A proximal-perturbed Bregman ADMM for solving nonsmooth and nonconvex optimization problems *

Jianchao Bai †

Abstract. In this paper, we focus on a linearly constrained composite minimization problem whose objective function is possibly nonsmooth and nonconvex. Unlike the traditional construction of augmented Lagrangian function, we provide a proximal-perturbed augmented Lagrangian and then develop a new Bregman Alternating Direction Method of Multipliers (ADMM). Under mild assumptions, we show that the novel augmented Lagrangian residual can be bounded by the primal residuals plus a summable sequence. We further demonstrate that the augmented Lagrangian sequence converges to the limitation of objective sequence, and the iterative sequence converges to a stationary point of the problem. The sublinear convergence rate of the primal residuals are also established.

Keywords: nonconvex optimization, ADMM, Bregman distance, convergence complexity

Mathematics Subject Classification(2010): 65K10; 65Y20; 90C26

1 Introduction

The problem we are interested in this paper is the following potentially nonsmooth and nonconvex minimization problem

$$\min_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} F(x,y) := f_1(x) + f_2(x) + g_1(y) + g_2(y) \qquad \text{s.t.} \;\; Ax + By = b, \tag{1.1}$$

where $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $g_1 : \mathbb{R}^m \to \mathbb{R}$ are continuously differentiable functions (possibly nonconvex) with L_f-Lipschitz gradient and L_g-Lipschitz gradient respectively, $f_2(x)$ and $g_2(y)$ are proper lower semicontinuous functions (possibly nonsmooth), $A \in \mathbb{R}^{p\times n}$, $B \in \mathbb{R}^{p\times m}$ and $b \in \mathbb{R}^p$ are given. Hereafter, the symbols $\mathbb{R}, \mathbb{R}^n$ and $\mathbb{R}^{p\times n}$ denote the sets of real numbers, n dimensional real column vectors, and $m \times n$ real matrices, respectively, and the symbol $\nabla f(x)$ denotes the gradient of differentiable function f at x. We use $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ to denote the standard Euclidean norm and inner product, respectively. Throughout this paper, the solution set of the problem (1.1) is assumed to be nonempty.

A classical yet vital method for solving linearly constrained constrained problem in the form of (1.1) is the Augmented Lagrangian Method (ALM) proposed by Hestenes [11] and Powell

*This work was partially supported by the National Natural Science Foundation of China (12471298) and the Shaanxi Fundamental Science Research Project for Mathematics and Physics (23JSQ031).

†School of Mathematics and Statistics & MOE Key Laboratory for Complexity Science in Aerospace, Northwestern Polytechnical University, Xi'an 710129, China (`jianchaobai@nwpu.edu.cn`).

[19], and it recursively takes the following iterations:

$$\begin{cases} (x_{k+1}, y_{k+1}) = \arg \min\limits_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} \mathcal{L}_\beta(x, y, \lambda_k), \\ \lambda_{k+1} = \lambda_k - \beta\big(Ax_{k+1} + By_{k+1} - b\big), \end{cases}$$

where

$$\mathcal{L}_\beta(x, y, \lambda) = \underbrace{F(x, y) + \big\langle \lambda, Ax + By - b \big\rangle}_{L(x,y,\lambda)} + \frac{\beta}{2} \big\| Ax + By - b \big\|^2 \tag{1.2}$$

denotes the standard augmented Lagrangian function of (1.1), λ denotes the Lagrange multiplier, and $\beta > 0$ is the penalty parameter for the violation of the equality constraints.

As a first-order method, ALM has attracted increasing attention due to its applications in signal/image processing, stochastic learning, machine learning and so forth. Most of existing ALM-type methods were developed based on the classical augmented Lagrangian function, such as exact/inexact accelerated ALM [9, 12, 13, 18, 23] and stochastic ALM [1, 17] for solving equality constrained convex optimization problems, proximal ALM [16, 25] for solving linearly constrained nonconvex optimization problems, and splitting versions of ALM [8, 10] for solving multi-block separable structured minimization problems. Unlike using the traditional augmented Lagrangian function as in (1.2), a double-proximal ALM [3] were recently developed with convergence guaranteed and had been demonstrated to be efficient for solving some machine learning problems. Related to [3], a balanced ALM and a penalty dual-primal ALM [20] were developed for solving the optimization problems in the form of (1.1) and its multi-block extensions. More recently, by introducing an auxiliary variable for (1.1), a new ALM was proposed by Kim [15] based on a proximal-perturbed augmented Lagrangian function, and this method was also extended to solve the nonconvex optimization problem with nonlinear equality constraints [14]. One effective approach to establish the global convergence and sublinear convergence rate of ALM for convex minimization problems is to use variational analysis to characterize both the saddle-point of as well as the iterative sequence, see e.g. [1, 3, 10, 20]. However, for the optimization problems whose objective function is possibly nonconvex and nonsmooth, a practical yet useful technique is to construct a potential function related to the associated Lagrange function and then show the convergence by showing the monotonic decreasing property of potential function, we refer to [9, 13, 14, 25] for more details.

When the objective function of optimization problems has separable structures and many variables, such as (1.1), ALM does not make full use of these structures and hence could not take advantage of the special properties of each component objective function. Consequently, solving the involved subproblem of ALM becomes very difficult. An effective and practical approach to overcome such difficulty is the Alternating Direction Method of Multipliers (ADMM) which can be regarded as a splitting version of ALM. Although He, et al. [10] pointed out that the multi-block ADMM can be rewritten as a Jacobian decomposition of ALM, both of these two methods are based on the standard augmented Lagrangian function. A natrual question is: can we construct a different augmented Lagrangian function to develop a new ADMM for solving the nonconvex and nonsmooth minimization problem (1.1)?

In this paper, motivated by the above question, we will propose a new ADMM-type method based on the similar way of constructing augmented Lagrangian as in [14]. We further establish the convergence of the proposed method in terms of the corresponding augmented Lagrangian sequence as well as the iterative residual with respect to primal variable and constraint violation. More details on the features of our method are summarized in the forthcoming section.

