A proximal alternating direction method of multipliers with a proximal-perturbed Lagrangian function for nonconvex and nonsmooth structured optimization

Bohan Li* Pengjie Liu* Hu Shao* Ting Wu† Jiawei Xu‡

December 10, 2024

Abstract. Building on [J. Glob. Optim. 89 (2024) 899–926], we continue to focus on solving a nonconvex and nonsmooth structured optimization problem with linear and closed convex set constraints, where its objective function is the sum of a convex (possibly nonsmooth) function and a smooth (possibly nonconvex) function. Based on the traditional augmented Lagrangian construction, we introduce a proximal-perturbed Lagrangian function and propose a proximal alternating direction method of multipliers that leverages this new Lagrangian-based formulation. We establish that the iterative subsequence obtained by the proposed method converges to a stationary point under standard assumptions.

Key words: Nonconvex and nonsmooth structured optimization, Proximal alternating direction method of multipliers, Proximal-perturbed Lagrangian function, Theoretical convergence

Mathematics Subject Classification (2020): 65K05; 90C30

1 Introduction

In this study, we consider the nonconvex and nonsmooth structured optimization problem:

$$\min \theta_1(p) + \theta_2(q) \quad \text{s.t. } Ap + q = b, \quad p \in \mathcal{P}, \tag{1}$$

where the function $\theta_1 : \mathbb{R}^n \to \mathbb{R}$ is convex, proper lower semicontinuous, and possibly nonsmooth, $\theta_2 : \mathbb{R}^m \to \mathbb{R}$ is continuously differentiable and possibly nonconvex, $A \in \mathbb{R}^{m\times n}$ is a matrix, $b \in \mathbb{R}^m$ is a vector, and $\mathcal{P} \subseteq \mathbb{R}^n$ is a nonempty, closed, convex set. It is well known that separable optimization problems with linear constraints and closed convex set constraints have been extensively studied in the convex setting using splitting algorithms [1, 2, 3]. In fact, nonconvex scenarios are widely encountered in practical engineering applications. For structured problems of the form (1), most studies, such as [3, 4, 5, 6], focus on cases that ignore the closed convex set constraints. Research that simultaneously considers both linear constraints and closed convex set constraints in nonconvex

*School of Mathematics, Jiangsu Center for Applied Mathematics (CUMT), China University of Mining and Technology, Xuzhou 221116, China (han16878@163.com; liupengjie2019@163.com; shaohu@cumt.edu.cn)

†School of Mathematics, Nanjing University, Nanjing 210093, China (tingwu@nju.edu.cn)

‡School of Mathematics and Computational Science, Xiangtan University, Xiangtan 411105, China (xujiawei@smail.xtu.edu.cn)

structured problems remains relatively scarce. Recently, Yin et al. [7] proposed a partial Bregman alternating direction method of multipliers (ADMM) incorporating a general relaxation factor for solving structured problem (1) in the nonconvex setting, which has greatly inspired our work. Herein, we further investigate this problem and design a splitting algorithm with provable convergence to solve it.

On the other hand, Kim [8] observed that the convergence analysis of most existing augmented Lagrangian (AL) methods depends on the boundedness assumption of the dual iterates. To relax this restriction, Kim introduced a new Lagrangian-based formulation, namely the proximal-perturbed Lagrangian function (PPLF), and established the convergence of the AL method to a stationary solution based on standard assumptions without requiring the aforementioned boundedness condition of dual iterates. For the PPLF introduced by Kim [8], it does not include a quadratic penalty term for linear constraints, which structurally differs from the traditional AL function. This design avoids the need for additional and often stricter assumptions when handling linear constraints. Moreover, the penalty parameter and dual proximal parameter involved in the PPLF are relatively easy to select in numerical experiments and are not highly sensitive to the numerical results. Inspired by this formulation, Kim [9] introduced a primal-dual-based method for solving nonconvex composite problems under linear constraints. The solved objective function is the sum of a continuously differentiable (potentially nonconvex) function and a proper closed convex (potentially nonsmooth) function. In a subsequent study, Bai et al. [10] proposed a Bregman ADMM-type algorithm, leveraging the PPLF to solve separable nonconvex and nonsmooth optimization problems. It is worth mentioning that the problems studied in [9, 10] consider only linear constraints, while disregarding closed convex set constraints.

Following the PPLF considered in [8, 9, 10], we propose a proximal ADMM with a PPLF (denote PPLF-PADMM) to solve the structured problem (1) addressed in [7]. We further establish the theoretical convergence of PPLF-PADMM. Specifically, we obtain that the iterative subsequence generated by the PPLF-PADMM converges to a stationary point under standard assumptions.

