TAIWANESE JOURNAL OF MATHEMATICS
Vol. xx, No. x, pp. 1–22, xx 20xx
DOI: 10.11650/tjm/240603

A New Penalty Dual-primal Augmented Lagrangian Method and its Extensions

Xiaoqing Ou, Guolin Yu, Jie Liu, Jiawei Chen* and Zhaohan Liu

Abstract. In this paper, we propose a penalty dual-primal balanced-based augmented Lagrangian method for solving linearly constrained convex minimization problems. Convergence and convergence rate of the penalty dual-primal balanced-based augmented Lagrangian method are established by the tool of variational inequality. Further, we generalize the penalty dual-primal balanced-based augmented Lagrangian method to solve linearly constrained multi-block separable convex minimization problems with full splitting technique and partial splitting technique. Numerical results on the basic pursuit problem and the Lasso model are presented to illustrate the efficiency of the proposed methods.

1. Introduction

In this paper, we consider the following linearly constrained convex minimization problem

$$\min_x \{\theta(x) \mid Ax = b, x \in \mathscr{X}\}, \tag{1.1}$$

where $\theta\colon \mathbb{R}^n \to \mathbb{R}$ is a convex but not necessarily smooth function, $\mathscr{X}$ is a nonempty closed and convex set of $\mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ is a given matrix and $b \in \mathbb{R}^m$ is a known vector. The problem (1.1) is assumed to have solution throughout this paper.

There are many algorithms that can be used to solve problem (1.1), where a benchmark method is the Augmented Lagrangian Method (ALM) proposed in [17, 19]. It plays a significant role in both algorithmic design and practical applications for various convex optimization problems; see [2, 3, 7, 10, 11, 18, 20] and the references therein.

For a given iterate (x^k, λ^k), the iterative scheme of the classical ALM reads as

$$\begin{cases} x^{k+1} = \arg\min_x \{\mathscr{L}_\beta(x, \lambda^k) \mid x \in \mathscr{X}\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - b), \end{cases} \tag{1.2}$$

Received October 30, 2023; Accepted June 7, 2024.
Communicated by Jein-Shan Chen.
2020 *Mathematics Subject Classification.* 65K05, 65K15.
Key words and phrases. convex minimization, augmented Lagrangian method, penalty, convergence, variational inequality, Lasso.
*Corresponding author.

where $\mathscr{L}_\beta(x,\lambda) := \theta(x) - \lambda^\top(Ax-b) + \frac{\beta}{2}\|Ax-b\|^2$ is the augmented Lagrangian function of the problem (1.1), $\beta > 0$ is the penalty parameter, $\lambda \in \mathbb{R}^m$ is the associated Lagrange multiplier. Hereafter, x and λ are referred to the primal and dual variables respectively, and I and $\mathbf{0}$ are regarded as a identity matrix and a zero matrix with proper dimensions, respectively.

Ignoring some constant terms, the x-subproblem of (1.2) can be rewritten as

$$x^{k+1} = \arg\min_x \left\{ \theta(x) + \frac{\beta}{2}\left\| Ax - b - \frac{1}{\beta}\lambda^k \right\|^2 \;\middle|\; x \in \mathscr{X} \right\}.$$

It is obvious that the objective function θ, the coefficient matrix A, and the set $\mathscr{X}$ are all appear at the same time, so it is still difficult to be solved if without utilizing some linearization techniques or inner solvers. Some existing algorithms can be applied to decouple the objective function θ and coefficient matrix A, so as to alleviate the x-subproblem substantially, such as the linearized ALM [13] and primal-dual method [6]. In the above-mentioned algorithm, the x-subproblem only depends on θ and $\mathscr{X}$, and the proximity operator of the objective function θ, which is defined as

$$\mathrm{Prox}_\theta^\beta(x) := \arg\min_y \left\{ \theta(y) + \frac{\beta}{2}\|y - x\|^2 \;\middle|\; y \in \mathbb{R}^n \right\}, \quad \forall\, x \in \mathbb{R}^n$$

has a closed-form representation, where $\beta > 0$. In order to ensure convergence, there is an extra restriction on step-size , i.e., $\sigma > \beta\|A^\top A\|$, where $\sigma > 0$ and $\|A^\top A\|$ represent the spectral norm of $A^\top A$. Hence the step-size for solving (1.2) becomes small when $\|A^\top A\|$ is large, and so the convergence rate will be low. Recently, a balanced version of ALM was firstly proposed by He and Yuan [16], which has no limitation on step-size and takes the following iterative scheme

$$\text{(1.3)} \quad \text{(Balanced ALM)} \begin{cases} x^{k+1} = \arg\min_x \left\{ \theta(x) + \frac{1}{\beta}\left\| x - \left(x^k + \frac{1}{\beta}\right)A^\top\lambda^k \right\|^2 \mid x \in \mathscr{X} \right\}, \\ \lambda^{k+1} = \lambda^k - \left(\frac{1}{\beta}AA^\top + \delta I_m\right)^{-1}\{A(2x^{k+1} - x^k) - b\}, \end{cases}$$

where $\beta > 0$ and $\delta > 0$. It is clear that the x-subproblem of the balanced ALM decouples the objective function and the coefficient matrix without any extra condition. Namely, the parameter β does not depend on $\|A^\top A\|$, and the x-subproblem has a closed-form solution since its solution can be expressed as a proximal mapping. However, the balanced ALM will take much time to update λ^{k+1} since it needs to calculate the inverse of matrix $\frac{1}{\beta}AA^\top + \delta I_m$, i.e., the matrix $\left(\frac{1}{\beta}AA^\top + \delta I_m\right)^{-1}$, and in practice it will take an inner solver to tackle the dual subproblem or use the well-known Cholesky factorization to deal with an equivalent linear equation of dual problem. In this sense, a new penalty ALM was proposed in [1] to solve it, which reads as

$$\text{(1.4)} \quad \begin{cases} x^{k+1} = \arg\min_x \left\{ \theta(x) - \langle \lambda^k, Ax - b\rangle + \frac{\beta}{2}\|A(x - x^k)\|^2 + \frac{1}{2}\|x - x^k\|_Q^2 \mid x \in \mathscr{X} \right\}, \\ \lambda^{k+1} = \lambda^k - \beta\left[A(2x^{k+1} - x^k) - b\right], \end{cases}$$

where $\beta > 0$, $Q \succ 0$ is an arbitrarily given positive-defined matrix, the term $\frac{\beta}{2}\|A(x-x^k)\|^2$ can be treated as a penalty term, while the quadratic term $\frac{1}{2}\|x-x^k\|_Q^2$ can be regarded as a penalty term.

Both the balanced ALM and the new Penalty ALM update the new iterate by a primal-dual order. Exploiting the variational inequality structure of the balanced ALM, a dual-primal version of the balanced ALM was proposed in [22]. The proposed method [22] generates the new iterates by a dual-primal order and enjoys the same computational difficulty with the original primal-dual balanced ALM, which reads as

$$\begin{cases} \lambda^{k+1} = \lambda^k - \left(\frac{1}{\beta}AA^\top + \delta I_m\right)^{-1}(Ax^k - b), \\ x^{k+1} = \arg\min_x \left\{\theta(x) + \frac{\beta}{2}\left\|x - \left\{x^k + \frac{1}{\beta}A^\top(2\lambda^{k+1} - \lambda^k)\right\}\right\|^2 \;\middle|\; x \in \mathscr{X}\right\}, \end{cases} \tag{1.5}$$

where $\beta > 0$ and $\delta > 0$.