2 Development of 2P-ADMM

Inspired by the new Lagrangian-based first-order method [14, 15], by introducing a similar perturbation variable $z \in \mathbb{R}^p$ such that $z = 0$, we reformulate the problem (1.1) into the following equivalent double-constrained problem

$$\min_{x\in\mathbb{R}^n, y\in\mathbb{R}^m, z\in\mathbb{R}^p} F(x,y) \qquad \text{s.t.} \;\; Ax + By - b = z, \;\; z = 0. \tag{2.1}$$

Define the Proximal-Perturbed (2P) augmented Lagrangian of (2.1) as

$$\mathcal{L}_\beta(x,y,\lambda,z,\mu) = F(x,y) + \langle \lambda, Ax + By - b - z\rangle + \langle \mu, z\rangle + \frac{\alpha}{2}\|z\|^2 - \frac{\sigma}{2}\|\lambda - \mu\|^2, \tag{2.2}$$

where $\lambda, \mu \in \mathbb{R}^p$ are the Lagrange multipliers associated with the equality constraints, $\alpha > 0$ is a penalty parameter and $\sigma > 0$ denotes a proximal parameter.

To predigest discussion, we simply denote $\mathcal{L}_\beta(x,y,\lambda,z,\mu)$ by $\mathcal{L}_\beta(w)$ where $w = (x,y,\lambda,z,\mu)$. Special comments are given regarding this new 2P augmented Lagrangian function:

(i) Unlike the standard augmented Lagrangian (1.2), we exploit a proximal term $\frac{\sigma}{2}\|\lambda-\mu\|^2$ in (2.2), in stead of the widely-used quadratic penalty for the constraint $Ax+By-b = z$, to ensure the strongly concavity of $\mathcal{L}_\beta(w)$ w.r.t. the Lagrange multipliers λ (for fixed μ) and μ (for fixed λ). This technique is helpful for simplifying the update of Lagrange multipliers. Besides, minimizing the new function $\mathcal{L}_\beta(w)$ w.r.t. each primal variable can enjoy the proximity operator of $f_2(x)$ or $g_2(y)$, when each subproblem exploits a customized proximal term including the general Bregman distance;

(ii) Because $\mathcal{L}_\beta(w)$ is smooth and strongly convex about z, there exists a unique solution for given (λ, μ). More specifically, by minimizing $\mathcal{L}_\beta(w)$ w.r.t. variable z, we can derive

$$z(\lambda,\mu) = \frac{\lambda - \mu}{\alpha}, \tag{2.3}$$

which implies $\lambda = \mu$ at the unique solution $z^* = 0$. Based on the relationship in (2.3), we thus add the smoothing proximal term $-\frac{\beta}{2}\|\lambda-\mu\|^2$ to the Lagrangian in (2.2).

Now, plugging the certain relationship (2.3) into (2.2) results in

$$\mathcal{L}_\beta(w) = L(x,y,\lambda) - \frac{1}{2\beta}\|\lambda - \mu\|^2 \tag{2.4}$$

with $\beta = \frac{\alpha}{1+\alpha\sigma}$. Clearly, the function $L_\beta(w)$ is strongly concave about λ for given (x,y,μ), so there exists a unique maximizer, denoted by $\lambda(x,y,\mu)$, namely,

$$\lambda(x,y,\mu) = \arg\max_{\lambda\in\mathbb{R}^p} L_\beta(w) = \mu + \beta(Ax + By - b).$$

Note that directly minimizing $L_\beta(w)$ about the primal variables x and y is still challenging since it does not make the full advantages of each nonsmooth objective function in (1.1) as well as the sparable structure of the problem. To tackle these obstacles, we first employ an approximation to $\mathcal{L}_\beta(w)$ as the following

$$\begin{aligned}\tilde{\mathcal{L}}_\beta(w, v_1, v_2) := \; & f_2(x) + g_2(y) + \mathcal{B}_{\phi_1}(x, v_1) + \mathcal{B}_{\phi_2}(y, v_2) \\ & + \bar{\mathcal{L}}_\beta(w) + \langle \nabla_x \bar{\mathcal{L}}_\beta(w), x - v_1\rangle + \langle \nabla_y \bar{\mathcal{L}}_\beta(w), y - v_2\rangle,\end{aligned}$$

where $\bar{\mathcal{L}}_\beta(w)$ is the smooth part of $\mathcal{L}_\beta(w)$:

$$\bar{\mathcal{L}}_\beta(w) = f_1(x) + g_1(y) + \langle \lambda, Ax + By - b - z \rangle + \langle \mu, z \rangle + \frac{\alpha}{2}\|z\|^2 - \frac{\sigma}{2}\|\lambda - \mu\|^2,$$

and $\mathcal{B}_{\phi_i}$ represents the Bregman distance [5, 24] defined as

$$\mathcal{B}_{\phi_i}(x, v) := \phi_i(x) - \phi_i(v) - \langle \nabla \phi_i(v), x - v \rangle, \quad i = 1, 2,$$

for any $x, v \in \mathbb{R}^n$. Here $\phi_i(\cdot)$ is a continuously differentiable function with L_{ϕ_i}-Lipschitz gradient and satisfies

$$\mathcal{B}_{\phi_i}(x, v) \geq \frac{\theta_i}{2}\|x - v\|^2, \quad i = 1, 2.$$

This type of proxima term (i.e., Bregman distance) is exploited to simplify the subproblems when it is not easy to solve or does not admit a closed-form solution. A particular choice of ϕ_i is $\phi_i(\cdot) = \frac{\theta_i}{2}\|\cdot\|^2$, which makes the Bregman distance becomes $\mathcal{B}_{\phi_i}(x, v) = \frac{\theta_i}{2}\|x - v\|^2$ and hence the modulus θ_i can be regarded as proximal parameter.

