha\Gamma_{\check{k}}^{2\theta-1})^2} \le \cdots \le \frac{\Gamma_{\check{k}}}{(1+\alpha\Gamma_{\check{k}}^{2\theta-1})^{\frac{k-\check{k}}{2}}}.$$

Let $\tau := \left(\frac{1}{1+\alpha\Gamma_{\check{k}}^{2\theta-1}}\right)^{1/2} \in (0,1)$. Then

$$\Gamma_k \le \frac{\max\{\Gamma_j : 0 \le j \le \check{k}\}}{\tau^{\check{k}-k}} = \frac{\max\{\Gamma_j : 0 \le j \le \check{k}\}}{\tau^{\check{k}}}\tau^k = O(\tau^k), \qquad \forall\, k \ge \check{k}.$$

(iii) Case $\theta \in (\frac{1}{2},1)$. Then $\nu := 1-2\theta < 0$. From (3.1), we get

$$\alpha \le (\Gamma_{k-2}-\Gamma_k)\Gamma_k^{-2\theta}, \qquad \forall\, k \ge \check{k}. \tag{3.5}$$

Define $h(t) = t^{-2\theta}$ for $t \in [0,+\infty)$. Clearly, h is monotonically decreasing as $h'(t) = -2\theta t^{-(1+2\theta)} < 0$. This further gives $h(\Gamma_{k-2}) \le h(\Gamma_k)$ for any $k \ge \check{k}$ as Γ_k is monotonically decreasing, and so $h(\Gamma_{k-2}) \le h(t),\ t \in [\Gamma_k, \Gamma_{k-2}]$. We consider two situations as follows.

• Assume that $h(\Gamma_k) < 2h(\Gamma_{k-2})$ for any $k \ge \check{k}$. This, together with (3.5), gets

$$\begin{aligned}\alpha < 2(\Gamma_{k-2}-\Gamma_k)h(\Gamma_{k-2}) =& 2h(\Gamma_{k-2})\int_{\Gamma_k}^{\Gamma_{k-2}} 1dt \le 2\int_{\Gamma_k}^{\Gamma_{k-2}} h(t)dt\\ =& 2\int_{\Gamma_k}^{\Gamma_{k-2}} t^{-2\theta}dt = \frac{2}{1-2\theta}(\Gamma_{k-2}^{1-2\theta}-\Gamma_k^{1-2\theta}),\end{aligned}$$

which implies that

$$0 < \frac{\alpha(2\theta-1)}{2} < \Gamma_k^{1-2\theta} - \Gamma_{k-2}^{1-2\theta}.$$

Denote $\hat{\mu} = \frac{\alpha(2\theta-1)}{2} > 0$, one has

$$0 < \hat{\mu} < \Gamma_k^{\nu} - \Gamma_{k-2}^{\nu}, \qquad \forall\, k \ge \check{k}. \tag{3.6}$$

• Assume that $h(\Gamma_k) \ge 2h(\Gamma_{k-2})$. Then $\Gamma_k^{-2\theta} \ge 2\Gamma_{k-2}^{-2\theta}$. This is equivalent to $\frac{1}{2}\Gamma_{k-2}^{2\theta} \ge \Gamma_k^{2\theta}$, which by raising both sides to the power $\frac{1}{2\theta}$ and letting $q = (1/2)^{\frac{1}{2\theta}}$ leads to $q\Gamma_{k-2} \ge \Gamma_k$. Since $q^\nu\Gamma_{k-2}^\nu \le \Gamma_k^\nu$, we have $(q^\nu-1)\Gamma_{k-2}^\nu \le \Gamma_k^\nu - \Gamma_{k-2}^\nu$. In view of $q^\nu - 1 > 0$ and $\Gamma_p \to 0^+$ as $p \to +\infty$, there exists $\bar{\mu}$ such that $(q^\nu-1)\Gamma_{k-2}^\nu > \bar{\mu}$ for any $k \ge \check{k}$. Thus, we obtain

$$0 < \bar{\mu} \le \Gamma_k^\nu - \Gamma_{k-2}^\nu,\ \forall\ k \ge \check{k}. \tag{3.7}$$

In both situations above, we get

$$0 < \min\{\hat{\mu},\ \bar{\mu}\} =: \mu \le \Gamma_i^\nu - \Gamma_{i-2}^\nu, \qquad \forall\, i \ge \check{k},$$

where $\hat{\mu}$ and $\bar{\mu}$ are defined as in (3.6) and (3.7), respectively. Summing up inequality above for $i = \check{k}+1, \cdots, k+1 (\ge \check{k}+1)$, we get

$$\sum_{i=\check{k}+1}^{k+1}(\Gamma_i^\nu - \Gamma_{i-2}^\nu) = (\Gamma_{k+1}^\nu + \Gamma_k^\nu) - (\Gamma_{\check{k}}^\nu + \Gamma_{\check{k}-1}^\nu) \ge \mu(k-\check{k}+1).$$