Observe that the original primal-dual balanced ALM also will take much time to update λ^{k+1}, and in practice it will take an inner solver or use the well-known Cholesky factorization to deal with an equivalent linear equation of dual problem being the same as the balanced ALM. Motivated by the works [1,16,22], our main purpose is to alleviate the difficulty for solving dual-subproblem of the original primal-dual balanced ALM (1.5) by utilizing the novel penalty technique proposed in [1]. We propose a penalty dual-primal ALM combines a novel penalty technique with updating the new iterates in a dual-primal order, as follows:

Algorithm 1.1 (The novel penalty dual-primal ALM)**.** *Let $\beta > 0$ and $Q \succ 0$ be an arbitrarily given positive-defined matrix. Then the new iterate $\omega^{k+1} = (x^{k+1}, \lambda^{k+1})$ is generated with $\omega^k = (x^k, \lambda^k)$ via the following steps:*

$$\begin{cases} \lambda^{k+1} = \lambda^k - \beta(Ax^k - b), \\ x^{k+1} = \arg\min_x \left\{\theta(x) - \langle 2\lambda^{k+1} - \lambda^k, Ax - b\rangle + \frac{\beta}{2}\|A(x - x^k)\|^2 \right. \\ \qquad\qquad \left. + \frac{1}{2}\|x - x^k\|_Q^2 \;\middle|\; x \in \mathscr{X}\right\}. \end{cases} \tag{1.6}$$

In (1.6), the quadratic term $\frac{\beta}{2}\|A(x-x^k)\|^2$ can be treated as a penalty term, while the quadratic term $\frac{1}{2}\|x-x^k\|_Q^2$ can be regarded as a regularization term.

Clearly, the x-subproblem of (1.6) can be rewritten equivalently as

$$x^{k+1} = \arg\min_x \left\{\theta(x) - \langle 2\lambda^{k+1} - \lambda^k, Ax - b\rangle + \frac{1}{2}\|x - x^k\|_{\beta A^\top A + Q}^2 \;\middle|\; x \in \mathscr{X}\right\}.$$

If $Q := \tau I - \beta A^\top A$ with $\tau > \beta\|A^\top A\|$, the x-update is reduced to

$$x^{k+1} = \arg\min_x \left\{\theta(x) + \frac{\tau}{2}\left\|x - x^k - \frac{1}{\tau}A^\top(2\lambda^{k+1} - \lambda^k)\right\|^2 \;\middle|\; x \in \mathscr{X}\right\}.$$

If $Q := \beta(\tau I - A^\top A)$ with $\tau > \|A^\top A\|$, the x-update is reduced to

$$x^{k+1} = \arg\min_x \left\{ \theta(x) + \frac{\tau\beta}{2} \left\| x - x^k - \frac{1}{\tau} A^\top (2\lambda^{k+1} - \lambda^k) \right\|^2 \;\Big|\; x \in \mathscr{X} \right\},$$

which has a closed-form solution by proximity operator of the objective function $\theta(x)$.

The dual update of (1.6) is similar to one of [6] and is comparatively much easier than that of the dual-primal balanced ALM [22]. Compared with some existing splitting algorithms, the convergence of this penalty dual-primal ALM (1.6) does not depend on the value of $\|A^\top A\|$. We also raise the extension versions of the proposed penalty dual-primal ALM (1.6) to tackle the multi-block separable convex minimization problem with both linear equality and inequality constraints.

The paper is organized as follows. In Section 2, we recall some preliminaries. In Section 3, we establish the convergence analysis of the penalty dual-primal ALM. We extend the proposed method to solve the multiple-block separable convex problems and show its convergence analysis in Section 4. In Section 5, we further give the partial splitting version and its convergence analysis. In Section 6, we present some computational experiments. Finally, we give the conclusions.

2. Preliminaries

In this section, we recall some fundamental variational inequality characterization and lemmas. Let $\mathbb{R}^n$ be the n-dimensional Euclidean space with inner product $\langle x, y\rangle = x^\top y = \sum_{i=1}^n x_i y_i$ and norm $\|x\| = \sqrt{\langle x, x\rangle}$, where $x, y \in \mathbb{R}^n$ and $\top$ stands for the transpose operation.

We first recall the optimality condition of the model (1.1) in the lens of variational inequality; see, e.g., [12, 14, 15]. The Lagrangian function of model (1.1) is defined as

$$\mathscr{L}(x, \lambda) := \theta(x) - \lambda^\top (Ax - b), \tag{2.1}$$

where $\lambda \in \mathbb{R}^m$ is the Lagrange multiplier. For the simplicity, we set $\Omega := \mathscr{X} \times \Lambda$ and $\Lambda := \mathbb{R}^m$.

The pair $(x^*, \lambda^*) \in \Omega$ is called a saddle point of Lagrangian function (2.1) which means that x^* is a solution point of (1.1), if it satisfies

$$\mathscr{L}(x^*, \lambda) \le \mathscr{L}(x^*, \lambda^*) \le \mathscr{L}(x, \lambda^*), \quad \forall\, (x, \lambda) \in \Omega. \tag{2.2}$$

(2.2) can be separately rewritten as the following mixed variational inequalities

$$\begin{aligned} &x^* \in \mathscr{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^\top (-A^\top \lambda^*) \ge 0, \quad \forall\, x \in \mathscr{X}, \\ &\lambda^* \in \Lambda, \quad (\lambda - \lambda^*)^\top (Ax^* - b) \ge 0, \quad \forall\, \lambda \in \Lambda, \end{aligned}$$

which can be further reformulated as the following compact form

$$\omega^* \in \Omega, \quad \theta(x) - \theta(x^*) + (\omega - \omega^*)^\top F(\omega^*) \geq 0, \quad \forall\, \omega \in \Omega, \tag{2.3}$$

where

$$\omega = \begin{pmatrix} x \\ \lambda \end{pmatrix} \quad \text{and} \quad F(\omega) = \begin{pmatrix} -A^\top \lambda \\ Ax - b \end{pmatrix}. \tag{2.4}$$

Clearly, the operator $F(\omega)$ defined by (2.4) is affine with a skew-symmetric matrix satisfying

$$(\omega - \widetilde{\omega})^\top (F(\omega) - F(\widetilde{\omega})) = 0, \quad \forall\, \omega, \widetilde{\omega} \in \Omega. \tag{2.5}$$

We denote by Ω^* the solution set of the variational inequality (2.3).

The following basic lemma will be used frequently for our further discussions.

Lemma 2.1. [16] *Let $\mathscr{X}$ be a closed convex set and let $\theta, f \colon \mathbb{R}^n \to \mathbb{R}$ be convex functions. If f is differentiable, and the solution set of the minimization problem*

$$\min_x \{\theta(x) + f(x) \mid x \in \mathscr{X}\}$$

is nonempty, then it holds that

$$x^* \in \arg\min_x \{\theta(x) + f(x) \mid x \in \mathscr{X}\}$$

if and only if

$$x^* \in \mathscr{X}, \quad \theta(x) - \theta(x^*) + (x - x^*)^\top \nabla f(x^*) \geq 0, \quad \forall\, x \in \mathscr{X}.$$

We recall the so called Fejér convergence theorem, which will be used in our convergence analysis.

Definition 2.2. A sequence $\{u^k\} \subset \mathbb{R}^n$ is called Fejér convergent with respect to a nonempty subset U of $\mathbb{R}^n$ if, for every $u \in U$,

$$\|u^{k+1} - u\| \leq \|u^k - u\|, \quad \forall\, k \in \mathbb{N}.$$

The following result can be derived from Theorem 1 of [5].

Lemma 2.3. *If $\{u^k\} \subset \mathbb{R}^n$ is a Fejér convergent sequence with respect to a nonempty set U, then $\{u^k\}$ is bounded. Furthermore, if a cluster point $\overline{u}$ of $\{u^k\}$ belongs to U, then $\lim_{k\to\infty} u^k = \overline{u}$.*

3. Convergence analysis

In this section, we establish the convergence analysis of the proposed penalty dual-primal ALM, following the analogous analysis method in [16]. We prove the following lemma which plays a key role in convergence analysis of Algorithm 1.1.