Input: $\alpha \gg 1, \sigma \in (0,1), \beta = \frac{\alpha}{1+\alpha\sigma}, r \in (0.9, 1), \theta_1 > L_f$ and $\theta_2 > L_g$.
Initialization: $w_0 = (x_0, y_0, z_0, \lambda_0, \mu_0)$ and $\delta_0 \in (0, 1]$.
For $k = 0, 1, \cdots$

1. $x_{k+1} = \arg\min_{x \in \mathbb{R}^n} \left\{ f_2(x) + \langle x - x_k, \nabla_x \bar{\mathcal{L}}_\beta(x_k, y_k, \lambda_k, z_k, \mu_k) \rangle + \mathcal{B}_{\phi_1}(x, x_k) \right\}$;
2. $y_{k+1} = \arg\min_{y \in \mathbb{R}^m} \left\{ g_2(y) + \langle y - y_k, \nabla_y \bar{\mathcal{L}}_\beta(x_{k+1}, y_k, \lambda_k, z_k, \mu_k) \rangle + \mathcal{B}_{\phi_2}(y, y_k) \right\}$;
3. $\mu_{k+1} = \mu_k + \tau_k(\lambda_k - \mu_k)$ with $\tau_k = \frac{\delta_k}{1+\|\lambda_k - \mu_k\|^2}$;
4. $\lambda_{k+1} = \mu_{k+1} + \beta(Ax_{k+1} + By_{k+1} - b)$;
5. $z_{k+1} = \frac{\lambda_{k+1} - \mu_{k+1}}{\alpha}$;
6. $\delta_{k+1} = r\delta_k$;

End
Output (x_{k+1}, y_{k+1}).

Algorithm 2.1: A Proximal-Perturbed Lagrangian Method (2P-ADMM)

Based on the above preliminaries and the splitting solving idea w.r.t. primal variables x and y, we propose the following 2P-based Alternating Direction Method of Multipliers (2P-ADMM) whose framework is described in Algorithm 2.1. In fact, both x-subproblem and y-subproblem can be simplified as

$$\begin{cases} x_{k+1} = \arg\min_{x \in \mathbb{R}^n} \left\{ f_2(x) + \langle x - x_k, \nabla f_1(x_k) + A^\top \lambda_k \rangle + \mathcal{B}_{\phi_1}(x, x_k) \right\}, \\ y_{k+1} = \arg\min_{y \in \mathbb{R}^m} \left\{ g_2(y) + \langle y - y_k, \nabla g_1(y_k) + B^\top \lambda_k \rangle + \mathcal{B}_{\phi_2}(y, y_k) \right\}. \end{cases} \tag{2.5}$$

Hence, the fifth step in Algorithm 2.1 does on work and can be neglected when carrying out experiments. The updating formula of δ_{k+1} implies $\delta_k = r^k\delta_0$. By the region $r \in (0, 1)$ and $\delta_0 \in (0, 1]$, we know the sequence $\{\delta_k\}$ is summable. Due to this fact, the choice of τ_k can guarantee the boundedness of $\{\mu_k\}$, which in turn guarantees the boundedness of $\{\lambda_k\}$.

3 Convergence analysis

3.1 Technical preliminarily

In this subsection, we prepare several lemmas that will be used in the convergence analysis of the augmented Lagrangian sequence as in (2.2)(equivalently (2.4)) and the iterative sequence. Throughout this paper, similar to the assumptions in [2], we make the following assumptions:

(A1) $\bar{f}_1 = \inf_x \left\{ f_1(x) - \frac{1}{2L_f} \left\| \nabla f_1(x) \right\|^2 \right\} > -\infty$ and $\bar{g}_1 = \inf_y \left\{ g_1(y) - \frac{1}{2L_g} \left\| \nabla g_1(y) \right\|^2 \right\} > -\infty$;

(A2) $\lim_{\|x\| \to \infty} \inf f_2(x) = +\infty$ and $\lim_{\|y\| \to \infty} \inf g_2(y) = +\infty$.

Lemma 3.1 *Let $\{\mu_k\}$, $\{z_k\}$ and $\{\lambda_k\}$ be the sequences generated by Algorithm 2.1. Then, these three sequences are bounded.*

Proof. By the update of μ_{k+1}, we have

$$\left\| \mu_{k+1} \right\| = \left\| \mu_0 + \sum_{i=0}^{k} \tau_i \big(\lambda_i - \mu_i \big) \right\| \leq \|u_0\| + \sum_{i=0}^{+\infty} \frac{\delta_i}{\|\lambda_i - \mu_i\|^2 + 1} \|\lambda_i - \mu_i\| \leq \|u_0\| + \frac{1}{2} \sum_{i=0}^{\infty} \delta_i < +\infty,$$

which shows that the sequence $\{\mu_k\}$ is bounded. Combine the update of μ_{k+1} and z_{k+1}, we have $z_k = \frac{1}{\alpha \tau_k}(\mu_{k+1} - \mu_k)$, which, by the boundedness of μ_{k+1}, shows that $\{z_k\}$ is also bounded.

Besides, the update of μ_{k+1} gives

$$\lambda_k = \frac{1}{\tau_k} \mu_{k+1} + \Big(1 - \frac{1}{\tau_k} \Big) \mu_k,$$

which means λ_k is a combination of μ_{k+1} and μ_k. Combine this relationship with the boundedness of $\{\mu_k\}$ to ensure that $\{\lambda_k\}$ is a bounded sequence. ■

The above lemma as well as the following lemma will be used to investigate some properties of the iterative sequence $\{w\}$ generated by Algorithm 2.1.