This, together with $\Gamma_{k-1} \geq \Gamma_k$ (for any k) and $\nu < 0$, further gives

$$\frac{\mu}{2}(k - \check{k} + 1) \leq \Gamma_{k+1}^{\nu} - \Gamma_{\check{k}-1}^{\nu} \leq \Gamma_{k+1}^{\nu}, \qquad \forall\, k \geq \check{k}.$$

Then,

$$\Gamma_{k+1} \leq \left[\frac{\mu}{2}(k - \check{k} + 1)\right]^{1/\nu} = \left[\frac{\mu}{2}(k - \check{k} + 1)\right]^{1/(1-2\theta)} = O((k+1)^{1/(1-2\theta)}).$$

So, the claim (iii) is proved. □

Lemma 3.2. *Suppose that the Assumption H holds, $\hat{\mathcal{L}}_\beta(\hat{w})$ is a KL function, and w^* is the unique limit point of $\{w^k\}$. Then, there exists $\check{k} \geq 2$ such that*

$$\|w^k - w^*\| = O(\max\{\varphi(\Gamma_k),\ \sqrt{\Gamma_{k-2}}\}), \qquad \forall\, k \geq \check{k}, \tag{3.8}$$

where $\Gamma_k = \hat{\mathcal{L}}_\beta(\hat{w}^k) - \hat{\mathcal{L}}_\beta(\hat{w}^*)$ with $\hat{\mathcal{L}}_\beta(\hat{w}^*) = \lim_{k\to+\infty} \hat{\mathcal{L}}_\beta(\hat{w}^k)$, $\varphi \in \Xi_\eta$ with $\eta > 0$.

Proof. Let $k_0 \geq 1$, such that $\{\Gamma_k\}_{k\geq k_0}$ is a monotonically decreasing sequence, and from (3.3), there exists $\check{k} \geq k_0 + 1$ such that, for any $k \geq \check{k}$,

$$\|\mathbf{x}_{[1:m]}^k - \mathbf{x}_{[1:m]}^{k-1}\| + \|y^k - y^{k-1}\| + \|y^{k-1} - y^{k-2}\| \leq \sqrt{\frac{3}{\delta}(\Gamma_{k-2} - \Gamma_k)} \leq \sqrt{\frac{3}{\delta}\Gamma_{k-2}}. \tag{3.9}$$

On the other hand, for $q \geq k$, one has

$$\begin{aligned}\sum_{i=k}^{q} \|\mathbf{x}_{[1:m]}^{i+1} - \mathbf{x}_{[1:m]}^{i}\| \geq &\sum_{i=k}^{q}(\|\mathbf{x}_{[1:m]}^{i} - \mathbf{x}_{[1:m]}^{*}\| - \|\mathbf{x}_{[1:m]}^{i+1} - \mathbf{x}_{[1:m]}^{*}\|)\\ =&\|\mathbf{x}_{[1:m]}^{k} - \mathbf{x}_{[1:m]}^{*}\| - \|\mathbf{x}_{[1:m]}^{q+1} - \mathbf{x}_{[1:m]}^{*}\|.\end{aligned}$$

Letting $q \to +\infty$ in the relation above, we give

$$\sum_{i=k}^{+\infty} \|\mathbf{x}_{[1:m]}^{i+1} - \mathbf{x}_{[1:m]}^{i}\| \geq \|\mathbf{x}_{[1:m]}^{k} - \mathbf{x}_{[1:m]}^{*}\| - \lim_{q\to+\infty} \|\mathbf{x}_{[1:m]}^{q+1} - \mathbf{x}_{[1:m]}^{*}\| = \|\mathbf{x}_{[1:m]}^{k} - \mathbf{x}_{[1:m]}^{*}\|. \tag{3.10}$$

Similarly,

$$\sum_{i=k}^{+\infty} \|y^{i+1} - y^i\| \geq \|y^k - y^*\|. \tag{3.11}$$

To proceed, from (2.42), (3.10) and (3.11), we have

$$\begin{aligned}\|\mathbf{x}_{[1:m]}^{k} - \mathbf{x}_{[1:m]}^{*}\| + \|y^k - y^*\| \leq &\Big(\|y^k - y^{k-1}\| + \frac{1}{2}\|\mathbf{x}_{[1:m]}^{k} - \mathbf{x}_{[1:m]}^{k-1}\| + \frac{1}{2}\|y^{k-1} - y^{k-2}\|\Big)\\ &+ \frac{4\zeta}{\delta}\varphi(\hat{\mathcal{L}}_\beta(\hat{w}^k) - \hat{\mathcal{L}}_\beta(\hat{w}^*)).\end{aligned}$$