Lemma 3.1. *Let $\{\omega^k\}$ be the sequence generated by Algorithm* 1.1*. Then $\omega^{k+1}\in\Omega$ and*

$$\theta(x)-\theta(x^{k+1})+(\omega-\omega^{k+1})^\top F(\omega^{k+1})\geq(\omega-\omega^{k+1})^\top H(\omega^k-\omega^{k+1}),\quad\forall\,\omega\in\Omega, \tag{3.1}$$

where

$$H=\begin{pmatrix}\beta A^\top A+Q & -A^\top\\ -A & \frac{1}{\beta}I\end{pmatrix}. \tag{3.2}$$

Proof. For the x-subproblem in (1.6), it follows from Lemma 2.1 that $x^{k+1}\in\mathscr{X}$,

$$\theta(x)-\theta(x^{k+1})+(x-x^{k+1})^\top\big\{-A^\top(2\lambda^{k+1}-\lambda^k)+(\beta A^\top A+Q)(x^{k+1}-x^k)\big\}\geq 0$$

for all $x\in\mathscr{X}$, which can be equivalently reformulated as

$$\begin{aligned}&x^{k+1}\in\mathscr{X},\quad\theta(x)-\theta(x^{k+1})+(x-x^{k+1})^\top(-A^\top\lambda^{k+1})\\&\qquad\geq(x-x^{k+1})^\top\big\{-A^\top(\lambda^k-\lambda^{k+1})+(\beta A^\top A+Q)(x^k-x^{k+1})\big\},\quad\forall\,x\in\mathscr{X}.\end{aligned} \tag{3.3}$$

For the λ-subproblem in (1.6), we have

$$Ax^k-b+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)=0,$$

which implies that

$$(\lambda-\lambda^{k+1})^\top\left\{Ax^{k+1}-b-A(x^{k+1}-x^k)+\frac{1}{\beta}(\lambda^{k+1}-\lambda^k)\right\}\geq 0,\quad\forall\,\lambda\in\Lambda,$$

and so,

$$\begin{aligned}&(\lambda-\lambda^{k+1})^\top(Ax^{k+1}-b)\\&\geq(\lambda-\lambda^{k+1})^\top\left\{-A(x^k-x^{k+1})+\frac{1}{\beta}(\lambda^k-\lambda^{k+1})\right\},\quad\forall\,\lambda\in\Lambda.\end{aligned} \tag{3.4}$$

Combining (3.3) and (3.4), we have

$$\begin{aligned}&\theta(x)-\theta(x^{k+1})+(\omega-\omega^{k+1})^\top\begin{pmatrix}-A^\top\lambda^{k+1}\\ Ax^{k+1}-b\end{pmatrix}\\&\geq(\omega-\omega^{k+1})^\top\begin{pmatrix}\beta A^\top A+Q & -A^\top\\ -A & \frac{1}{\beta}I\end{pmatrix}(\omega^k-\omega^{k+1}).\end{aligned}$$

Consequently, one has

$$\theta(x)-\theta(x^{k+1})+(\omega-\omega^{k+1})^\top F(\omega^{k+1})\geq(\omega-\omega^{k+1})^\top H(\omega^k-\omega^{k+1}).$$

□

Convergence of the penalty dual-primal ALM depends on the positive definiteness of the matrix H which is proved by the following proposition.

Proposition 3.2. *The matrix H defined in* (3.2) *is positive definite.*

Proof. Note that

$$\begin{aligned} H &= \begin{pmatrix} \beta A^\top A + Q & -A^\top \\ -A & \frac{1}{\beta} I \end{pmatrix} = \begin{pmatrix} \beta A^\top A & -A^\top \\ -A & \frac{1}{\beta} I \end{pmatrix} + \begin{pmatrix} Q & 0 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} -\sqrt{\beta} A^\top \\ \frac{1}{\sqrt{\beta}} I \end{pmatrix} \begin{pmatrix} -\sqrt{\beta} A & \frac{1}{\sqrt{\beta}} I \end{pmatrix} + \begin{pmatrix} Q & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

Then, we have

$$\begin{aligned} \omega^\top H \omega &= (x, \lambda) \left(\begin{pmatrix} -\sqrt{\beta} A^\top \\ \frac{1}{\sqrt{\beta}} I \end{pmatrix} \begin{pmatrix} -\sqrt{\beta} A & \frac{1}{\sqrt{\beta}} I \end{pmatrix} + \begin{pmatrix} Q & 0 \\ 0 & 0 \end{pmatrix} \right) \begin{pmatrix} x \\ \lambda \end{pmatrix} \\ &= (x, \lambda) \begin{pmatrix} -\sqrt{\beta} A^\top \\ \frac{1}{\sqrt{\beta}} I \end{pmatrix} \begin{pmatrix} -\sqrt{\beta} A & \frac{1}{\sqrt{\beta}} I \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} + (x, \lambda) \begin{pmatrix} Q & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} \\ &= \left(\begin{pmatrix} -\sqrt{\beta} A & \frac{1}{\sqrt{\beta}} I \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} \right)^\top \begin{pmatrix} -\sqrt{\beta} A & \frac{1}{\sqrt{\beta}} I \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} + \|x\|_Q^2 \\ &= \left(\frac{1}{\sqrt{\beta}} \lambda - \sqrt{\beta} A x \right)^\top \left(\frac{1}{\sqrt{\beta}} \lambda - \sqrt{\beta} A x \right) + \|x\|_Q^2 \\ &= \left\| \frac{1}{\sqrt{\beta}} \lambda - \sqrt{\beta} A x \right\|^2 + \|x\|_Q^2 > 0, \quad \forall\, \omega = (x, \lambda)^\top \neq 0, \end{aligned}$$

and hence the matrix H is positive definite. □

The following lemma is also the basis of convergence analysis of the proposed novel penalty dual-primal ALM (1.6).

Lemma 3.3. *Let $\{\omega^k\}$ be the sequence generated by Algorithm* 1.1. *Then, we obtain*

$$\begin{aligned} &\theta(x) - \theta(x^{k+1}) + (\omega - \omega^{k+1})^\top F(\omega) \\ &\geq \frac{1}{2} \big(\|\omega - \omega^{k+1}\|_H^2 - \|\omega - \omega^k\|_H^2 \big) + \frac{1}{2} \|\omega^k - \omega^{k+1}\|_H^2, \quad \forall\, \omega \in \Omega. \end{aligned} \tag{3.5}$$

Proof. It follows from (2.5) that

$$(\omega - \omega^{k+1})^\top F(\omega^{k+1}) = (\omega - \omega^{k+1})^\top F(\omega), \quad \forall\, \omega \in \Omega.$$

Together with (3.1) yields that

$$\omega^{k+1} \in \Omega, \quad \theta(x) - \theta(x^{k+1}) + (\omega - \omega^{k+1})^\top F(\omega) \geq (\omega - \omega^{k+1})^\top H(\omega^k - \omega^{k+1}) \tag{3.6}$$

for all $\omega \in \Omega$. Note that for any $a, b, c, d \in \mathbb{R}^{n+m}$,

$$(a-b)^\top H(c-d) = \frac{1}{2}\{\|a-d\|_H^2 - \|a-c\|_H^2\} + \frac{1}{2}\{\|c-b\|_H^2 - \|d-b\|_H^2\}.$$

Letting $a = \omega$, $b = d = \omega^{k+1}$ and $c = \omega^k$ in the above equality, we obtain

$$(\omega - \omega^{k+1})^\top H(\omega^k - \omega^{k+1}) = \frac{1}{2}\{\|\omega - \omega^{k+1}\|_H^2 - \|\omega - \omega^k\|_H^2\} + \frac{1}{2}\|\omega^k - \omega^{k+1}\|_H^2. \tag{3.7}$$

Consequently, the desired result follows from (3.6) and (3.7). □

The following result shows that $\{\omega^k\}$ is a Fejér convergence with respect to the set Ω^*.

Theorem 3.4. *Let $\{\omega^k\}$ be the sequence generated by Algorithm* 1.1*. Then,*

$$\|\omega^{k+1} - \omega^*\|_H^2 \leq \|\omega^k - \omega^*\|_H^2 - \|\omega^k - \omega^{k+1}\|_H^2, \quad \forall\, \omega^* \in \Omega^*, \tag{3.8}$$

and $\{\omega^k\}$ is Fejér convergent with respect to Ω^*.