2 Preliminaries

Throughout this paper, for any symmetric matrix $F \in \mathbb{R}^{n\times n}$, we denote $\|p\|_F^2 := p^\top F p$. For simplicity, let $\Delta_{\min}(F)$ represent the smallest singular value of the linear operator F. Let $\Lambda := (p, q, z, \lambda, \nu)$, and use $\mathcal{L}_\beta(\Lambda_k)$ to represent $\mathcal{L}_\beta(p_k, q_k, z_k, \lambda_k, \nu_k)$. Assume a convex function $\theta : \mathbb{R}^n \to \mathbb{R}$, its subdifferential at $p \in \mathbb{R}^n$ is described as $\partial\theta(p) := \{\xi \in \mathbb{R}^n \mid \theta(q) \geq \theta(p) + \langle \xi, q - p\rangle,$
$\forall\, q \in \mathbb{R}^n\}$. For a nonempty convex set $\mathcal{C} \subseteq \mathbb{R}^n$, its normal cone at $p \in \mathcal{C}$ is given by $N_{\mathcal{C}}(p) = \{\xi \in \mathbb{R}^n \mid \langle \xi, u - p\rangle \leq 0,\ \forall\, u \in \mathcal{C}\}$.

Lemma 2.1 [11] Let $\theta : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function with the gradient $\nabla\theta$ that is L_θ-Lipschitz continuous. Then for any $q, \hat{q} \in \mathbb{R}^n$, we have $|\theta(q) - \theta(\hat{q}) - \langle \nabla\theta(\hat{q}), q - \hat{q}\rangle| \leqslant \frac{L_\theta}{2}\|q - \hat{q}\|^2$.

Proposition 2.1 [12] Let $\theta_1, \theta_2 : \mathbb{R}^n \to \mathbb{R}$ be two functions, and let $\mathcal{C} \subseteq \mathbb{R}^n$ be a nonempty closed convex set.
(i) If the function θ_1 is locally Lipschitz continuous at $p^* \in \mathbb{R}^n$ and reaches its local minimum at p^*, then p^* is a stationary point of θ_1, i.e., $0 \in \partial\theta_1(p^*)$.
(ii) If both θ_1 and θ_2 are subdifferentially regular at p, it follows that $\partial(\theta_1 + \theta_2)(p) = \partial\theta_1(p) + \partial\theta_2(p)$. As a result, if θ_1 is subdifferentially regular at p and θ_2 is continuously differentiable at p, it follows that $\partial(\theta_1 + \theta_2)(p) = \partial\theta_1(p) + \nabla\theta_2(p)$.

(iii) $p^* \in \mathcal{C}$ is a minimizer of the problem $\min\{\theta_1(p) \mid p \in \mathcal{C}\}$, where θ_1 is a convex function, if and only if $0 \in \partial\theta_1(p^*) + N_{\mathcal{C}}(p^*)$, or equivalently, there exists $\xi^* \in \partial\theta_1(p^*)$ such that $\langle \xi^*, p - p^* \rangle \geq 0$ for any $p \in \mathcal{C}$.

Definition 2.1 We say that $(p^*, q^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ is a stationary point of structured problem (1) if it satisfies

$$0 \in \partial\theta_1(p^*) + A^\top \lambda^* + N_{\mathcal{P}}(p^*),\ p^* \in \mathcal{P};\ \nabla\theta_2(q^*) + \lambda^* = 0;\ Ap^* + q^* = b,$$

or equivalently, there exists $\xi^* \in \partial\theta_1(p^*)$ such that

$$p^* \in \mathcal{P},\ \langle \xi^* + A^\top \lambda^*, p - p^* \rangle \geq 0,\ \forall\, p \in \mathcal{P};\ \nabla\theta_2(q^*) + \lambda^* = 0;\ Ap^* + q^* = b.$$

3 The Proposed Algorithm and Its Theoretical Analysis

Inspired by [8, 9, 10], which apply PPLF to the design of first-order algorithms, we reformulate problem (1) by introducing perturbation variable $z \in \mathbb{R}^m$ and letting $Ap + q - b = z$ and $z = 0$. Then, we have

$$\min_{p\in\mathbb{R}^n,\ q\in\mathbb{R}^m} \ \theta_1(p) + \theta_2(q) \ \text{ s.t. } Ap + q - b = z,\ z = 0,\ \ \forall\, p \in \mathcal{P}.$$