Together with (3.9) and the definition of Γ_k, one has

$$\|\mathbf{x}_{[1:m]}^{k} - \mathbf{x}_{[1:m]}^{*}\| + \|y^k - y^*\| \leq \tfrac{4\zeta}{\delta}\varphi(\Gamma_k) + \sqrt{\tfrac{3}{\delta}\Gamma_{k-2}} = O(\max\{\varphi(\Gamma_k),\ \sqrt{\Gamma_{k-2}}\}). \tag{3.12}$$

Additionally, by relations (1.15) and (2.20), we get

$$\|\lambda^k - \lambda^*\| = \Big\|\nabla g(y^k) - \nabla g(y^*) + \beta(1-s)\Big(\sum_{i=1}^m A_i(x_i^k - x_i^*) + (y^k - y^*)\Big) + [\nabla\hat{\phi}(y^k) - \nabla\hat{\phi}(y^*)] - [\nabla\hat{\phi}(y^{k-1}) - \nabla\hat{\phi}(y^*)]\Big\|,$$

which implies,

$$\begin{aligned}\|\lambda^k - \lambda^*\| =& O(\|\mathbf{x}_{[1:m]}^k - \mathbf{x}_{[1:m]}^*\|) + O(\|y^k - y^*\|) + O(\|y^{k-1} - y^*\|)\\ =& O(\max\{\varphi(\Gamma_k),\ \sqrt{\Gamma_{k-2}}\}).\end{aligned}$$

Together with (3.12), we obtain

$$\|w^k - w^*\| = O(\max\{\varphi(\Gamma_k),\ \sqrt{\Gamma_{k-2}}\}), \qquad \forall\, k \geq \check{k},$$

and relation (3.8) holds true. □

Theorem 3.3 (Convergence rate of sequence)**.** *Suppose that the Assumption H holds, and w^* is the unique limit point of $\{w^k\}$. If $\hat{\mathcal{L}}_\beta(\hat{w})$ satisfies the KŁ property at $\hat{w}^*$, and the associated function*

$$\varphi(t) = t^{1-\theta} : [0,\eta) \to [0,+\infty) \text{ with } \theta \in [0,1),$$

then there exists $\check{k} \geq 2$ such that the following results hold:

(i) if $\theta = 0$, then w^k converges to w^* in a finite number of iterations;
(ii) if $\theta \in (0, \frac{1}{2}]$, there exists $\hat{\tau} \in [0,1)$ such that $\|w^k - w^*\| = O(\hat{\tau}^k)$ for all $k \geq \check{k}$;
(iii) if $\theta \in (\frac{1}{2}, 1)$, then $\|w^k - w^*\| = O(k^{\frac{1-\theta}{1-2\theta}})$ for all $k \geq \check{k}$.

Proof. Since $\hat{\mathcal{L}}_\beta(\hat{w}^k)$ converges to $\hat{\mathcal{L}}_\beta(\hat{w}^*)$, $\lim\limits_{k\to+\infty} \hat{w}^k \to \hat{w}^*$, and $\hat{\mathcal{L}}_\beta(\hat{w})$ satisfies the KŁ property at $\hat{w}^*$, we conclude that there exist $\epsilon > 0$, $\eta > 0$, $\varphi \in \Xi_\eta$ and $\check{k} \geq 2$ such that for any $k \geq \check{k}$, $d(\hat{w}, \hat{w}^*) < \epsilon$, $\hat{\mathcal{L}}_\beta(\hat{w}^*) < \hat{\mathcal{L}}_\beta(\hat{w}^k) < \hat{\mathcal{L}}_\beta(\hat{w}^*) + \eta$, and the KŁ property

$$\varphi'(\Gamma_k) \cdot d(0, \partial\hat{\mathcal{L}}_\beta(\hat{w}^k)) \geq 1 \tag{3.13}$$

holds. Now, let $\theta \in [0,1)$ and $\varphi(t) = t^{1-\theta}$, then $\varphi'(t) = (1-\theta)t^{-\theta}$. This together with (3.13), further yields $\Gamma_k^\theta \leq d(0, \partial\hat{\mathcal{L}}_\beta(\hat{w}^k))$, $\forall\ k \geq \check{k}$. This implies that $\hat{\mathcal{L}}_\beta(\hat{w})$ satisfies the Łojasiewicz property at $\hat{w}^*$ for all $k \geq \check{k}$ with $c_l = 1$.