Proof. Set $\omega = \omega^* \in \Omega^*$ in (3.5). Then we get

$$\begin{aligned} &\|\omega^k - \omega^*\|_H^2 - \|\omega^{k+1} - \omega^*\|_H^2 - \|\omega^k - \omega^{k+1}\|_H^2 \\ &\geq 2\{\theta(x^{k+1}) - \theta(x^*) + (\omega^{k+1} - \omega^*)^\top F(\omega^*)\}, \quad \forall\, \omega^* \in \Omega^*. \end{aligned} \tag{3.9}$$

Since $\omega^* \in \Omega^*$ and $\omega^{k+1} \in \Omega$, according to (2.3) and (2.4), we have

$$\theta(x^{k+1}) - \theta(x^*) + (\omega^{k+1} - \omega^*)^\top F(\omega^*) \geq 0.$$

Therefore, (3.8) follows from (3.9) immediately, and so,

$$\|\omega^{k+1} - \omega^*\|_H \leq \|\omega^k - \omega^*\|_H. \tag{3.10}$$

This together with Definition 2.2 yields that $\{\omega^k\}$ is Fejér convergent with respect to Ω^*. □

Now, we prove the convergence and the worst-case $O(1/T)$ convergence rate of $\{\omega^k\}$ generated by Algorithm 1.1, where T denotes the total iteration counter.

Theorem 3.5. *Let $\{\omega^k\}$ be the sequence generated by Algorithm* 1.1 *and H be defined in* (3.2)*. Then, the following assertions hold:*

(i) *the sequence $\{\omega^k\}$ converges to some $\overline{\omega} \in \Omega^*$;*

(ii) *for any integer number $T > 0$, we have*

$$\widetilde{\omega}_T \in \Omega, \quad \theta(\widetilde{x}_T) - \theta(x) + (\widetilde{\omega}_T - \omega)^\top F(\omega) \leq \frac{1}{2(T+1)}\|\omega - \omega^0\|_H^2, \quad \forall\, \omega \in \Omega,$$

where $\widetilde{\omega}_T := \frac{1}{T+1}\sum_{k=0}^{T} \omega^{k+1}$.

Proof. (i) From (3.10), it follows that $\|\omega^{k+1}-\omega^*\|_H \le \|\omega^1-\omega^*\|_H$ and $\{\|\omega^{k+1}-\omega^*\|_H\}$ is convergent for $\omega^* \in \Omega^*$ and so, the sequence $\{\omega^k\}$ is bounded. Let $\overline{\omega}$ be a cluster point of $\{\omega^k\}$ and $\{\omega^{k_j}\}$ be a subsequence converging to $\overline{\omega}$. It follows from (3.1) that

$$\omega^{k_j} \in \Omega, \quad \theta(x)-\theta(x^{k_j})+(\omega-\omega^{k_j})^\top F(\omega^{k_j}) \ge (\omega-\omega^{k_j})^\top H(\omega^{k_j-1}-\omega^{k_j}), \quad \forall\, \omega \in \Omega.$$

Again, from (3.8), we have

$$\lim_{k\to\infty} \|\omega^k-\omega^{k+1}\|_H = 0. \tag{3.11}$$

Since the matrix H is positive definite, it follows from (3.11) and the lower semicontinuity of $\theta(\cdot)$ and the continuity of $F(\cdot)$ that

$$\overline{\omega} \in \Omega, \quad \theta(x)-\theta(\overline{x})+(\omega-\overline{\omega})^\top F(\overline{\omega}) \ge 0, \quad \forall\, \omega \in \Omega.$$

So, $\overline{\omega} \in \Omega^*$. Then we deduce from Lemma 2.3 and Theorem 3.4 that $\lim_{k\to\infty}\omega^k = \overline{\omega} \in \Omega^*$.

(ii) The proof is similar to that of [16, Theorem 3.5] and so is omitted. □

4. Splitting version of the penalty dual-primal ALM

In this section, we consider the splitting version of the dual-primal balanced ALM (1.6) for solving the following linearly constrained multi-block separable convex optimization problem

$$\min_x \left\{ \sum_{i=1}^p \theta_i(x_i) \;\Big|\; \sum_{i=1}^p A_i x_i = b, x_i \in \mathscr{X}_i \right\}, \tag{4.1}$$

where $x = (x_i)_{i=1}^p$, $p \ge 1$, $\sum_{i=1}^p n_i = n$, $\theta_i\colon \mathbb{R}^{n_i} \to \mathbb{R}$, $i = 1,\dots,p$, are convex but not necessarily smooth functions, $\mathscr{X}_i \subset \mathbb{R}^{n_i}$, $i = 1,\dots,p$, are nonempty closed convex sets, $A_i \in \mathbb{R}^{m\times n_i}$, $i = 1,\dots,p$, are given matrices and $b \in \mathbb{R}^m$ is a known vector. The model (4.1) has been applied to distributed optimization, statistical learning and Potts models; see, e.g., [4,21,23].

4.1. Splitting version of the penalty dual-primal ALM

In this subsection, we extend the penalty dual-primal ALM to solve the multi-block separable convex optimization problem (4.1) and propose a splitting version of (1.6) below.

Algorithm 4.1 (The splitting penalty dual-primal balanced ALM)**.** *Let $i = 1,\dots,p$, and $x_i \in \mathscr{X}_i$, $\beta_i > 0$, $Q_i \succ 0$ be arbitrarily given positive-defined matrices. Then the new iterate $\omega^{k+1} = (x^{k+1},\lambda^{k+1})$ is generated with $\omega^k = (x^k,\lambda^k)$ via the following steps:*

$$\begin{cases} \lambda^{k+1} = \lambda^k - \sum_{i=1}^p \beta_i \big(\sum_{i=1}^p A_i x_i^k - b \big), \\ x_i^{k+1} = \arg\min_{x_i} \big\{ \theta_i(x_i) - \langle 2\lambda^{k+1}-\lambda^k, A_i x_i - b\rangle + \frac{\beta_i}{2}\|A_i(x_i-x_i^k)\|^2 \\ \qquad\qquad + \frac{1}{2}\|x_i - x_i^k\|_{Q_i}^2 \big\}, \quad i = 1,2,\dots,p. \end{cases} \tag{4.2}$$

4.2. Variational inequality characterization of the splitting version

In order to analyze convergence, we also present the optimality conditions of the model (4.2) in the framework of variational inequality. Firstly, we reuse the letters and set $\Omega := \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_p \times \Lambda$.

Analogous to the analysis in Section 2, we deduce that the optimality conditions of (4.2) can be equivalently characterized by the following mixed variational inequalities

$$\omega^* \in \Omega, \quad \theta(x) - \theta(x^*) + (\omega - \omega^*)^\top F(\omega^*) \geq 0, \quad \forall\, \omega \in \Omega, \tag{4.3}$$

where

$$\theta(x) = \sum_{i=1}^{p} \theta_i(x_i), \quad \omega = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \quad \text{and} \quad F(\omega) = \begin{pmatrix} -A_1^\top \lambda \\ \vdots \\ -A_p^\top \lambda \\ \sum_{i=1}^{p} A_i x_i - b \end{pmatrix}.$$

Similarly, it is easy to verify that the operator $F(\cdot)$ is affine with a skew-symmetric matrix satisfying

$$(\omega - \widetilde{\omega})^\top (F(\omega) - F(\widetilde{\omega})) = 0, \quad \forall\, \omega, \widetilde{\omega} \in \Omega.$$

We also denote by Ω^* the set of solutions of the mixed variational inequalities (4.3).

4.3. Convergence analysis of the splitting version

According to the similar analysis route in Section 3, we next give the essential lemmas and the key proposition below.

Lemma 4.2. *Let $\{\omega^k\}$ be the sequence generated by Algorithm 4.1. Then $\omega^{k+1} \in \Omega$ and*

$$\theta(x) - \theta(x^{k+1}) + (\omega - \omega^{k+1})^\top F(\omega^{k+1}) \geq (\omega - \omega^{k+1})^\top H(\omega^k - \omega^{k+1}), \quad \forall\, \omega \in \Omega,$$

where

$$H = \begin{pmatrix} \beta_1 A_1^\top A_1 + Q_1 & \cdots & \mathbf{0} & -A_1^\top \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \beta_p A_p^\top A_p + Q_p & -A_p^\top \\ -A_1 & \cdots & -A_p & \sum_{i=1}^{p} \frac{1}{\beta_i} I \end{pmatrix}. \tag{4.4}$$

Proof. It follows from Lemma 3.1 by setting $\theta(x) := \sum_{i=1}^{n} \theta_i(x)$ and $A := [A_1 \cdots A_n]$ directly. □

Proposition 4.3. *The matrix H defined by (4.4) is positive definite.*

Proof. The proof is the same as that of Proposition 3.2. □

Similar to Section 3, the following lemmas and theorems on the convergence and convergence rate of the sequence $\{\omega^k\}$ generated by Algorithm 4.1 can be derived based on Lemma 4.2 and Proposition 4.3.