For the unique solution $z^* = 0$, the problem mentioned above is equivalent to (1). Now, we present the PPLF of problem (1):

$$\mathcal{L}_\beta(\Lambda) = \theta_1(p) + \theta_2(q) + \langle \lambda, Ap + q - b - z \rangle + \langle \nu, z \rangle + \frac{\gamma}{2}\|z\|^2 - \frac{\beta}{2}\|\lambda - \nu\|^2, \tag{2}$$

where $\lambda \in \mathbb{R}^m$ and $\nu \in \mathbb{R}^m$ are the Lagrange multipliers corresponding to $Ap + q - b = z$ and $z = 0$, respectively. Here, $\gamma > 0$ is a penalty parameter, and $\beta > 0$ is a dual proximal parameter. Notice that the structure of $\mathcal{L}_\beta(\Lambda)$ is different from that of the standard AL function. Specifically, it lacks a penalty term for enforcing the linear constraint $Ap + q - b = z$. Instead, only the auxiliary constraint $z = 0$ is penalized with a quadratic term, while the constraint $Ap + q - b = z$ is relaxed into the objective function using its associated multiplier λ. Additionally, the negative quadratic term $-\frac{\beta}{2}\|\lambda - \nu\|^2$ ensures that $\mathcal{L}_\beta$ is smooth and strongly concave in λ for a fixed ν, and vice versa. Due to the strong convexity of $\mathcal{L}_\beta(\Lambda)$ with respect to the perturbation variable z, a unique solution exists for any given (λ, ν). Minimizing $\mathcal{L}_\beta$ with respect to z yields $z(\lambda, \nu) = (\lambda - \nu)/\gamma$, which implies $\lambda = \nu$ at the unique solution $z^* = 0$. This relationship between λ and ν at $z^* = 0$ motivates the inclusion of the term $-\frac{\beta}{2}\|\lambda - \nu\|^2$ in the PPLF (2). Substituting $z(\lambda, \nu)$ into $\mathcal{L}_\beta(\Lambda)$ leads to the reduced PPLF, i.e.,

$$\mathcal{L}_\beta(p, q, z(\lambda, \nu), \lambda, \nu) = \theta_1(p) + \theta_2(q) + \langle \lambda, Ap + q - b \rangle - \frac{1}{2\rho}\|\lambda - \nu\|^2. \tag{3}$$

Since $\mathcal{L}_\beta(p, q, z(\lambda, \nu), \lambda, \nu)$ is strongly concave in λ for given (p, q, ν), there exists a unique maximizer, denoted by $\lambda(p, q, \nu)$. Maximizing the reduced PPLF (3) with respect to λ yields $\lambda(p, q, \nu) = \arg\max_{\lambda\in\mathbb{R}^m} \mathcal{L}_\beta(p, q, z(\lambda, \nu), \lambda, \nu) = \nu + \rho(Ap + q - b)$, from which the λ-update step in (4d) is derived.

To proceed, based on the above analysis, we present the proposed PPLF-PADMM, which leverages the features of the PPLF to solve problem (1). Throughout the proof and analysis, we suppose that the sequence $\{\Lambda_k\}$ is produced by the PPLF-PADMM and $\mathcal{L}_\beta(\Lambda_0) < +\infty$. The specific iterative steps are outlined as follows:

Algorithm 1 PPLF-PADMM for solving problem (1).

Input: $\gamma \gg 1$, $\beta \in (0,1)$, $\rho := \frac{\gamma}{1+\gamma\beta}$, $r \in (0.9,1)$, $\eta > L_{\theta_2} + 3\rho + \frac{2\rho^2}{\gamma}$, and $\frac{\Delta_{\min}(F)}{2} - \left(\frac{3}{2} + \frac{1}{1+\gamma\beta}\right)\rho\|A\|^2 > 0$.

Initialize: $p_0, q_0, z_0, \lambda_0, \nu_0$ and δ_0.

for $k = 0, 1, ..., K-1$ **do**

$$p_{k+1} = \arg\min_p \left\{ \mathcal{L}_\beta(p, q_k, z_k, \lambda_k, \nu_k) + \frac{1}{2}\|p - p_k\|_F^2 \mid p \in \mathcal{P} \right\}, \tag{4a}$$

$$q_{k+1} = \arg\min_q \left\{ \langle \nabla_q \mathcal{L}_\beta(p_{k+1}, q_k, z_k, \lambda_k, \nu_k), q - q_k \rangle + \frac{\eta}{2}\|q - q_k\|^2 \right\}, \tag{4b}$$

$$\nu_{k+1} = \nu_k + \tau_k(\lambda_k - \nu_k) \text{ with } \tau_k = \frac{\delta_k}{\|\lambda_k - \nu_k\|^2 + 1}, \tag{4c}$$

$$\lambda_{k+1} = \nu_{k+1} + \rho(Ap_{k+1} + q_{k+1} - b), \tag{4d}$$

$$z_{k+1} = \frac{\lambda_{k+1} - \nu_{k+1}}{\gamma}, \quad \delta_{k+1} = r\delta_k. \tag{4e}$$

end for

Output: $\Lambda_{k+1} := (p_{k+1}, q_{k+1}, z_{k+1}, \lambda_{k+1}, \nu_{k+1})$.