Lemma 3.2 *Let $\{\mu_k\}$ and $\{\lambda_k\}$ be the sequences generated by Algorithm 2.1. Then, we have*

$$\left\| \mu_{k+1} - \mu_k \right\|^2 \leq \frac{\delta_k^2}{4}, \qquad \left\| \lambda_k - \mu_k \right\|^2 \leq \frac{\delta_k}{\tau_k}, \tag{3.1}$$

and

$$\left\| \lambda_{k+1} - \lambda_k \right\|^2 \leq 2 \left\| \lambda_{k+1} - \mu_{k+1} \right\|^2 + 2 \left\| \mu_{k+1} - \lambda_k \right\|^2. \tag{3.2}$$

Proof. By the way of updating μ_{k+1} and λ_{k+1}, we have

$$\left\| \mu_{k+1} - \mu_k \right\|^2 = \tau_k^2 \left\| \lambda_k - \mu_k \right\|^2 \leq \frac{\delta_k^2}{\|\lambda_k - \mu_k\|^2 + 2 + \frac{1}{\|\lambda_k - \mu_k\|^2}} \leq \frac{\delta_k^2}{4},$$

where the first inequality uses the definition of τ_k and the last inequality uses the fact that $a + b \geq 2\sqrt{ab}$ for any $a, b \geq 0$. Using the the definition of τ_k again, it holds that

$$\tau_k \left\| \lambda_k - \mu_k \right\|^2 = \frac{\delta_k}{1 + \frac{1}{\|\lambda_k - \mu_k\|^2}} \leq \delta_k.$$

The result in (3.2) follows directly by the the fact $(a+b)^2 \leq 2a^2 + 2b^2$ for any a and b. ■

By the update of τ_k, we know $\tau_k \in (0,1)$ which is a bounded sequence. Based on its lower bound, next we provide some core properties related to the sequence $\{\mathcal{L}_\beta(w_k)\}$, which further establishes that both the iterative residual and the constraint residual converge to zero.

Theorem 3.1 *Let $\bar{\tau} > 0$ be the lower bound of $\{\tau_k\}$ and $\{w_k := (x_k, y_k, \lambda_k, z_k, \mu_k)\}$ be the sequence generated by Algorithm 2.1. Then, the following hold:*

(i) The sequence $\{\mathcal{L}_\beta(w_{k+1})\}$ defined in (2.4) satisfies

$$\mathcal{L}_\beta(w_{k+1}) \leq \mathcal{L}_\beta(w_k) - \frac{\theta_1 - L_f}{2}\big\|x_{k+1} - x_k\big\|^2 - \frac{\theta_2 - L_g}{2}\big\|y_{k+1} - y_k\big\|^2 + \frac{\delta_{k+1} + \delta_k}{\beta\bar{\tau}}; \tag{3.3}$$

(ii) Under the assumptions (A1)-(A2), the sequence $\{\mathcal{L}_\beta(w_k)\}$ is convergent. Moreover,

$$\lim_{k\to\infty} \big\|w_{k+1} - w_k\big\| = 0 \quad \text{and} \quad \lim_{k\to\infty} \big\|Ax_{k+1} + By_{k+1} - b\big\| = 0. \tag{3.4}$$

Proof. To prove the assertion (i), we split $\mathcal{L}_\beta(w_{k+1}) - \mathcal{L}_\beta(w_k)$ into three residuals:

$$\begin{aligned} \mathcal{L}_\beta(w_{k+1}) - \mathcal{L}_\beta(w_k) =& \mathcal{L}_\beta(x_{k+1}, y_k, z_k, \lambda_k, \mu_k) - \mathcal{L}_\beta(x_k, y_k, z_k, \lambda_k, \mu_k) && (3.5)\\ &+ \mathcal{L}_\beta(x_{k+1}, y_{k+1}, z_k, \lambda_k, \mu_k) - \mathcal{L}_\beta(x_{k+1}, y_k, z_k, \lambda_k, \mu_k) && (3.6)\\ &+ \mathcal{L}_\beta(x_{k+1}, y_{k+1}, z_{k+1}, \lambda_{k+1}, \mu_{k+1}) - \mathcal{L}_\beta(x_{k+1}, y_{k+1}, z_k, \lambda_k, \mu_k). && (3.7) \end{aligned}$$

According to the equivalent expression of x_{k+1}-subproblem as in (2.5), we have

$$f_2(x_{k+1}) + \big\langle \nabla f_1(x_k) + A^\top \lambda_k, x_{k+1} - x_k \big\rangle + \mathcal{B}_{\phi_1}(x_{k+1}, x_k) \leq f_2(x_k),$$

implying that

$$f_2(x_{k+1}) - f_2(x_k) + \big\langle A^\top \lambda_k, x_k - x_{k+1} \big\rangle \leq \big\langle \nabla f_1(x_k), x_k - x_{k+1} \big\rangle - \frac{\theta_1}{2}\|x_{k+1} - x_k\|^2,$$

which, by using the Lipchitz continuity of f_1:

$$f_1(x_{k+1}) - f_1(x_k) \leq \big\langle \nabla f_1(x_k), x_{k+1} - x_k \big\rangle + \frac{L_f}{2}\big\|x_{k+1} - x_k\big\|^2$$

gives

$$\begin{aligned} \mathcal{L}_\beta(x_{k+1}, y_k, z_k, \lambda_k, \mu_k) - \mathcal{L}_\beta(x_k, y_k, z_k, \lambda_k, \mu_k) &= f_1(x_{k+1}) - f_1(x_k)\\ + f_2(x_{k+1}) - f_2(x_k) + \big\langle A^\top \lambda_k, x_{k+1} - x_k \big\rangle &\leq -\frac{\theta_1 - L_f}{2}\big\|x_{k+1} - x_k\big\|^2. \end{aligned} \tag{3.8}$$

Similarly, the Lipchitz continuity of g_1 yields

$$g_1(y_{k+1}) - g_1(y_k) \leq \big\langle \nabla g_1(y_k), y_{k+1} - y_k \big\rangle + \frac{L_g}{2}\big\|y_{k+1} - y_k\big\|^2,$$

which, by using the following property from the y_{k+1}-subproblem:

$$g_2(y_{k+1}) - g_2(x_k) + \big\langle B^\top \lambda_k, y_{k+1} - y_k \big\rangle \leq \big\langle \nabla g_1(y_k), y_k - y_{k+1} \big\rangle - \frac{\theta_2}{2}\big\|y_{k+1} - y_k\big\|^2,$$

gives

$$\mathcal{L}_\beta(x_{k+1}, y_{k+1}, z_k, \lambda_k, \mu_k) - \mathcal{L}_\beta(x_{k+1}, y_k, z_k, \lambda_k, \mu_k) \le -\frac{\theta_2 - L_f}{2}\big\|y_{k+1} - y_k\big\|^2. \tag{3.9}$$