On the other hand, the associated function $\varphi(t) = t^{1-\theta} : [0,\eta) \to [0,+\infty)$ with $\theta \in [0,1)$, then $\varphi'(t) > 0$. Together with Lemma 3.2 and $\{\Gamma_k\}$ is a deceasing sequence, where $\Gamma_k = \hat{\mathcal{L}}_\beta(\hat{w}^k) - \hat{\mathcal{L}}_\beta(\hat{w}^*)$ with $\hat{\mathcal{L}}_\beta(\hat{w}^*) = \lim\limits_{k\to+\infty} \hat{\mathcal{L}}_\beta(\hat{w}^k)$, the relation (3.8) leads to

$$\|w^k - w^*\| = O(\max\{\varphi(\Gamma_{k-2}),\ \sqrt{\Gamma_{k-2}}\}), \qquad \forall\, k \geq \check{k}.$$

Then, together with the definition of φ, we have

$$\|w^k - w^*\| = O(\max\{\Gamma_{k-2}^{1-\theta},\ \sqrt{\Gamma_{k-2}}\}), \qquad \forall\, k \geq \check{k}. \tag{3.14}$$

Based on the relation (3.14) and the fact that $\hat{\mathcal{L}}_\beta(\hat{w})$ satisfies the Łojasiewicz property at $\hat{w}^*$ for all $k \geq \check{k}$, and with the help of the relevant conclusions of Theorem 3.1, the proof is further completed.

(i) If $\theta = 0$, then $\Gamma_k \to 0$ in a finite numbers of iterations. Then, from (3.14), w^k converges to w^* in a finite numbers of iterations.

(ii) If $\theta \in (0, \frac{1}{2}]$, then by (3.14), we further give $\max\{\Gamma_{k-2}^{1-\theta},\ \sqrt{\Gamma_{k-2}}\} = \sqrt{\Gamma_{k-2}}$. Let $\hat{\tau} := \tau^{\frac{1}{2}}$. By Theorem 3.1 (ii) for any $k \geq \check{k}$ it holds $\|w^k - w^*\| = O(\hat{\tau}^k)$.

(iii) If $\theta \in (\frac{1}{2}, 1)$, then for any $k \geq \check{k}$, from (3.14) and Theorem 3.1 (iii), we get $\|w^k - w^*\| = O(k^{\frac{1-\theta}{1-2\theta}})$ holds. □

4 Numerical Experiments

In this section, we conduct numerical experiments to show the performances of BPSADMM. All experiments are run in MATLAB R2014b on a 64-bit Dell laptop with Intel(R) Xeon(R) E-2186M CPU (2.90 GHz), 128.00 GB RAM and Windows 10 operating system.

4.1 Implementation Details

Testing model. We consider the following three-block robust principal component analysis (PCA) nonconvex optimization model[4, 24] from the application of matrix decomposition,

$$\min_{L,S,T} \|L\|_{\circledast} + \gamma\|S\|_{1/2}^{1/2} + \frac{\omega}{2}\|T - M\|_F^2, \qquad \text{s.t.} \quad L + S = T, \tag{4.1}$$

where $M \in \mathbb{R}^{p\times n}$ is the given observation matrix, $L, S, T \in \mathbb{R}^{p\times n}$ are decision making matrices (variable), low-rank term $\|L\|_{\circledast} := \sum\limits_{i=1}^{\min(p,n)} |\sigma_i(L)|^{1/2}$ with $\sigma_i(L)$ is the singular value of the low-rank matrix L, sparse term $\|S\|_{1/2}^{1/2} := \sum\limits_{i=1}^{p}\sum\limits_{j=1}^{n} |S_{ij}|^{1/2}$, γ is a trade-off parameter between $\|L\|_{\circledast}$ and $\|S\|_{1/2}^{1/2}$, ω is a penalty parameter related to the noise level, and $\|\cdot\|_F$ denotes the Frobenius norm of matrix, i.e., $\|A\|_F = \|(a_{ij})\|_F = \sqrt{\sum\limits_{i=1}^{p}\sum\limits_{j=1}^{n} |a_{ij}|^2}$.

The model (4.1) is a three-block case of the problem (1.2) discussed in this work. $\|\cdot\|_2$ norm and inner product operator in vector space $\mathbb{R}^n$ are extended to Frobenius norm and $\langle A, B\rangle = \sum\limits_{i=1}^{p}\sum\limits_{j=1}^{n} a_{ij}b_{ij}$ in matrix space $\mathbb{R}^{p\times n}$, respectively, and the ALF of model (4.1) is given by

$$\mathcal{L}_\beta(L,S,T,\lambda) = \|L\|_{\circledast} + \gamma\|S\|_{1/2}^{1/2} + \frac{\omega}{2}\|T-M\|_F^2 - \langle\lambda, L+S-T\rangle + \frac{\beta}{2}\|L+S-T\|_F^2.$$

Applying BPSADMM to solve model (4.1), and according to different positions of multiplier $\lambda^{k+\frac{1}{2}}$, the following two iterative schemes are obtained.