Lemma 3.1 The sequences $\{\nu_k\}$, $\{\lambda_k\}$, and $\{z_k\}$ are bounded.

Proof From the ν-update step (4c), we deduce that

$$\begin{aligned}\|\nu_{k+1}\| &= \|\nu_0 + \sum_{i=0}^{k} \tau_i(\lambda_i - \nu_i)\| \leq \|\nu_0\| + \sum_{i=0}^{+\infty} \frac{\delta_i}{\|\lambda_i - \nu_i\|^2 + 1}\|\lambda_i - \nu_i\| \\ &\leq \|\nu_0\| + \sum_{i=0}^{+\infty} \frac{\delta_i}{\|\lambda_i - \nu_i\| + \frac{1}{\|\lambda_i - \nu_i\|}} \leq \|\nu_0\| + \frac{1}{2}\sum_{i=0}^{+\infty} \delta_i.\end{aligned}$$

Note that $\sum_{k=0}^{+\infty} \delta_k$ is convergent, where $\delta_k = r^k\delta_0$ and $r \in (0.9, 1)$. Moreover, since $\|\nu_0\| < +\infty$, $\{\nu_k\}$ is bounded. Furthermore, it follows from (4c) that $\lambda_k = \frac{1}{\tau_k}(\nu_{k+1} - \nu_k) + \nu_k$. Combining (4c) and (4e), we obtain $z_k = \frac{\nu_{k+1} - \nu_k}{\gamma\tau_k}$. Since $\{\nu_k\}$ is bounded, we conclude that both $\{\lambda_k\}$ and $\{z_k\}$ are bounded.

Lemma 3.2 The following four relationships hold.

$$\|\nu_{k+1} - \nu_k\|^2 = \tau_k^2\|\lambda_k - \nu_k\|^2 \leq \frac{\delta_k^2}{4}, \quad \tau_k\|\lambda_k - \nu_k\|^2 \leq \delta_k, \;\; \|\nu_{k+1} - \lambda_k\|^2 = (1 - \tau_k)^2\|\lambda_k - \nu_k\|^2, \tag{5}$$

$$\|\lambda_{k+1} - \lambda_k\|^2 \leq 3\rho^2\|A\|^2\|p_{k+1} - p_k\|^2 + 3\rho^2\|q_{k+1} - q_k\|^2 + 3\|\nu_{k+1} - \nu_k\|^2. \tag{6}$$

Proof It immediately follows from the ν-update step (4c) that

$$\|\nu_{k+1} - \nu_k\|^2 = \tau_k^2\|\lambda_k - \nu_k\|^2 = \frac{\delta_k^2}{\|\lambda_k - \nu_k\|^2 + 2 + \frac{1}{\|\lambda_k - \nu_k\|^2}} \leq \frac{\delta_k^2}{4},$$

and the first relation in (5) holds. By $\tau_k = \frac{\delta_k}{\|\lambda_k - \nu_k\|^2 + 1} \leq 1$, where $\delta_k \in (0, 1]$, and combining $\tau_k\|\lambda_k - \nu_k\|^2 = \frac{\delta_k}{1 + 1/\|\lambda_k - \nu_k\|^2} \leq \delta_k$, we obtain the second relation in (5). Using (4c), we obtain

$\|\lambda_k - \nu_{k+1}\| = \|\lambda_k - \nu_k - \tau_k(\lambda_k - \nu_k)\| = (1-\tau_k)\|\lambda_k - \nu_k\|$, and squaring both sides yields the third relation in (5). Finally, using the λ-update (4d) and the fact that $(a+b+c)^2 \le 3(a^2+b^2+c^2)$ for any $a, b, c \in \mathbb{R}^m$, we obtain (6).

Lemma 3.3 Suppose that the gradient of θ_2 satisfies the Lipschitz condition. Then, for any $k \ge 0$, we have

$$\mathcal{L}_\beta(\Lambda_{k+1}) - \mathcal{L}_\beta(\Lambda_k) \le -c_1\|p_{k+1} - p_k\|^2 - c_2\|q_{k+1} - q_k\|^2 - c_3\|z_{k+1}\|^2 + \widehat{\delta_k}, \tag{7}$$

where $c_1 := \frac{\Delta_{\min}(F)}{2} - \left(\frac{3}{2} + \frac{1}{1+\gamma\beta}\right)\rho\,\|A\|^2$, $c_2 := \frac{1}{2}\left(\eta - \left(\mathcal{L}_{\theta_2} + 3\rho + \frac{2\rho^2}{\gamma}\right)\right)$, $c_3 := \frac{1}{2\gamma}$, and $\widehat{\delta_k} := \frac{3\delta_k^2}{8\rho} + \frac{\delta_k}{\rho} - \frac{\tau_k\delta_k}{2\rho}$.