Notice that

$$\begin{aligned} &\mathcal{L}_\beta(x_{k+1}, y_{k+1}, z_{k+1}, \lambda_{k+1}, \mu_{k+1}) - \mathcal{L}_\beta(x_{k+1}, y_{k+1}, z_k, \lambda_k, \mu_k) \\ =&\big\langle \lambda_{k+1} - \lambda_k, Ax_{k+1} + By_{k+1} - b\big\rangle - \frac{1}{2\beta}\big\|\lambda_{k+1} - \mu_{k+1}\big\|^2 + \frac{1}{2\beta}\big\|\lambda_k - \mu_k\big\|^2 \end{aligned} \tag{3.10}$$

and

$$\big\|\lambda_k - \mu_{k+1}\big\|^2 = \big\|\lambda_k - \mu_k + \mu_k - \mu_{k+1}\big\|^2 = (1-\tau_k)^2\big\|\lambda_k - \mu_k\big\|^2 \le \big\|\lambda_k - \mu_k\big\|^2. \tag{3.11}$$

By using $\lambda_{k+1} - \mu_{k+1} = \beta(Ax_{k+1} + By_{k+1} - b)$, $z_k = \frac{1}{\alpha}(\lambda_k - \mu_k)$, and applying the identity

$$\big\langle a - b, a\big\rangle = \frac{1}{2}\|a-b\|^2 + \frac{1}{2}\|a\|^2 - \frac{1}{2}\|b\|^2$$

with $(a, b) = (\lambda_{k+1} - \mu_{k+1}, \lambda_k - \mu_{k+1})$ to (3.10), we have

$$\begin{aligned} &\mathcal{L}_\beta(x_{k+1}, y_{k+1}, z_{k+1}, \lambda_{k+1}, \mu_{k+1}) - \mathcal{L}_\beta(x_{k+1}, y_{k+1}, z_k, \lambda_k, \mu_k) \\ =&\frac{1}{2\beta}\big\|\lambda_{k+1} - \lambda_k\big\|^2 - \frac{1}{2\beta}\big\|\mu_{k+1} - \lambda_k\big\|^2 + \frac{1}{2\beta}\big\|\lambda_k - \mu_k\big\|^2 \\ \overset{(3.2)}{\le}&\frac{1}{\beta}\big\|\lambda_{k+1} - \mu_{k+1}\big\|^2 + \frac{1}{2\beta}\big\|\mu_{k+1} - \lambda_k\big\|^2 + \frac{1}{2\beta}\big\|\lambda_k - \mu_k\big\|^2 \overset{(3.1),(3.11)}{\le} \frac{\delta_{k+1} + \delta_k}{\beta\bar{\tau}}. \end{aligned} \tag{3.12}$$

So, combining the above inequalities (3.8),(3.9) and (3.12) yields the desired result (3.3).

To prove the result (ii), we first show that $\{w^k\}$ is bounded. It follows from (3.3) together with the conditions $\theta_1 > L_f$ and $\theta_2 > L_g$ that

$$\begin{aligned} &\mathcal{L}_\beta(w_0) + \frac{\delta_0}{\beta\bar{\tau}}\sum_{j=0}^{k}\big(r^{j+1} + r^j\big) \ge \mathcal{L}_\beta(w_k) + \frac{\delta_{k+1} + \delta_k}{\beta\bar{\tau}} \ge \mathcal{L}_\beta(w_{k+1}) \\ =&F(x_{k+1}, y_{k+1}) + \big\langle\lambda_{k+1}, Ax_{k+1} + By_{k+1} - b\big\rangle - \frac{1}{2\beta}\big\|\lambda_{k+1} - \mu_{k+1}\big\|^2 \\ =&F(x_{k+1}, y_{k+1}) + \frac{1}{\beta}\big\langle\lambda_{k+1}, \lambda_{k+1} - \mu_{k+1}\big\rangle - \frac{1}{2\beta}\big\|\lambda_{k+1} - \mu_{k+1}\big\|^2 \\ =&F(x_{k+1}, y_{k+1}) + \frac{1}{2\beta}\big\|\lambda_{k+1}\big\|^2 - \frac{1}{2\beta}\big\|\mu_{k+1}\big\|^2 \\ \ge&\Big(f_1(x_{k+1}) - \frac{1}{2L_f}\big\|\nabla f_1(x_{k+1})\big\|^2\Big) + \Big(g_1(y_{k+1}) - \frac{1}{2L_g}\big\|\nabla g_1(y_{k+1})\big\|^2\Big) \\ &+ f_2(x_{k+1}) + g_2(y_{k+1}) + \frac{1}{2\beta}\big\|\lambda_{k+1}\big\|^2 - \frac{1}{2\beta}\big\|\mu_{k+1}\big\|^2 \\ \ge&\bar{f}_1 + \bar{g}_1 + f_2(x_{k+1}) + g_2(y_{k+1}) + \frac{1}{2\beta}\big\|\lambda_{k+1}\big\|^2 - \frac{1}{2\beta}\big\|\mu_{k+1}\big\|^2, \end{aligned}$$

where the last inequality uses (A1). Then, combining the above relationship with (A2), Lemma 3.1 as well as $r < 1$, we conclude that both $\{x_k\}$ and $\{y_k\}$ are bounded. Consequently, the whole sequence $\{w_k\}$ is bounded. Because $\{w_k\}$ is bounded, the sequence $\{\mathcal{L}_\beta(w_k)\}$ is also bounded from below and there exists at least one limit point. Without loss of generality, let $\{w_{k_j}\}$ be a subsequence of $\{w_k\}$ and w^* be its limit point. Then, the lower semicontinuity of $\{\mathcal{L}_\beta(w_k)\}$ implies $\mathcal{L}_\beta(w^*) \leq \lim_{j\to\infty} \inf \mathcal{L}_\beta(w_{k_j})$ So, $\{\mathcal{L}_\beta(w_{k_j})\}$ is bounded from below and hence is convergent.