BPSADMM-I:

$$\begin{cases} L^{k+1} = \arg\min\limits_{L}\left\{\mathcal{L}_\beta(L,S^k,T^k,\lambda^k) + \frac{\nu_1}{2}\|L-L^k\|_F^2\right\},\\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(L^{k+1}+S^k-T^k),\\ S^{k+1} = \arg\min\limits_{S}\left\{\mathcal{L}_\beta(L^{k+1},S,T^k,\lambda^{k+\frac{1}{2}}) + \frac{\nu_2}{2}\|S-S^k\|_F^2\right\},\\ T^{k+1} = \arg\min\limits_{T}\left\{\mathcal{L}_\beta(L^{k+1},S^{k+1},T,\lambda^{k+\frac{1}{2}}) + \frac{\nu_3}{2}\|T-T^k\|_F^2\right\},\\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(L^{k+1}+S^{k+1}-T^{k+1}). \end{cases} \tag{4.2}$$

BPSADMM-II:

$$\begin{cases} L^{k+1} = \arg\min\limits_{L}\left\{\mathcal{L}_\beta(L,S^k,T^k,\lambda^k) + \frac{\nu_1}{2}\|L-L^k\|_F^2\right\}\\ S^{k+1} = \arg\min\limits_{S}\left\{\mathcal{L}_\beta(L^{k+1},S,T^k,\lambda^k) + \frac{\nu_2}{2}\|S-S^k\|_F^2\right\},\\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(L^{k+1}+S^{k+1}-T^k),\\ T^{k+1} = \arg\min\limits_{T}\left\{\mathcal{L}_\beta(L^{k+1},S^{k+1},T,\lambda^{k+\frac{1}{2}}) + \frac{\nu_3}{2}\|T-T^k\|_F^2\right\},\\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(L^{k+1}+S^{k+1}-T^{k+1}). \end{cases} \tag{4.3}$$

Referring to [27], and noticing that the smooth quadratic characteristic of T-subproblem in (4.2)/(4.3), the iterations above can be further reduced to

BPSADMM-I:

$$\begin{cases} L^{k+1} = \arg\min_L \Big\{ \|L\|_{\circledast} - \langle \lambda^k,\ L\rangle + \frac{\beta}{2}\|L + S^k - T^k\|_F^2 + \frac{\nu_1}{2}\|L - L^k\|_F^2 \Big\} \\ \qquad = \mathcal{H}\Big(\frac{\beta\left(T^k - S^k\right) + \lambda^k + \nu_1 L^k}{\beta + \nu_1}, \frac{1}{\beta + \nu_1} \Big), \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(L^{k+1} + S^k - T^k), \\ S^{k+1} = \arg\min_S \Big\{ \gamma\|S\|_{1/2}^{1/2} - \langle \lambda^{k+\frac{1}{2}},\ S\rangle + \frac{\beta}{2}\|L^{k+1} + S - T^k\|_F^2 + \frac{\nu_2}{2}\|S - S^k\|_F^2 \Big\} \\ \qquad = \mathcal{H}\Big(\frac{\beta\left(T^k - L^{k+1}\right) + \lambda^{k+\frac{1}{2}} + \nu_2 S^k}{\beta + \nu_2}, \frac{\gamma}{\beta + \nu_2} \Big), \\ T^{k+1} = \arg\min_T \Big\{ \frac{\omega}{2}\|T - M\|_F^2 + \langle \lambda^{k+\frac{1}{2}},\ T\rangle + \frac{\beta}{2}\|L^{k+1} + S^{k+1} - T\|_F^2 + \frac{\nu_3}{2}\|T - T^k\|_F^2 \Big\} \\ \qquad = \frac{\beta\left(L^{k+1} + S^{k+1}\right) + \omega M - \lambda^{k+\frac{1}{2}} + \nu_3 T^k}{\omega + \beta + \nu_3}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(L^{k+1} + S^{k+1} - T^{k+1}), \end{cases} \tag{4.4}$$