Lemma 4.4. *Let $\{\omega^k\}$ be the sequence generated by Algorithm* 1.1*. Then, we obtain*

$$\begin{aligned}&\theta(x) - \theta(x^{k+1}) + (\omega - \omega^{k+1})^\top F(\omega)\\ \geq\ & \frac{1}{2}\big(\|\omega - \omega^{k+1}\|_H^2 - \|\omega - \omega^k\|_H^2\big) + \frac{1}{2}\|\omega^k - \omega^{k+1}\|_H^2, \quad \forall\, \omega \in \Omega,\\ &\|\omega^{k+1} - \omega^*\|_H^2 \leq \|\omega^k - \omega^*\|_H^2 - \|\omega^k - \omega^{k+1}\|_H^2, \quad \forall\, \omega^* \in \Omega^*,\end{aligned}$$

and $\{\omega^k\}$ is Fejér convergent with respect to Ω^*.

Proof. It follows from Theorem 3.4 by setting $\theta(x) := \sum_{i=1}^n \theta_i(x)$ and $A := [A_1 \cdots A_n]$. □

Theorem 4.5. *Let $\{\omega^k\}$ be the sequence generated by Algorithm* 4.1 *and H be defined in* (4.4)*. Then, the following assertions hold:*

(i) *the sequence $\{\omega^k\}$ converges to some $\overline{\omega} \in \Omega^*$;*

(ii) *for any iterate number $T > 0$, we have*

$$\widetilde{\omega}_T \in \Omega, \quad \theta(\widetilde{x}_T) - \theta(x) + (\widetilde{\omega}_T - \omega)^\top F(\omega) \leq \frac{1}{2(T+1)}\|\omega - \omega^0\|_H^2, \quad \forall\, \omega \in \Omega,$$

where $\widetilde{\omega}_T := \frac{1}{T+1}\sum_{k=0}^T \omega^{k+1}$.

Proof. The proof is same as that of Theorem 3.5 and so is omitted. □

5. Partial proximal strategy of the penalty dual-primal ALM

The penalty dual-primal ALM (1.6) can be generalized to its splitting version (4.2) when the background issue changes from the one-block to the multiple-block case. If the functions θ_i have nice properties such as strong convexity and differentiability for some $i \in \{1, 2, \ldots, p\}$, then the x_i-subproblems do not need to add the proximal regularity terms. So, it is necessary to design the penalty dual-primal ALM (1.6) for solving linearly constrained multiple-block separable minimization problems (4.1) by the partial proximal strategy.

5.1. Partial proximal penalty dual-primal ALM

Without loss of generality, we only add the proximal matrix terms to the former p_1 subproblems, where $1 \leq p_1 \leq p$, and $p_2 = p - p_1$. For succinctness of notations, we adopt the notations in Section 4.

A penalty dual-primal ALM with partial proximal regularization terms is proposed to solve the multiple-block model (4.1).

Algorithm 5.1 (The partial proximal penalty dual-primal ALM)**.** *Let $\beta_i > 0$ and $Q_i \succ 0$ be arbitrarily given positive-defined matrixes, $i = 1, 2, \ldots, p$ and $x_i \in \mathscr{X}_i$. Then the new iterate $\omega^{k+1} = (x^{k+1}, \lambda^{k+1})$ is generated with $\omega^k = (x^k, \lambda^k)$ via the following steps:*

$$\begin{cases} \lambda^{k+1} = \lambda^k - \sum_{i=1}^p \beta_i \big(\sum_{i=1}^p A_i x_i^k - b \big), \\ x_i^{k+1} = \arg\min_{x_i} \big\{ \theta(x_i) - \langle 2\lambda^{k+1} - \lambda^k, A_i x_i - b \rangle + \frac{1}{2} \| x_i - x_i^k \|^2_{\beta_i A_i^\top A_i + Q_i} \big\}, \quad i = 1, \ldots, p_1, \\ x_i^{k+1} = \arg\min_{x_i} \big\{ \theta(x_i) - \langle 2\lambda^{k+1} - \lambda^k, A_i x_i - b \rangle + \frac{\beta_i}{2} \| A_i (x_i - x_i^k) \|^2 \big\}, \quad i = p_1 + 1, \ldots, p. \end{cases}$$

5.2. Convergence analysis of the partial proximal penalty dual-primal ALM

According to the same variational inequality characterization and the same analysis route in Section 4, we next give the essential lemmas and the key proposition below.

Lemma 5.2. *Let $\{\omega^k\}$ be the sequence generated by Algorithm 5.1. Then $\omega^{k+1} \in \Omega$ and*

$$\theta(x) - \theta(x^{k+1}) + (\omega - \omega^{k+1})^\top F(\omega^{k+1}) \geq (\omega - \omega^{k+1})^\top H (\omega^k - \omega^{k+1}), \quad \forall \, \omega \in \Omega,$$

where H is defined by

(5.1)

$$H = \begin{pmatrix} \beta_1 A_1^\top A_1 + Q_1 & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & -A_1^\top \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \beta_{p_1} A_{p_1}^\top A_{p_1} + Q_{p_1} & \mathbf{0} & \cdots & \mathbf{0} & -A_{p_1}^\top \\ \mathbf{0} & \cdots & \mathbf{0} & \beta_{p_1+1} A_{p_1+1}^\top A_{p_1+1} & \cdots & \mathbf{0} & -A_{p_1+1}^\top \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \beta_p A_p^\top A_p & -A_p^\top \\ -A_1 & \cdots & -A_{p_1} & -A_{p_1+1} & \cdots & -A_p & \sum_{i=1}^p \frac{1}{\beta_i} I \end{pmatrix}.$$

Proof. The proof is same as that of Lemma 4.2 and so is omitted. □

Proposition 5.3. *The matrix H defined by* (5.1) *is positive definite.*

Proof. The proof is same as that of Proposition 4.3 and so is omitted. □

Similar to Section 4, we can obtain the convergence and convergence rate of the sequence $\{\omega^k\}$ generated by Algorithm 5.1 based on Lemma 5.2 and Proposition 5.3.

Lemma 5.4. *Let $\{\omega^k\}$ be the sequence generated by Algorithm 5.1. Then, we obtain*

$$\begin{aligned} & \theta(x) - \theta(x^{k+1}) + (\omega - \omega^{k+1})^\top F(\omega) \\ \geq & \frac{1}{2} \big(\|\omega - \omega^{k+1}\|_H^2 - \|\omega - \omega^k\|_H^2 \big) + \frac{1}{2} \|\omega^k - \omega^{k+1}\|_H^2, \quad \forall \, \omega \in \Omega, \\ & \|\omega^{k+1} - \omega^*\|_H^2 \leq \|\omega^k - \omega^*\|_H^2 - \|\omega^k - \omega^{k+1}\|_H^2, \quad \forall \, \omega^* \in \Omega^*, \end{aligned}$$

and $\{\omega^k\}$ is Fejér convergent with respect to Ω^*.

Proof. The proof is same as that of Lemma 3.3 and Theorem 3.4 and so is omitted. □

Theorem 5.5. *Let $\{\omega^k\}$ be the sequence generated by Algorithm 5.1 and H be defined by (5.1). Then, the following assertions hold:*

(i) *the sequence $\{\omega^k\}$ converges to some $\overline{\omega} \in \Omega^*$;*

(ii) *for any iterate number $T > 0$, we have*

$$\widetilde{\omega}_T \in \Omega, \quad \theta(\widetilde{x}_T) - \theta(x) + (\widetilde{\omega}_T - \omega)^\top F(\omega) \leq \frac{1}{2(T+1)} \|\omega - \omega^0\|_H^2, \quad \forall\, \omega \in \Omega,$$

where $\widetilde{\omega}_T := \frac{1}{T+1} \sum_{k=0}^{T} \omega^{k+1}$.