Let $\underline{\mathcal{L}}_\beta$ be the lower bound of $\{\mathcal{L}_\beta(w_k)\}$. Then, by the result (3.3) again, we deduce

$$\sum_{k=0}^{\infty}\Big(\frac{\theta_1 - L_f}{2}\big\|x_{k+1}-x_k\big\|^2 + \frac{\theta_2 - L_g}{2}\big\|y_{k+1}-y_k\big\|^2\Big) \leq \mathcal{L}_\beta(w_0) - \underline{\mathcal{L}}_\beta + \frac{2\delta_0}{\beta\bar{\tau}(1-r)} < +\infty, \quad (3.13)$$

where the last inequality holds by the fact

$$\sum_{k=0}^{\infty}\delta_k \leq \frac{\delta_0}{1-r} < +\infty. \quad (3.14)$$

Then, it follows from (3.13) together with the condition $\theta_1 > L_f$ and $\theta_2 > L_g$ that

$$\lim_{k\to\infty}\big\|x_{k+1}-x_k\big\|^2 = 0, \quad \text{and} \quad \lim_{k\to\infty}\big\|y_{k+1}-y_k\big\|^2 = 0. \quad (3.15)$$

Summarizing the inequalities in (3.1) over $k = 0, 1, \cdots, \infty$ together with (3.14) shows

$$\lim_{k\to\infty}\big\|\mu_{k+1}-\mu_k\big\|^2 = 0, \quad \text{and} \quad \lim_{k\to\infty}\big\|\lambda_k-\mu_k\big\|^2 = 0. \quad (3.16)$$

Combine the following relationship from (3.2):

$$\big\|\lambda_{k+1}-\lambda_k\big\|^2 \leq 2\big\|\lambda_{k+1}-\mu_{k+1}\big\|^2 + 4\big(\big\|\mu_{k+1}-\mu_k\big\|^2 + \big\|\mu_k-\lambda_k\big\|^2\big)$$

with (3.16) immediately ensures

$$\lim_{k\to\infty}\big\|\lambda_{k+1}-\lambda_k\big\|^2 = 0. \quad (3.17)$$

Besides, the update of z_{k+1} gives

$$\big\|z_{k+1}-z_k\big\|^2 \leq \frac{2}{\alpha^2}\Big(\big\|\lambda_{k+1}-\lambda_k\big\|^2 + \big\|\mu_{k+1}-\mu_k\big\|^2\Big),$$

which, by (3.17) and the first result in (3.16), further implies $\lim\limits_{k\to\infty}\|z_{k+1}-z_k\|^2 = 0$. As a results, combine this limitation and (3.16)-(3.17) to confirm the first result in (3.4). The second result in (3.4) is clearly from $\lim\limits_{k\to\infty}\|\lambda_k-\mu_k\|^2 = 0$ and the update of λ_{k+1}. ■

3.2 Convergence and convergence rate

In the following, the distance from any point x to the set Ω is defined as $\text{dist}(x,\Omega) := \inf_{\bar{x}\in\Omega}\|x-\bar{x}\|$. Based on this definition, we first give an estimation on dist $\big(\mathbf{0}, \partial L_\beta(w_{k+1})\big)$ by the iterative residuals, and then analyze the convergence of the iterative sequence $\{w\}$ and its convergence rate. Similar analysis can be found in e.g. [4, 6, 7, 22]. Hereafter, $\mathbf{0}$ stands for a zero vector with proper dimensions. For a proper lower semi-continuous function h, its *(limiting-) subdifferential* [21, Definition 8.3 (b)] at $x \in \mathrm{dom}h$, denoted as $\partial h(x)$, is defined as

$$\partial h(x) := \left\{ \nu \in \mathbb{R}^n : \exists x^k \to x, h(x^k) \to h(x), \nu^k \to \nu \text{ with } \nu^k \in \hat{\partial} h(x^k) \right\}, \tag{3.18}$$

where $\hat{\partial} h(x)$ denotes the *regular subdifferential* [21, Definition 8.3 (a)] of h at $\mathbf{x}$ given as

$$\hat{\partial} h(x) := \left\{ \nu \in \mathbb{R}^n : \liminf_{\bar{x} \to x, \bar{x} \neq x} \frac{h(\bar{x}) - h(x) - \langle \nu, \bar{x} - x \rangle}{\|\bar{x} - x\|} \geq 0 \right\}.$$

Corollary 3.1 *Let $\{w_k = (v_k, z_k, \mu_k)\}$ be the sequence generated by Algorithm 2.1. Then, for every $k \geq 0$, the following hold:*

(i) There exists a F^* such that

$$\lim_{k \to \infty} L_\beta(w_{k+1}) = \lim_{k \to \infty} L(v_{k+1}) = \lim_{k \to \infty} F(x_{k+1}, y_{k+1}) = F^*.$$

(ii) It holds that $\lim_{k \to \infty} \mathrm{dist}\left(\mathbf{0}, \partial L_\beta(w_{k+1})\right) = \mathrm{dist}\left(\mathbf{0}, \partial L(v_{k+1})\right) = 0$.

Proof. Note that

$$\begin{aligned} F(x_{k+1}, y_{k+1}) &= L_\beta(w_{k+1}) - \left\langle \lambda_{k+1}, Ax_{k+1} + By_{k+1} - b \right\rangle + \frac{1}{2\beta} \left\| \lambda_{k+1} - \mu_{k+1} \right\|^2 \\ &= L(x_{k+1}, y_{k+1}, \lambda_{k+1}) - \left\langle \lambda_{k+1}, Ax_{k+1} + By_{k+1} - b \right\rangle, \end{aligned}$$

which ensures the conclusion (i) by the second item of Theorem 3.1, and (3.16).