and **BPSADMM-II:**

$$\begin{cases} L^{k+1} = \arg\min_L \Big\{ \|L\|_{\circledast} - \langle \lambda^k,\ L\rangle + \frac{\beta}{2}\|L + S^k - T^k\|_F^2 + \frac{\nu_1}{2}\|L - L^k\|_F^2 \Big\} \\ \qquad = \mathcal{H}\Big(\frac{\beta\left(T^k - S^k\right) + \lambda^k + \nu_1 L^k}{\beta + \nu_1}, \frac{1}{\beta + \nu_1} \Big), \\ S^{k+1} = \arg\min_S \Big\{ \gamma\|S\|_{1/2}^{1/2} - \langle \lambda^k,\ S\rangle + \frac{\beta}{2}\|L^{k+1} + S - T^k\|_F^2 + \frac{\nu_2}{2}\|S - S^k\|_F^2 \Big\} \\ \qquad = \mathcal{H}\Big(\frac{\beta\left(T^k - L^{k+1}\right) + \lambda^k + \nu_2 S^k}{\beta + \nu_2}, \frac{\gamma}{\beta + \nu_2} \Big), \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(L^{k+1} + S^{k+1} - T^k), \\ T^{k+1} = \arg\min_T \Big\{ \frac{\omega}{2}\|T - M\|_F^2 + \langle \lambda^{k+\frac{1}{2}},\ T\rangle + \frac{\beta}{2}\|L^{k+1} + S^{k+1} - T\|_F^2 + \frac{\nu_3}{2}\|T - T^k\|_F^2 \Big\} \\ \qquad = \frac{\beta\left(L^{k+1} + S^{k+1}\right) + \omega M - \lambda^{k+\frac{1}{2}} + \nu_3 T^k}{\omega + \beta + \nu_3}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(L^{k+1} + S^{k+1} - T^{k+1}), \end{cases} \tag{4.5}$$

where $\mathcal{H}(\cdot, \frac{1}{\beta+\nu})$ is the half shrinkage operator[27].

On the other hand, for comparison, applying BADMM[24] to model (4.1) yields that

$$\begin{cases} L^{k+1} = \arg\min_L \Big\{ \mathcal{L}_\beta(L, S^k, T^k, \lambda^k) + \frac{\rho}{2}\|L - L^k\|_F^2 \Big\}, \\ S^{k+1} = \arg\min_S \Big\{ \mathcal{L}_\beta(L^{k+1}, S, T^k, \lambda^k) + \frac{\rho}{2}\|S - S^k\|_F^2 \Big\}, \\ T^{k+1} = \arg\min_T \Big\{ \mathcal{L}_\beta(L^{k+1}, S^{k+1}, T, \lambda^k) + \frac{\rho}{2}\|T - T^k\|_F^2 \Big\}, \\ \lambda^{k+1} = \lambda^k - \beta(L^{k+1} + S^{k+1} - T^{k+1}). \end{cases}$$

Similar to the processing technique of (4.2)/(4.3), the BADMM iteration above can be further simplified as

$$\begin{cases} L^{k+1} = \mathcal{H}\Big(\dfrac{\beta\left(T^k - S^k\right) + \lambda^k + \rho L^k}{\beta + \rho}, \dfrac{1}{\beta + \rho}\Big), \\ S^{k+1} = \mathcal{H}\Big(\dfrac{\beta\left(T^k - L^{k+1}\right) + \lambda^k + \rho S^k}{\beta + \rho}, \dfrac{\gamma}{\beta + \rho}\Big), \\ T^{k+1} = \dfrac{\beta\left(L^{k+1} + S^{k+1}\right) + \omega M - \lambda^k + \rho T^k}{\omega + \beta + \rho}, \\ \lambda^{k+1} = \lambda^k - \beta(L^{k+1} + S^{k+1} - T^{k+1}). \end{cases} \tag{4.6}$$

Parameters setting. In our experiments, taking the observation matrix $M = L^* + S^* + N$, where L^* and S^* are the original low-rank matrix and sparse matrix that we want to recover by the model (4.1), respectively, and N is a Gaussian noise matrix. In addition, "rank" denotes the rank of the low-rank matrix L, and "spr" represents the sparsity ratio of the sparse matrix S. The MATLAB codes that generate matrix M are shown below.

- $L = \text{randn}(p,\ \text{rank}) * \text{randn}(\text{rank},\ n)$;
- $S = \text{zeros}(p,\ n);\ q = \text{randperm}(p{*}n);\ K = \text{round}(\text{spr} * \text{p} * \text{n});\ S(q(1:K)) = \text{randn}(K,\ 1)$;
- $\varrho = 0.01;\ \ N = \text{randn}(p,\ n) * \varrho$;
- $T = L + S;\ M = T + N$.

Specifically, we set $p = n = 100, 300$ and 500, and for each dimension, four decomposition models are tested respectively, namely,

$$(\text{rank}, \text{spr}) = (25, 0.05),\ (25, 0.1),\ (30, 0.05),\ (30, 0.1).$$

Moreover, we set $\gamma = \frac{0.1}{\sqrt{p}},\ \omega = 2,\ \beta = 5.5$. The parameters of BPSADMM-I and BPSADMM-II are chosen as $r = s = 1,\ \nu_1 = 0.3$ and $\nu_2 = \nu_3 = 15$. For BADMM (4.6), the parameters are the same as[24], that is $\rho = 0.3$. The matrices L, S, T are initialized by zero matrices. We terminate the iteration when the number of iterations Itr > 300 is satisfied.