Proof. The proof is same as that of Theorem 3.5 and so is omitted. □

Remark 5.6. Compared with the balanced ALM (1.3) and the dual-primal balanced ALM (1.5) for linearly constrained one-block convex optimization problem, Algorithms 1.1, 4.1 and 5.1 do not need to calculate the inverse of the matrix $\frac{1}{\beta}AA^\top + \delta I_m$ and so, the updating time of λ^{k+1} in Algorithms 1.1, 4.1 and 5.1 is less than that of the balanced ALM (1.3) and the dual-primal balanced ALM (1.5). Secondly, Algorithms 1.1, 4.1 and 5.1 are the primal-dual ALM with penalty term which is distinct with the dual-primal balanced ALM (1.5). Besides, Algorithms 4.1 and 5.1 are applicable to the linearly constrained multi-block convex optimization problems. The difference between Algorithm 1.1 and the penalty ALM (1.4) is the calculation order of x^k and λ^k. The numerical results reported in Section 6 show that Algorithm 1.1 is slightly better than the balanced ALM (1.3), the penalty ALM (1.4) and the dual-primal balanced ALM (1.5) given in [1, 16, 22], respectively.

6. Numerical experiments

In this section, LASSO and the basic pursuit problem, which are extensively applied to image processing, statistical learning, compress sensing and machine learning, are solved by the proposed algorithms. All code are written in Matlab and all experiments are performed in Matlab R2015b on a workstation with an Intel(R) Core(TM) i7-8550U CPU(1.80GHz) and 8GB RAM.

We firstly apply the penalty dual-primal balanced ALM (1.6) to solve the basic pursuit problem, which is an equality constrained ℓ_1-norm minimization problem, and compare it with the primal-dual balanced ALM proposed in [22] and the balanced ALM proposed in [16].

Example 6.1. The basic pursuit problem (BPP) is given as follows:

$$\min_x \left\{ \|x\|_1 \mid Ax = b, x \in \mathbb{R}^n \right\}, \tag{6.1}$$

where $\|x\|_1 = \sum_{i=1}^n |x_i|$ is the ℓ_1-norm of a vector, $A \in \mathbb{R}^{m\times n}$ $(m < n)$ is a given matrix and $b \in \mathbb{R}^m$ is a given vector.

For the (BPP), the iterative scheme of the proposed method (1.6) reads as follows:

$$\begin{cases} \lambda^{k+1} = \lambda^k - \beta(Ax^k - b), \\ x^{k+1} = \arg\min_x \left\{ \|x\|_1 - \langle 2\lambda^{k+1} - \lambda^k, Ax - b\rangle + \frac{1}{2}\|x - x^k\|^2_{\beta A^\top A + Q} \right\}. \end{cases}$$

In particular, let $Q = \tau I - \beta A^\top A$ where $\tau > \beta\|A^\top A\|$, the iterate scheme above could be converted to

$$\begin{cases} \lambda^{k+1} = \lambda^k - \beta(Ax^k - b), \\ x^{k+1} = \arg\min_x \left\{ \|x\|_1 + \frac{\tau}{2}\left\|x - x^k - \frac{1}{\tau}A^\top(2\lambda^{k+1} - \lambda^k)\right\|^2 \right\}. \end{cases} \tag{6.2}$$

Then the solution of the x-subproblem in (6.2) is given by the following explicit form

$$x^{k+1} = S_{1/\tau}\left[x^k + \frac{1}{\tau}A^\top(2\lambda^{k+1} - \lambda^k)\right],$$

where $S_\delta(t)$ is the soft thresholding operator [9] defined by

$$(S_\delta(t))_i := (1 - \delta/|t_i|)_+ \cdot t_i, \quad i = 1, 2, \ldots, m. \tag{6.3}$$

Following the same rules, the iterative scheme of the dual-primal balanced ALM [22] for solving (6.1) reads as follows:

$$\text{(DP-ALM)} \quad \begin{cases} \lambda^{k+1} = \lambda^k - \left(\frac{1}{\tau}AA^\top + \delta I_m\right)^{-1}(Ax^k - b), \\ x^{k+1} = \arg\min_x \left\{ \|x\|_1 + \frac{\tau}{2}\left\|x - x^k - \frac{1}{\tau}A^\top(2\lambda^{k+1} - \lambda^k)\right\|^2 \right\}, \end{cases}$$

and the iterative scheme of the balanced ALM [16] for solving (6.1) reads as follows:

$$\text{(Balanced ALM)} \quad \begin{cases} x^{k+1} = \arg\min_x \left\{ \|x\|_1 + \frac{\tau}{2}\left\|x - x^k - \frac{1}{\tau}A^\top\lambda^k\right\|^2 \right\}, \\ \lambda^{k+1} = \lambda^k - \left(\frac{1}{\tau}AA^\top + \delta I_m\right)^{-1}\left(A(2x^{k+1} - x^k) - b\right). \end{cases}$$

We generate the data by the same way as in [22]: the matrix A is generated from independently normal distribution $\mathcal{N}(0, 1)$; for all tested algorithm the initial point (x^0, λ^0) is randomly generated, and we take the following toned values of parameters for the mentioned experiments:

(1) the penalty dual-primal balanced ALM (PDP-ALM): $\beta := 0.001$ and $\tau = 2.5$;

(2) the dual-primal balanced ALM (DP-ALM) and the balanced ALM (B-ALM): $\delta = 1000$ and $\tau = 2.5$.

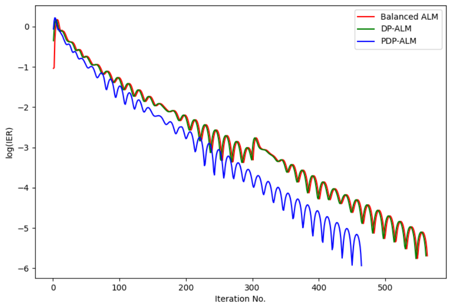
The termination criteria is defined by

$$R(k) = \max\left\{\|x^{k+1} - x^k\|, \|\lambda^{k+1} - \lambda^k\|\right\} < 10^{-7}.$$

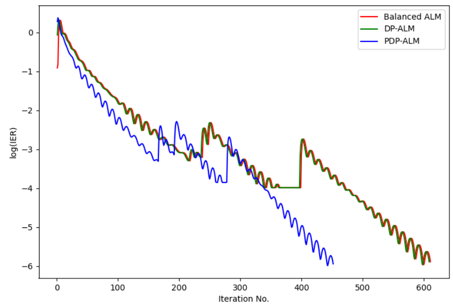
Table 6.1 lists the number of iterations and runtime in seconds respectively of the PDP-ALM, the DP-ALM and the B-ALM for solving the basic pursuit problem with different dimension $m \times n$ of A, and $CR = \|Ax - b\|^2$ stands for constrained residual. From the numerical experimental result, it is clear that the PDP-ALM has much better numerical performs than the DP-ALM and B-ALM both in the number of iterations and runtime. To further visualize the numerical results, we also plot the convergence curves versus iteration numbers of some representative examples in Figure 6.1, which also shows that the proposed method has a better performance than DP-ALM and B-ALM in the number of iterations.

Table 6.1: The number of iterations and runtime of (PDP-ALM), (DP-ALM) and (B-ALM) for solving (BPP).