The first-order optimality condition of x_{k+1}-subproblem implies

$$\mathbf{0} \in \partial f_2(x_{k+1}) + \nabla f_1(x_k) + A^\mathsf{T} \lambda_k + \nabla \phi_1(x_{k+1}) - \nabla \phi_1(x_k).$$

Combining it with the reformulation (2.4) to have

$$e_{k+1}^x \in \partial_x \mathcal{L}_\beta(w_{k+1}),$$

where $e_{k+1}^x := \nabla f_1(x_{k+1}) - \nabla f_1(x_k) + \nabla \phi_1(x_k) - \nabla \phi_1(x_{k+1}) + A^\mathsf{T}(\lambda_{k+1} - \lambda_k)$. Similarly, we have from the first-order optimality condition of y_{k+1}-subproblem that

$$e_{k+1}^y \in \partial_y \mathcal{L}_\beta(w_{k+1}).$$

where $e_{k+1}^y := \nabla g_1(y_{k+1}) - \nabla g_1(y_k) + \nabla \phi_2(y_k) - \nabla \phi_2(y_{k+1}) + B^\mathsf{T}(\lambda_{k+1} - \lambda_k)$. Besides, it follows from the λ-update that

$$\nabla_\lambda \mathcal{L}_\beta(w_{k+1}) = (Ax_{k+1} + By_{k+1} - b) - \frac{1}{\beta}(\lambda_{k+1} - \mu_{k+1}) = \mathbf{0}$$

and $\nabla_\mu \mathcal{L}_\beta(w_{k+1}) = -\frac{1}{\beta}(\mu_{k+1} - \lambda_{k+1}) := e_{k+1}^\mu$. Hence, the following relationship holds:

$$e_{k+1} := \left(e_{k+1}^x, e_{k+1}^y, \mathbf{0}, e_{k+1}^\mu\right) \in \partial L_\beta(w_{k+1}).$$

Next, we simplify the computation of each component of e_{k+1}. By the Lipschitz continuity of f_1 and ϕ_1, we have

$$\begin{aligned}\|e^x_{k+1}\| &\le \|\nabla f_1(x_{k+1}) - \nabla f_1(x_k)\| + \|\nabla \phi_1(x_{k+1}) - \nabla \phi_1(x_k)\| + \|A\|\|\lambda_{k+1} - \lambda_k\| \\ &\le (L_f + L_{\phi_1})\|x_{k+1} - x_k\| + \|A\|\|\lambda_{k+1} - \lambda_k\|.\end{aligned}$$

Analogously, we have by the Lipschitz continuity of g_1 and ϕ_2 that

$$\begin{aligned}\|e^y_{k+1}\| &\le \|\nabla g_1(y_{k+1}) - \nabla g_1(y_k)\| + \|\nabla \phi_2(y_{k+1}) - \nabla \phi_2(y_k)\| + \|B\|\|\lambda_{k+1} - \lambda_k\| \\ &\le (L_g + L_{\phi_2})\|y_{k+1} - y_k\| + \|B\|\|\lambda_{k+1} - \lambda_k\|.\end{aligned}$$

Combining the last two results, the equality $\|e^\mu_{k+1}\| = \frac{1}{\beta}\|\lambda_{k+1} - \mu_{k+1}\|$ and the relationships

$$\begin{cases}\partial_x \mathcal{L}_\beta(w_{k+1}) = \partial_x L(v_{k+1}), \quad \partial_y \mathcal{L}_\beta(w_{k+1}) = \partial_y L(v_{k+1}), \\ \partial_\lambda \mathcal{L}_\beta(w_{k+1}) = \partial_x L(v_{k+1}) - \dfrac{1}{\beta}(\lambda_{k+1} - \mu_{k+1}),\end{cases}$$

to obtain

$$\text{dist}\left(\mathbf{0}, \partial L(v_{k+1})\right) \le \text{dist}\left(\mathbf{0}, \mathcal{L}_\beta(w_{k+1})\right) + \frac{1}{\beta}\|\lambda_{k+1} - \mu_{k+1}\|$$

and

$$\text{dist}\left(\mathbf{0}, \partial L_\beta(w_{k+1})\right) \le \|e_{k+1}\| \le c\big(\|x_{k+1} - x_k\| + \|y_{k+1} - y_k\| + \|\lambda_{k+1} - \lambda_k\| + \|\lambda_{k+1} - \mu_{k+1}\|\big)$$

with $c = \max\{L_f + L_{\phi_1}, L_g + L_{\phi_2}, \|A\| + \|B\|, \frac{1}{\beta}\}$. Then, we confirm the result (ii) by the first equality in (3.4). ■

Corollary 3.1 shows that the objective sequence of (1.1) is convergent, but it does not point the convergence of the iterative sequence as well as its convergence rate. In what follows, we not only show that any limit point of $\{v_k = (x_k, y_k, \lambda_k)\}$ converges to a stationary point of (1.1) as defined by (3.19), but also establishes the sublinear convergence rate of the iterative residuals of the primal variables. We say $(x^*, y^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ is a stationary point of (1.1) if $\mathbf{0} \in \partial L(x, y, \lambda)$, that is,

$$\mathbf{0} \in \nabla f_1(x^*) + \partial f_2(x^*) + A^\top \lambda^*, \ \mathbf{0} \in \nabla g_1(y^*) + \partial g_2(y^*) + B^\top \lambda^*, \ Ax^* + By^* = b. \tag{3.19}$$

Theorem 3.2 *Let $\{w_k = (v_k, z_k, \mu_k)\}$ be the sequence generated by Algorithm 2.1. Then,*

(i) Any limit point v^* of the sequence $\{v_k\}$ is a stationary point of (1.1);

(ii) For any integer $k \ge 1$, there exist $j \le k$ and $\zeta_1, \zeta_1 > 0$ such that

$$\|x_{j+1} - x_j\|^2 \le \frac{\zeta_0}{\zeta_1(k+1)} \quad \textit{and} \quad \|y_{j+1} - y_j\|^2 \le \frac{\zeta_0}{\zeta_2(k+1)},$$

where $\zeta_0 = \mathcal{L}_\beta(w_0) - \underline{\mathcal{L}}_\beta + \frac{2\delta_0}{\beta\bar{\tau}(1-r)}$ with $\underline{\mathcal{L}}_\beta$ being the lower bound of $\{\mathcal{L}_\beta(w_k)\}$.