Proof From the p-update of (4a), we see that

$$\mathcal{L}_\beta(p_{k+1}, q_k, z_k, \lambda_k, \nu_k) - \mathcal{L}_\beta(\Lambda_k) \le -\frac{1}{2}\|p_{k+1} - p_k\|_F^2 \le -\frac{\Delta_{\min}(F)}{2}\|p_{k+1} - p_k\|^2. \tag{8}$$

By the definition of PPLF in (2), we have

$$\begin{aligned}&\mathcal{L}_\beta\left(p_{k+1}, q_{k+1}, z_k, \lambda_k, \nu_k\right) - \mathcal{L}_\beta(p_{k+1}, q_k, z_k, \lambda_k, \nu_k) = \theta_2(q_{k+1}) - \theta_2(q_k) + \langle \lambda_k, q_{k+1} - q_k\rangle \\ &\le \langle \nabla\theta_2(q_k), q_{k+1} - q_k\rangle + \frac{L_{\theta_2}}{2}\|q_{k+1} - q_k\|^2 + \langle \lambda_k, q_{k+1} - q_k\rangle \le -\frac{1}{2}(\eta - L_{\theta_2})\|q_{k+1} - q_k\|^2,\end{aligned} \tag{9}$$

where the first inequality applies Lemma 2.1, and the second inequality follows from the iteration of the q-subproblem (4b). Next, we start by noting that

$$\begin{aligned}&\mathcal{L}_\beta(p_{k+1}, q_{k+1}, z_k, \lambda_{k+1}, \nu_{k+1}) - \mathcal{L}_\beta(p_{k+1}, q_{k+1}, z_k, \lambda_k, \nu_k) \\ &= \underbrace{\langle \lambda_{k+1} - \lambda_k, Ap_{k+1} + q_{k+1} - b\rangle}_{(A)} + \underbrace{\langle (\lambda_k - \nu_k) - (\lambda_{k+1} - \nu_{k+1}), z_k\rangle}_{(B)} \\ &\quad - \frac{\beta}{2}\|\lambda_{k+1} - \nu_{k+1}\|^2 + \frac{\beta}{2}\|\lambda_k - \nu_k\|^2.\end{aligned} \tag{10}$$

By using the update steps (4c) and (4d), we have $\lambda_{k+1} - \nu_{k+1} = \rho(Ap_{k+1} + q_{k+1} - b)$ and $z_k = \frac{1}{\gamma}(\lambda_k - \nu_k)$. Applying the identity $\langle a - b, a\rangle = \frac{1}{2}\|a-b\|^2 + \frac{1}{2}\|a\|^2 - \frac{1}{2}\|b\|^2$ to (A) and (B) with $a = \lambda_k - \nu_k$ and $b = \lambda_{k+1} - \nu_{k+1}$, we obtain

$$(A) = \frac{1}{2\rho}\|\lambda_{k+1} - \lambda_k\|^2 + \frac{1}{2\rho}\|\lambda_{k+1} - \nu_{k+1}\|^2 - \frac{1}{2\rho}\|\nu_{k+1} - \lambda_k\|^2, \tag{11}$$

$$\begin{aligned}(B) &= \frac{1}{2\gamma}\|\rho\, A(p_k - p_{k+1}) + \rho(q_k - q_{k+1}))\|^2 + \frac{1}{2\gamma}\|\lambda_k - \nu_k\|^2 - \frac{1}{2\gamma}\|\lambda_{k+1} - \nu_{k+1}\|^2 \\ &\le \frac{\rho^2\|A\|^2}{\gamma}\|p_{k+1} - p_k\|^2 + \frac{\rho^2}{\gamma}\|q_{k+1} - q_k\|^2 + \frac{1}{2\gamma}\|\lambda_k - \nu_k\|^2 - \frac{1}{2\gamma}\|\lambda_{k+1} - \nu_{k+1}\|^2.\end{aligned} \tag{12}$$