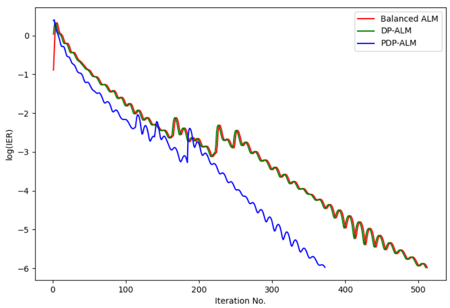
	PDP-ALM			DP-ALM			B-ALM		
$m \times n$	Iter.	Time	CR	Iter.	Time	CR	Iter.	Time	CR
300×500	465	0.11	2.86e-4	562	0.16	3.64e-4	564	0.18	3.64e-4
400×600	503	0.15	1.49e-4	669	0.28	1.82e-4	671	0.28	1.82e-4
450×750	453	0.18	6.60e-5	610	0.39	1.01e-4	612	0.35	1.01e-4
500×900	373	0.20	4.63e-5	511	0.39	5.98e-5	513	0.39	5.98e-5
500×1000	1294	0.73	7.25e-5	1969	1.50	2.29e-4	1971	1.47	2.29e-4
600×1150	843	0.70	1.77e-4	1263	1.40	1.65e-4	1266	1.34	1.70e-4
700×1300	672	0.75	8.90e-5	1073	1.51	1.15e-4	1074	1.50	1.15e-4
800×1450	455	0.65	4.63e-5	704	1.26	1.53e-4	706	1.31	1.48e-4
900×1600	547	1.03	5.99e-5	943	2.21	8.24e-5	945	2.27	8.27e-5
1000×1750	876	1.95	4.32e-5	1523	4.14	6.37e-5	1525	4.22	6.40e-5
1100×1900	725	1.83	4.78e-5	1179	3.81	8.53e-5	1181	3.89	8.58e-5
1100×2000	694	1.85	4.47e-5	1263	4.24	9.12e-5	1265	4.30	9.13e-5
1200×2150	406	1.31	3.85e-5	602	3.65	5.38e-5	896	3.88	1.10e-4
1300×2300	420	1.52	6.66e-5	762	3.53	6.27e-5	760	3.58	7.01e-5



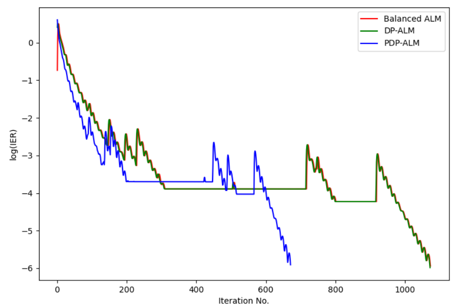
(a) $m \times n = 300 \times 500$



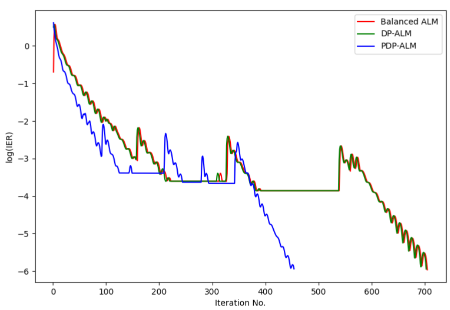
(b) $m \times n = 450 \times 750$



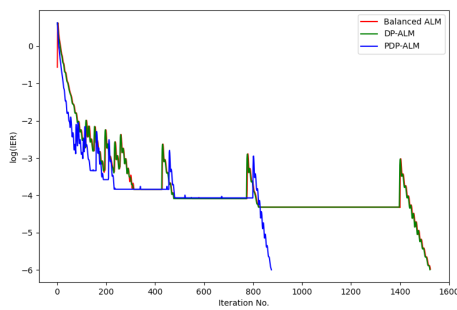
(c) $m \times n = 500 \times 900$



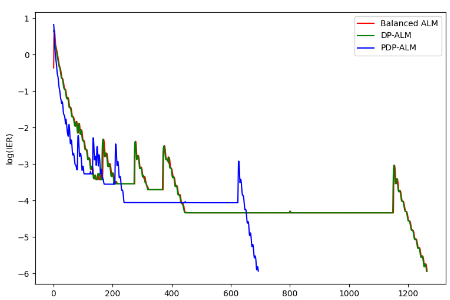
(d) $m \times n = 700 \times 1300$



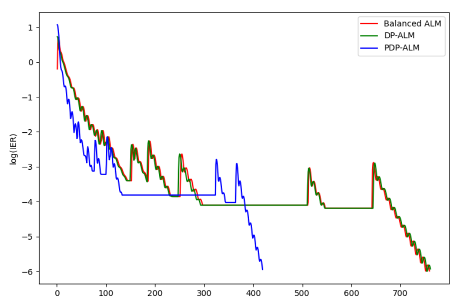
(e) $m \times n = 800 \times 1450$



(f) $m \times n = 1000 \times 1750$



(g) $m \times n = 1100 \times 2000$



(h) $m \times n = 1300 \times 2300$

Figure 6.1: Convergence curves of the PDP-ALM, DP-ALM and the balanced ALM compared with iteration number under various dimension of A.

We now apply the splitting penalty dual-primal balanced ALM (4.2) to solve the well-known LASSO model and compare it with the linearized ADMM with proximal method (shortly, PL-ADMM) [1] and the positive-indefinite proximal ADMM (shortly, PIPL-ADMM) [9]. The PL-ADMM [1] is also an extension of penalty ALM.

Example 6.2. The LASSO model is formulated as follows:

$$\min_y \frac{1}{2}\|Ay-b\|^2+\sigma\|y\|_1, \tag{6.4}$$

where $\|y\|_1 := \sum_{i=1}^n |y_i|$, $A \in \mathbb{R}^{m\times n}$ is a design matrix usually with $m \ll n$, m is the number of date point, n is the number of features, $b \in \mathbb{R}^m$ is the response vector and $\sigma > 0$ is a regularization parameter. By a new auxiliary variable x, (6.4) can be rewritten as the form

$$\min_{x,y}\left\{\frac{1}{2}\|x-b\|^2+\sigma\|y\|_1 \;\middle|\; x-Ay=0, x\in\mathbb{R}^m, y\in\mathbb{R}^n\right\}. \tag{6.5}$$

The problem (6.5) is a special case of (4.1). Then iterative scheme of the splitting penalty dual-primal balanced ALM (4.2) for solving (6.5) reads as follows:

$$\begin{cases} \lambda^{k+1} = \lambda^k - \beta_1(x-Ay)-\beta_2(x-Ay), \\ x^{k+1} = \arg\min_x \left\{\frac{1}{2}\|x-b\|^2 - \langle 2\lambda^{k+1}-\lambda^k, x-b\rangle + \frac{1}{2}\|x-x^k\|^2_{\beta_1 I+Q_1}\right\}, \\ y^{k+1} = \arg\min_y \left\{\sigma\|y\|_1 - \langle 2\lambda^{k+1}-\lambda^k, -Ay-b\rangle + \frac{1}{2}\|y-y^k\|^2_{\beta_2 A^\top A+Q_2}\right\}. \end{cases} \tag{6.6}$$

In particular, let $Q_1 = \tau_1 I - \beta_1 I$ with $\tau_1 > \beta_1\|I\|$, $Q_2 = \tau_2 I - \beta_2 A^\top A$ with $\tau_2 > \beta_2\|A^\top A\|$, the iterate (6.6) could be converted to

$$\begin{cases} \lambda^{k+1} = \lambda^k - \beta_1(x-Ay)-\beta_2(x-Ay), \\ x^{k+1} = \arg\min_x \left\{\frac{1}{2}\|x-b\|^2 + \frac{\tau_1}{2}\left\|x-x^k-\frac{1}{\tau_1}(2\lambda^{k+1}-\lambda^k)\right\|^2\right\}, \\ y^{k+1} = \arg\min_y \left\{\sigma\|y\|_1 + \frac{\tau_2}{2}\left\|y-y^k-\frac{1}{\tau_2}A^\top(2\lambda^{k+1}-\lambda^k)\right\|^2\right\}. \end{cases}$$

Moreover, $(\lambda^{k+1}, x^{k+1}, y^{k+1})$ has the following explicit form

$$\begin{cases} \lambda^{k+1} = \lambda^k - \beta_1(x-Ay)-\beta_2(x-Ay), \\ x^{k+1} = \frac{1}{\tau_1}\left[\tau_1 x^k + (2\lambda^{k+1}-\lambda^k)\right] \\ y^{k+1} = S_{\sigma/\tau_2}\left[y^k + \frac{1}{\tau_2}A^\top(2\lambda^{k+1}-\lambda^k)\right], \end{cases}$$

where $S_\delta(t)$ is the soft threshold operator defined by (6.3).