Proof. For any limit point $w^* = (v^*, z^*, \mu^*)$ of the sequence $\{w_k\}$, it follows from the second conclusion of Corollary 3.1 together with the definition of the limiting-subdifferential $\partial L(v^*)$ and the definition of (3.19) that the conclusion (i) holds.

Secondly, for any $k > 0$, we have from (3.3) and (3.14) that

$$\sum_{j=0}^{k}\Big(\frac{\theta_1 - L_f}{2}\|x_{j+1} - x_j\|^2 + \frac{\theta_2 - L_g}{2}\|y_{j+1} - y_j\|^2\Big) \le \mathcal{L}_\beta(w_0) - \underline{\mathcal{L}}_\beta + \frac{2\delta_0}{\beta\bar{\tau}(1-r)} = \zeta_0,$$

which indicates that there exists a $j \le k$ such that

$$\|x_{j+1} - x_j\|^2 \le \frac{\zeta_0}{(k+1)(\theta_1 - L_f)} \quad \text{and} \quad \|y_{j+1} - y_j\|^2 \le \frac{\zeta_0}{(k+1)(\theta_2 - L_g)}.$$

These inequalities with $\zeta_1 = \theta_1 - L_f > 0, \zeta_2 = \theta_2 - L_g > 0$ confirms the conclusion (ii). ■

References

[1] J. Bai, D. Han, H. Sun, H. Zhang, Convergence on a symmetric accelerated stochastic ADMM with larger stepsizes, CSIAM Trans. Appl. Math., 3, 448-479 (2022)

[2] J. Bai, K. Guo, J. Liang, Y. Jing, H. So, Accelerated symmetric ADMM and its applications in large-scale signal processing, J. Comp. Math., 42: 1605-1626, (2024)

[3] J. Bai, L. Jia, Z. Peng, A new insight on augmented Lagrangian method with applications in machine learning, J. Sci. Comput., 99:53, (2024)

[4] J. Bai, M. Zhang, H. Zhang, An inexact ADMM for separable nonconvex and nonsmooth optimization, Optimization Online, https://optimization-online.org/?p=28132, (2024)

[5] L. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, USSR Comput. Math. Math. Phys., 7: 200-217, (1967)

[6] K. Guo, D. Han, D. Wang, T. Wu, Convergence of ADMM for multi-block nonconvex separable optimization models, Front. Math. China, 12: 1139-1162, (2017)

[7] K. Guo, D. Han, T. Wu, Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints, INT J. Comput. Math., 94: 1653-1669, (2017)

[8] D. Han, W. Kong, W. Zhang, A partial splitting augmented Lagrangian method for low patch-rank image decomposition, J. Math. Imaging Vis., 51: 145-160, (2015)

[9] B. He, X. Yuan, On the acceleration of augmented Lagrangian method for linearly constrained optimization, Optimization Online, https://optimization-online.org/2010/10/2760, (2010)

[10] B. He, H. Xu, X. Yuan, On the proximal jacobian decomposition of ALM for multiple-block separable convex minimization problems and its relationship to ADMM, J. Sci. Comput., 66: 1204-1217, (2016)

[11] M. Hestenes, Multiplier and gradient methods, J. Optim. Theory Appl. 4: 303-320, (1969)

[12] M. Kang, M. Kang, M. Jung, Inexact accelerated augmented Lagrangian methods, Comput. Optim. Appl., 62: 373-404, (2015)

[13] Y. Ke, C. Ma, An accelerated augmented Lagrangian method for linearly constrained convex programming with the rate of convergence $\mathcal{O}(1/k^2)$, Appl. Math. (A Journal of Chinese Universities), 32: 117-126, (2017)

[14] J. Kim, A new Lagrangian-based first-order method for nonconvex constrained optimization, Oper. Res. Lett., 51: 357-363, (2024)

[15] J. Kim, A Lagrangian-based method with "false penalty" for linearly constrained nonconvex composite optimization, arXiv:2306.11299, (2023)

[16] W. Kong, J. Melo, R. Monteiro, Iteration-complexity of an inner accelerated inexact proximal augmented Lagrangian method based on the classical Lagrangian function, SIAM J. Optim. 33: 181-210, (2023)

[17] Y. Li, M. Zhao, W. Chen, Z. Wen, A stochastic composite augmented Lagrangian method for reinforcement learning, SIAM J. Optim., 33:921-949, (2023)

[18] Y. Liu, X. Liu, S. Ma. On the non-ergodic convergence rate of an inexact augmented Lagrangian framework for composite convex programming, Math. Oper. Res., 44: 632-650, (2019)

[19] M. Powell, A method for nonlinear constraints in minimization problems, In Optimization edited by R. Fletcher, pp. 283-298, Academic Press, New York, (1969)

[20] X. Qu, G. Yu, J. Liu, J. Chen, Z. Liu, A new penalty dual-primal augmented Lagrangian method and its extensions, Taiwanese J. Math., 28: 1223-1244, (2024)

[21] R. Rockafellar, R. Wets, Variational Analysis, Springer, New York, (1998)

[22] Z. Wu, M. Li, D. Wang, D. Han, A symmetric alternating direction method of multipliers for separable nonconvex minimization problems, Asia-Pac. J. Oper. Res., 34: 1750030, (2017)

[23] Y. Xu, Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming, SIAM J. Optim., 27: 1459-1484, (2017)

[24] J. Yin, C. Tang, J. Jian, Q. Huang, A partial Bregman ADMM with a general relaxation factor for structured nonconvex and nonsmooth optimization, J. Global Optim., 89: 899-926 (2024)

[25] J. Zeng, W. Yin, D. Zhou, Moreau envelope augmented Lagrangian method for nonconvex optimization with linear constraints, J. Sci. Comput., 91: 61, (2022)