4.2 Numerical Results

Table 1. Comparison the performance of BPSADMM-I, BPSADMM-II and BADMM

Data		BPSADMM-I		BPSADMM-II		BADMM	
Problem size	(rank, spr)	Tcpu(s)	RelErr	Tcpu(s)	RelErr	Tcpu(s)	RelErr
$p = n = 100$	(25, 0.05)	1.38	1.304370e-02	1.39	2.211021e-02	1.45	4.868507e-02
	(25, 0.1)	1.35	3.514360e-02	1.37	4.047266e-02	1.47	5.815192e-02
	(30, 0.05)	1.36	1.745922e-02	1.40	2.622026e-02	1.46	5.481113e-02
	(30, 0.1)	1.36	3.631817e-02	1.38	4.310971e-02	1.46	6.391727e-02
$p = n = 300$	(25, 0.05)	11.36	1.713243e-02	12.22	2.058489e-02	13.49	3.390832e-02
	(25, 0.1)	11.22	4.396118e-02	12.13	4.421682e-02	13.03	3.732762e-02
	(30, 0.05)	11.37	1.785201e-02	12.20	2.142713e-02	13.13	3.713961e-02
	(30, 0.1)	11.27	4.170349e-02	12.19	4.206969e-02	12.98	4.028103e-02
$p = n = 500$	(25, 0.05)	31.15	2.573066e-02	32.81	2.742151e-02	35.11	3.420691e-02
	(25, 0.1)	31.45	5.110801e-02	32.94	5.018707e-02	34.51	4.101933e-02
	(30, 0.05)	31.23	2.371600e-02	33.44	2.567142e-02	35.70	3.696487e-02
	(30, 0.1)	31.09	4.658065e-02	33.36	4.619092e-02	34.89	4.284460e-02

In this subsection, we present some simulation results for the robust PCA model (4.1). To describe the quality of recovery, we use the relative error as a performance measure, i.e.,

$$\text{RelErr} := \frac{\|(\hat{L}, \hat{S}, \hat{T}) - (L^*, S^*, T^*)\|_F}{\|(L^*, S^*, T^*)\|_F + 1},$$

where $\hat{L}, \hat{S}$ and $\hat{T}$ be a numerical solution of model (4.1).

In the numerical experiment, we use BPSADMM-I (4.4), BPSADMM-II (4.5) and BADMM (4.6) to test the model (4.1), respectively. The main numerical results are reported in Table 1, and denote the computing time (seconds) of CPU as "Tcpu(s)". As can be seen from Table 1, for the "Tcpu(s)", BPSADMM-I and BPSADMM-II are significantly better than BADMM. In terms of relative error, the BPSADMM-I and BPSADMM-II are not inferior to BADMM. In addition, BPSADMM-I is superior to BPSADMM-II. Thus, from the numerical test results and analysis, it can be seen that the numerical effect of BPSADMM is better than BADMM for the given test problems.

5 Concluding Remarks

In this paper, combining the idea of symmetric ADMM (in two-block optimization) and the Bregman distance, we proposed a Bregman-style partially symmetric ADMM (BPSADMM). Under some suitable assumptions, it has shown that any limit point of the iteration sequence generated by BPSADMM is a critical point of the problem (1.2). Based on the potential function satisfies the KL property, we also proved that the iteration sequence generated by BPSADMM is strongly convergent. Moreover, we analyzed the convergence rate results of both the potential function and iteration sequences under Łojasiewicz property.

In the following research, we will attempt to establish a more representative method by adding a multiplier update formula after each subproblem. In addition, we also want to combine inertial technology to accelerate the iteration and make it get better numerical results.

Acknowledgments. The authors wish to thank the Editor-in-Chief and the two anonymous referees for their very professional reviews and quite useful suggestions, which greatly helped us to improve the original version of this paper.

References

[1] Attouch, H., Bolte, J. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Math. Program.,* 116: 5–16 (2009)
[2] Attouch, H., Bolte, J., Redont, P., Soubeyran, A. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality. *Math. Oper. Res.,* 35: 438–457 (2010)
[3] Bolte, J., Sabach, S., Teboulle, M. Proximal alternating linearized minimization or nonconvex and nonsmooth problems. *Math. Program.,* 146(1-2): 459–494 (2014)
[4] Candès E., Li X., Ma, Y., Wright, J. Robust principal component analysis? *J. ACM,* 58(3): 1–37 (2011)
[5] Chen, C. Some notes on the divergence example for multi-block alternating direction method of multipliers (in Chinese). *Oper. Res. Trans.,* 23(3): 135–140 (2019)
[6] Chen, C., He B., Ye, Y., Yuan, X. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. *Math. Program.,* 155(1-2): 57–79 (2016)
[7] Douglas, J., Rachford, H.H. On the numerical solution of heat conduction problems in two or three space variables. *Trans. Am. Math. Soc.,* 82: 421–439 (1956)
[8] Gabay, D. Chapter IX applications of the method of multipliers to variational inequalities. *Stud. Math. Appl.,* 15: 299–331 (1983)
[9] Guo, K., Han, D., Wang, D., Wu, T. Convergence of ADMM for multi-block nonconvex separable optimization models. *Front. Math. China,* 12(5): 1139–1162 (2017)
[10] Han, D., Yuan, X. A note on the alternating direction method of multipliers. *J. Optim. Theory Appl.,* 155(1): 227–238 (2012)
[11] He, B., Tao, M., Yuan, X. Alternating direction method with Gaussian back substitution for separable convex programming. *SIAM J. Optim.,* 22(2): 313–340 (2012)
[12] He, B., Yuan, X. Linearized alternating direction method with Gaussian back substitution for separable convex programming. *Numer. Algebr. Control Optim.,* 3(2): 247–260 (2013)
[13] Hestenes, M. Multiplier and gradient methods. *J. Optim. Theory Appl.,* 4: 303–320 (1969).
[14] Li G., Pong, T. Global convergence of splitting methods for nonconvex composite optimization. *SIAM J. Optim.,* 25: 2434–2460 (2015)