Substituting (11) and (12) into (10) and rearranging terms, we yield

$$
\begin{aligned}
&\mathcal{L}_\beta(p_{k+1}, q_{k+1}, z_k, \lambda_{k+1}, \nu_{k+1}) - \mathcal{L}_\beta(p_{k+1}, q_{k+1}, z_k, \lambda_k, \nu_k) \\
&\le \frac{1}{2\rho}\|\lambda_{k+1} - \lambda_k\|^2 + \frac{\rho^2\|A\|^2}{\gamma}\|p_{k+1} - p_k\|^2 + \frac{\rho^2}{\gamma}\|q_{k+1} - q_k\|^2 - \frac{1}{2\rho}\|\nu_{k+1} - \lambda_k\|^2 + \frac{1}{2\rho}\|\lambda_k - \nu_k\|^2 \\
&\le \frac{1}{2\rho}(3\rho^2\|A\|^2\|p_{k+1} - p_k\|^2 + 3\rho^2\|q_{k+1} - q_k\|^2 + 3\|\nu_{k+1} - \nu_k\|^2) + \frac{\rho\,\|A\|^2}{1+\gamma\beta}\|p_{k+1} - p_k\|^2 \\
&\quad + \frac{\rho^2}{\gamma}\|q_{k+1} - q_k\|^2 + \frac{1}{2\rho}(2\tau_k - \tau_k^2)\|\lambda_k - \nu_k\|^2 \\
&\le \left(\frac{3}{2} + \frac{1}{1+\gamma\beta}\right)\rho\,\|A\|^2\|p_{k+1} - p_k\|^2 + \left(\frac{3\rho}{2} + \frac{\rho^2}{\gamma}\right)\|q_{k+1} - q_k\|^2 + \frac{3\delta_k^2}{8\rho} + \frac{\delta_k}{\rho} - \frac{\tau_k\delta_k}{2\rho},
\end{aligned}
\tag{13}
$$

where the second inequality applies the third relation in (5) and (6), and the third inequality uses the first and second relations in (5). For the z-update expression, it notes that $\nabla_z\mathcal{L}_\beta(\Lambda_{k+1}) = 0$ since z_{k+1} minimizes $\mathcal{L}_\beta\left(p_{k+1}, q_{k+1}, z, \lambda_{k+1}, \nu_{k+1}\right)$. Hence, by the γ-strong convexity of $\mathcal{L}_\beta$ with respect to z, we have

$$
\mathcal{L}_\beta(\Lambda_{k+1}) - \mathcal{L}_\beta(p_{k+1}, q_{k+1}, z_k, \lambda_{k+1}, \nu_{k+1}) \le -\frac{\gamma}{2}\|z_{k+1} - z_k\|^2 \le -\frac{1}{2\gamma}\|z_{k+1}\|^2,
$$

where the last inequality follows from the existence of a sufficiently large positive γ such that $\|z_{k+1}\| \le \gamma\|z_{k+1} - z_k\|$ for any $k > 0$, which implies that $-\frac{\gamma}{2}\|z_{k+1} - z_k\|^2 \le -\frac{1}{2\gamma}\|z_{k+1}\|^2$. Combining the above inequality with (8), (9), and (13), we yield the desired result (7).

Theorem 3.1 Suppose the conditions in Lemma 3.3 hold, and let θ_1 and θ_2 satisfy the coercivity condition. Then the following hold:
(i) The sequences $\{p_k\}$ and $\{q_k\}$ are bounded. Moreover, $\mathcal{L}_\beta(\Lambda_k)$ is convergent.
(ii) $\lim_{k\to+\infty}(\|p_{k+1} - p_k\| + \|q_{k+1} - q_k\| + \|z_{k+1}\| + \|\lambda_{k+1} - \lambda_k\| + \|\nu_{k+1} - \nu_k\|) = 0$.
(iii) Any accumulation point (p^*, q^*, λ^*) of the sequence $\{(p_k, q_k, \lambda_k)\}$ is a stationary point of (1).

Proof (i) Using the inequality in (7), we have $\mathcal{L}_\beta(\Lambda_0) + \sum_{i=0}^{k-1}\left(\frac{3\delta_0^2}{8\rho}r^{2i} + \frac{\delta_0}{\rho}r^i\right) \ge \mathcal{L}_\beta(\Lambda_{k-1}) + \frac{3\delta_{k-1}^2}{8\rho} + \frac{\delta_{k-1}}{\rho} \ge \mathcal{L}_\beta(\Lambda_k)$. Based on the definition of PPLF in (2), we obtain

$$
\begin{aligned}
\mathcal{L}_\beta(w_k) &= \theta_1\left(p_k\right) + \theta_2\left(q_k\right) + \langle\lambda_k, Ap_k + q_k - b\rangle - \langle\lambda_k - \nu_k, z_k\rangle + \frac{\gamma}{2}\left\|z_k\right\|^2 - \frac{\beta}{2}\left\|\lambda_k - \nu_k\right\|^2 \\
&= \theta_1\left(p_k\right) + \theta_2\left(q_k\right) + \frac{1}{\rho}\left\langle\lambda_k, \lambda_k - \nu_k\right\rangle - \frac{1}{2\rho}\left\|\lambda_k - \nu_k\right\|^2 \\
&= \theta_1\left(p_k\right) + \theta_2\left(q_k\right) + \frac{1}{2\rho}\left\|\lambda_k\right\|^2 - \frac{1}{2\rho}\left\|\nu_k\right\|^2.
\end{aligned}
$$

Combining the fact that θ_1 and θ_2 satisfy the coercivity condition, i.e., $\lim_{\|p\|\to+\infty}\theta_1(p) = +\infty$ and $\lim_{\|q\|\to+\infty}\theta_2(q) = +\infty$, with the boundedness of the sequences $\{\nu_k\}$ and $\{\lambda_k\}$, we conclude that both $\{p_k\}$ and $\{q_k\}$ are bounded. Since $\mathcal{L}_\beta(\Lambda_k)$ is lower semicontinuous and nonincreasing, the sequence $\{\mathcal{L}_\beta(\Lambda_k)\}$ is nonincreasing and bounded below, hence $\{\mathcal{L}_\beta(\Lambda_k)\}$ converges to a finite value $\mathcal{L}_\beta^*$.
(ii) Suppose that Λ^* is an accumulation point of the sequence $\{\Lambda_k\}$, and consider a convergent subsequence, i.e., $\lim_{i\to\infty}\Lambda_{k_i} = \Lambda^*$. Summing (7) from $k = 0$ to $k = K-1$, using the result from

conclusion (i), and letting $c_0 := \min\{c_1, c_2, c_3\}$, we obtain

$$\begin{aligned}\sum_{k=0}^{K-1}\left(\|p_{k+1}-p_k\|^2+\|q_{k+1}-q_k\|^2+\|z_{k+1}\|^2\right) &\le \frac{1}{c_0}\left(\mathcal{L}_\beta(\Lambda_0)-\mathcal{L}_\beta(\Lambda_k)+\sum_{k=0}^{K-1}\widehat{\delta_k}\right)\\ &\le \frac{1}{c_0}\left(\mathcal{L}_\beta(\Lambda_0)-\mathcal{L}_\beta^*+\sum_{k=0}^{K-1}\widehat{\delta_k}\right).\end{aligned}$$

Since $\sum_{k=0}^{+\infty}\delta_k \le \frac{\delta_0}{2(1-r)} < +\infty$ and $\sum_{k=0}^{+\infty}\delta_k^2 \le \frac{\delta_0^2}{2(1-r^2)} < +\infty$, and $\tau_k \in (0,1)$, we have $\sum_{k=0}^{+\infty}\hat{\delta}_k < +\infty$. Then, by taking the limit as $K \to +\infty$, we deduce

$$\sum_{k=0}^{+\infty}\|p_{k+1}-p_k\|^2 < +\infty, \quad \sum_{k=0}^{+\infty}\|q_{k+1}-q_k\|^2 < +\infty, \quad \sum_{k=0}^{+\infty}\|z_{k+1}\|^2 < +\infty.$$

On the other hand, it can be deduced from (5) and $\sum_{k=0}^{+\infty}\delta_k^2 < +\infty$ that $\sum_{k=0}^{+\infty}\|\nu_{k+1}-\nu_k\|^2 \le \sum_{k=0}^{+\infty}\frac{\delta_k^2}{4} < +\infty$. Finally, by combining (6), we obtain

$$\sum_{k=0}^{+\infty}\|\lambda_{k+1}-\lambda_k\|^2 \le 3\rho^2\|A\|^2\sum_{k=0}^{+\infty}\|p_{k+1}-p_k\|^2 + 3\rho^2\sum_{k=0}^{+\infty}\|q_{k+1}-q_k\|^2 + 3\sum_{k=0}^{+\infty}\|\nu_{k+1}-\nu_k\|^2 < +\infty.$$

Therefore, conclusion (ii) follows immediately.
(iii) We begin by noting from the first-order optimality condition of PPLF-PADMM that $p_{k+1} \in \mathcal{P}$, and there exists $\xi_{k+1} \in \partial\theta_1(p_{k+1})$ such that

$$\begin{cases}\langle \xi_{k+1} + A^\top\lambda_k + F(p_{k+1}-p_k),\ p_{k+1}-p\rangle \le 0,\ \forall\, p \in \mathcal{P}, & \text{(14a)}\\ \nabla\theta_2(q_k)+\lambda_k = 0, \quad \rho(Ap_{k+1}+q_{k+1}-b) = \lambda_{k+1}-\nu_{k+1}, \quad \lambda_{k+1}-\nu_{k+1} = \gamma z_{k+1}, & \text{(14b)}\end{cases}$$

It is straightforward to see that (14a) is equivalent to $p_{k+1} \in \mathcal{P}$ and $\langle A^\top\lambda_k + F(p_{k+1}-p_k), p_{k+1}-p\rangle \le \langle \xi_{k+1}, p-p_{k+1}\rangle$. Notice that from the convexity of θ_1 and $\xi_{k+1} \in \partial\theta_1(p_{k+1})$, we have $\langle \xi_{k+1}, p-p_{k+1}\rangle \le \theta_1(p)-\theta_1(p_{k+1})$ for any $p \in \mathcal{P}$. Hence, we obtain

$$p_{k+1} \in \mathcal{P}, \quad \theta_1(p_{k+1}) + \langle \lambda_k, Ap_{k+1}\rangle + \langle F(p_{k+1}-p_k), p_{k+1}-p\rangle \le \theta_1(p) + \langle \lambda_k, Ap\rangle.$$