[15] Lions, P.L. Mercier, B. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.,* 16: 964–979 (1979)

[16] Liu, P., Jian, J., Xu, J., Ma, G. A linear approximation Bregman-type Peaceman-Rachford splitting method for nonconvex nonseparable optimization. *Acta. Math. Sin. Chin. Ser.,* 66(01): 75–94 (2023)

[17] Jian, J., Liu, P., Jiang, X. A partially symmetric regularized alternating direction method of multipliers for nonconvex multi-block optimization. *Acta. Math. Sin. Chin. Ser.,* 64(06): 1005–1026 (2021)

[18] Nesterov, Y. Introduction Lectures on Convex Optimization: a Basic Course. Springer Science & Business Media, New York, 2004

[19] Peaceman, D.W., Rachford, Jr.H.H. The numerical solution of parabolic and elliptic differential equations. *J. Soc. Indust. Appl. Math.*, 3: 28–41 (1955)

[20] Powell, M. A Method for Nonlinear Constraints in Minimization Problems. In: Fletcher, R. (ed.) Optimization, pp. 283–298. Academic Press, London, 1969

[21] Rockafellar, R. T., Wets, R. Variational Analysis. Springer Science & Business Media, Berlin, 2009

[22] Scheinberg, K., Ma, S., Goldfarb, D. Sparse inverse covariance selection via alternating linearization methods. *Adv. Neural Inform. Proc. Sys.,* 2101–2109 (2010)

[23] Wang, H., Banerjee, A. Bregman alternating direction method of multipliers. *Proc. Adv. Neural Inform. Proc. Sys.,* 2816–2824 (2014)

[24] Wang, F., Cao, W., Xu, Z. Convergence of multi-block Bregman ADMM for nonconvex composite problems. *Sci. China Inform. Sci.,* 61(12): 122101 (2018)

[25] Wang, F., Xu, Z., Xu, H. Convergence of Bregman alternating direction method with multipliers for nonconvex composite problems. arXiv:1410.8625 (2014)

[26] Xu, J., Chao, M. An inertial Bregman generalized alternating direction method of multipliers for nonconvex optimization. *J. Appl. Math. Comput.,* 68: 1– 27 (2022)

[27] Xu, Z., Chang, X., Xu, F., Zhang, H. $L_{1/2}$ regularization: A thresholding representation theory and a fast solver. *IEEE T. Neur. Net. Lear.,* 23(7): 1013–1027 (2012)

[28] Yang, L., Pong, T.K., Chen, X. Alternating direction method of multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction. *SIAM J. Imaging Sci.,* 10(1): 74–110 (2017)

[29] Zeng, J., Fang,J., Xu, Z. Sparse SAR imaging based on $L_{1/2}$ regularization. *Sci. China F.,* 55: 1755-1775 (2012)

[30] Zeng, J., Xu, Z., Zhang, B., Hong, W., Wu, Y. Accelerated $L_{1/2}$ regularization based SAR imaging via BCR and reduced Newton skills. *Signal Process.*, 93: 1831–1844 (2013)

[31] Zhang, C., Song, Y., Cai, X., Han, D. An extended proximal ADMM algorithm for three-block nonconvex optimization problems. *J. Comput. Appl. Math.,* 398: 113681 (2021)

[32] Zhang, C., Yang, Y., Wang, Z., Chen, Y. A linearized alternating direction method of multipliers for a special three-block nonconvex optimization problem of background/foreground extraction. *IEEE Access*, 8: 198886–198899 (2020)

[33] Zhu, M, Mihic, K, Ye, Y. On a randomized multi-block ADMM for solving selected machine learning problems. arXiv:1907.01995 (2019)