In the above-mentioned relationship, let $k := k_i$, and then take the limit as $i \to +\infty$, invoking $\lim_{i\to+\infty}\Lambda_{k_i} = \Lambda^*$, conclusion (ii), and using the closeness of $\mathcal{P}$, we obtain

$$p^* \in \mathcal{P}, \quad \theta_1(p^*) + \langle \lambda^*, Ap^*\rangle \le \theta_1(p) + \langle \lambda^*, Ap\rangle, \quad \forall\, p \in \mathcal{P},$$

which implies that $p^* = \arg\min_{p\in\mathcal{P}} \theta_1(p) + \langle \lambda^*, Ap\rangle$. According to the first-order optimality condition, there exists $\xi^* \in \partial\theta_1(p^*)$ such that

$$p^* \in \mathcal{P}, \quad \langle \xi^* + A^\top\lambda^*, p^*-p\rangle \le 0, \quad \forall\, p \in \mathcal{P}.$$

Using (14b), taking $k := k_i$, passing to the limit as $i \to \infty$, and invoking $\lim_{i\to\infty}\Lambda_{k_i} = \Lambda^*$, we conclude that $\nabla\theta_2(q^*) + \lambda^* = 0$ and $\rho(Ap^*+q^*-b) = \lambda^*-\nu^* = \gamma z^* = 0$. From the above results and Definition 2.1, it follows that any accumulation point (p^*, q^*, λ^*) of the sequence $\{(p_k, q_k, \lambda_k)\}$ is a stationary point of (1).

4 Conclusion

In this work, we address the problem of solving a nonconvex and nonsmooth structured optimization problem with linear and closed convex set constraints, where the objective is the sum of a convex (possibly nonsmooth) function and a smooth (possibly nonconvex) function. Building upon the augmented Lagrangian function, we introduce a new Lagrangian-based formulation and develop a PPLF-ADMM tailored to this problem. We demonstrate that the sequence obtained by the proposed PPLF-ADMM converges to a stationary point under standard assumptions. Our results are expected to provide a reference algorithm for efficiently solving composite practical optimization problems.

References

[1] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning with the alternating direction method of multipliers, Found. Trends Mach. Learn. **3**(1) 1-122 (2011).

[2] Bai, J.C., Hager, W.W., Zhang, H.: An inexact accelerated stochastic ADMM for separable convex optimization, Comput. Optim. Appl. **81** 479–518 (2022).

[3] Han, D.R.: A survey on some recent developments of alternating direction method of multipliers, J. Oper. Res. Soc. China **10** 1-52 (2022).

[4] Bai, J.C., Zhang, M., Zhang, H.C.: An inexact ADMM for separable nonconvex and nonsmotth optimization, Comput. Optim. Appl., Revision, (2024). https://optimization-online.org/?p=28132

[5] Liu, P.J., Jian, J.B., Ma, G.D.: A Bregman-style partially symmetric alternating direction method of multipliers for nonconvex multi-block optimization, Acta Math. Appl. Sin. Engl. Ser. **39** 354–380 (2023).

[6] Liu, P.J., Jian, J.B., Shao, H., Wang, X.Q., Xu, J.W., Wu, X.Y.: A Bregman-Style improved ADMM and its linearized version in the nonconvex setting: Convergence and rate analyses, J. Oper. Res. Soc. China **12** 298–340 (2024).

[7] Yin, J.H., Tang, C.M., Jian, J.B., Huang, Q.X.: A partial Bregman ADMM with a general relaxation factor for structured nonconvex and nonsmooth optimization, J. Glob. Optim. **89** 899–926 (2024).

[8] Kim, J.G.: A new Lagrangian-based first-order method for nonconvex constrained optimization, Oper. Res. Letters, **51**(3) 357-363 (2023).

[9] Kim, J.G.: A Lagrangian-based method with "False Penalty" for linearly constrained nonconvex composite optimization, 2023. https://arxiv.org/abs/2306.11299

[10] Bai, J.C., Wang, B.R.: A proximal-perturbed Bregman ADMM for solving nonsmooth and nonconvex optimization problems, 2024. https://optimization-online.org/wp-content/uploads/2024/12/BP-ALM12-2.pdf

[11] Nesterov, Y.: Introduction Lectures on Convex Optimization: A Basic Course. Springer Science & Business Media, Berlin, 2013.

[12] Rockafellar, R.T., Wets, R.J.B.: Variational Analysis, Springer Science & Business Media, 2009.