We generate the data by the same way as in [9]: we first choose $A_{ij} \sim \mathcal{N}(0,1)$ and then scaled the columns to have unit norm. We use the script 'sprandn' to generate a sparse vector y^* which have approximately density $= 100/n$ non-zeros entries taken from the normal distribution with zero mean and unit variance. We generate b via $b := Ay^* + e$, where e is a small white noise taken from $e \sim \mathscr{N}(0, 10^{-3}I)$. We choose the dimension of A is 1050×3500. We set the regularization parameter σ to 0.1, and for all tested algorithm the initial point (x^0, y^0, λ^0) is randomly generated, and we take the following toned values of parameters for the mentioned experiments:

(1) The splitting penalty dual-primal balanced ALM (PDP-ALM): $\beta_1 = \beta_2 := \frac{2-\tau_2}{\tau_2|\tau_2-1|}$ and $\tau_1 := \frac{|\tau_2-1|}{5\beta_1\tau_2} + \frac{4}{5}$;

(2) PL-ADMM [1] and PIPL-ADMM [9]: $\beta = \frac{2-\tau_2}{\tau_2|\tau_2-1|}$ and $\tau := \frac{|\tau_2-1|}{5\beta_1\tau_2} + \frac{4}{5}$,

and we all set $\tau_2 = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7$.

The termination criteria is defined by

$$\max\left\{\|x^{k+1} - x^k\|, \|y^{k+1} - y^k\|, \|\lambda^{k+1} - \lambda^k\|\right\} < 10^{-10}.$$

Table 6.2: The number of iterations and runtime of PL-ADMM, PL-ADMM and PDP-ALM for solving LASSO model.

	PL-ADMM		PIPL-ADMM		PDP-ALM					
τ_2	Iter.1	Time.1	Iter.2	Time.2	Iter.3	Time.3	$\frac{\text{Iter.3}}{\text{Iter.1}}$	$\frac{\text{Time.3}}{\text{Time.1}}$	$\frac{\text{Iter.3}}{\text{Iter.2}}$	$\frac{\text{Time.3}}{\text{Time.2}}$
0.05	376	29.35	204	16.16	101	3.53	0.27	0.12	0.50	0.22
0.10	380	29.63	417	33.98	111	3.92	0.29	0.13	0.27	0.12
0.15	384	30.13	397	31.66	118	4.15	0.31	0.14	0.30	0.13
0.20	399	32.97	410	34.58	123	4.34	0.31	0.13	0.30	0.13
0.25	413	32.18	409	33.57	127	4.54	0.31	0.14	0.31	0.14
0.30	409	31.92	406	34.08	130	4.56	0.32	0.14	0.32	0.13
0.35	420	33.02	404	32.64	132	4.65	0.31	0.14	0.33	0.14
0.40	420	32.87	397	32.32	134	4.76	0.32	0.14	0.34	0.15
0.45	431	33.82	392	31.34	136	4.78	0.32	0.14	0.35	0.15
0.50	438	34.67	392	35.86	137	4.83	0.31	0.14	0.35	0.13
0.55	443	34.69	395	31.34	138	4.85	0.31	0.14	0.35	0.15
0.60	452	35.34	395	31.78	138	4.87	0.31	0.14	0.35	0.15
0.65	407	23.41	394	31.38	138	3.53	0.34	0.15	0.35	0.11
0.70	413	23.74	400	32.67	137	3.53	0.33	0.15	0.34	0.11

Table 6.2 lists the number of iterations and runtime in seconds respectively of PL-ADMM, PL-ADMM and PDP-ALM for solving the LASSO model with different param-

eter τ_2. To further visualize the numerical results, we also plot the iterations results in terms of the various parameters τ_2 in Figure 6.2. From Table 6.2 and Figure 6.2, one can see that PDP-ALM has much better performs than PL-ADMM, PL-ADMM both in the number of iterations and runtime.

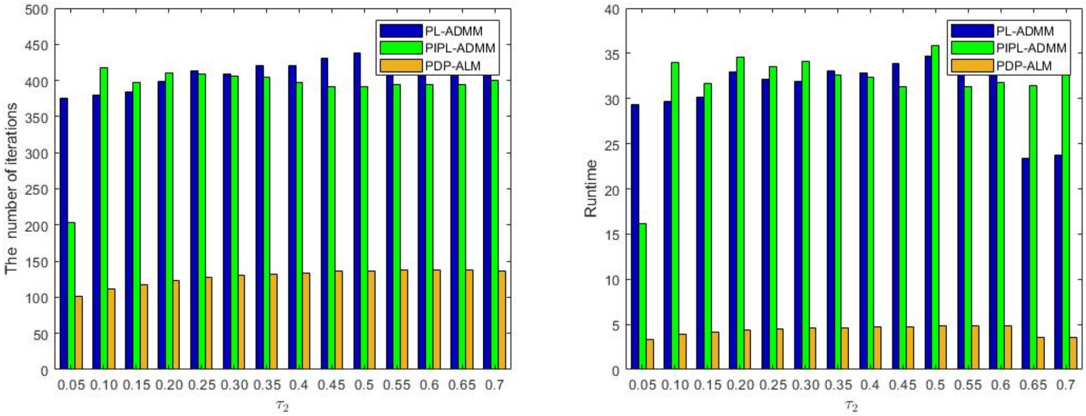


Figure 6.2: The number of iterations and average runtime of PDP-ALM, PL-ADMM and PIPL-ADMM with the different parameter τ_2.

7. Conclusions

A new penalty dual-primal augmented Lagrangian method for solving linearly constrained convex minimization problems is introduced based on the balanced technique [16]. Further, two extensions of the penalty dual-primal augmented Lagrangian method are proposed to solve the linearly constrained multiple-block separable convex minimization problems. The global convergence and sub-linear convergence rate of the proposed methods are established by using the tool of variational inequality. Numerical tests on the basic pursuit problem and the Lasso model are reported to show the efficiency of the proposed methods. In the future research, it is interesting to study the convergence of the penalty dual-primal augmented Lagrangian method when the involved problems are infeasible from the least violation; see [8]. Besides, it is also interesting to design dynamical systems (the so-called continuous time algorithms) for linearly constrained optimization problems and variational inequalities by the balanced technique.

Acknowledgments

The authors would like to express their sincere thanks to the editor and referees for the valuable comments and helpful suggestions, which help to improve the paper. They also would like to express their sincere thanks to Prof. Bingsheng He and Prof. Xiaoming Yuan [16] for their contributions in ALM and its variants as well as balanced technique, which inspires the paper. This research was partially supported by the Natural Science Foundation of China (12071379, 12361062, 62366001), the Natural Science Foundation of Chongqing (cstc2021jcyj-msxmX0925, cstc2022ycjh-bgzxm0097), and Natural Science Foundation of Ningxia Provincial of China (2023AAC02053), the Youth Project of Science and Technology Research Program of Chongqing Education Commission of China (KJQN202201802) and the Major Project of Science and Technology Research Program of Chongqing Education Commission of China.

References

[1] J. Bai, L. Jia and Z. Peng, *A new insight on augmented Lagrangian method with applications in machine learning*, J. Sci. Comput. **99** (2024), no. 2, Paper No. 53, 33 pp.

[2] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Computer Science and Applied Mathematics, Academic Press, New York, 1982.

[3] E. G. Birgin and J. M. Martínez, *Practical Augmented Lagrangian Methods for Constrained Optimization*, Fundamentals of Algorithms **10**, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2014.

[4] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn. **3** (2011), no. 1, 1–122.

[5] R. Burachik, L. M. G. Drummond, A. N. Iusem and B. F. Svaiter, *Full convergence of the steepest descent method with inexact line searches*, Optimization **32** (1995), no. 2, 137–146.

[6] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision **40** (2011), no. 1, 120–145.

[7] ———, *An introduction to continuous optimization for imaging*, Acta Numer. **25** (2016), 161–319.

[8] J. Chen and Y.-H. Dai, *Multiobjective optimization with least constraint violation: Optimality conditions and exact penalization*, J. Global Optim. **87** (2023), no. 2-4, 807–830.

[9] J. Chen, Y. Wang, H. He and Y. Lv, *Convergence analysis of positive-indefinite proximal ADMM with a Glowinski's relaxation factor*, Numer. Algorithms **83** (2020), no. 4, 1415–1440.

[10] M. Fortin and R. Glowinski, *Augmented Lagrangian Methods: Applications to the numerical solution of boundary Value Problems*, Studies in Mathematics and its Applications **15**, North-Holland, Amsterdam, 1983.

[11] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-splitting Methods in Nonlinear Mechanics*, SIAM Studies in Applied Mathematics **9**, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1989.

[12] G. Gu, B. He and X. Yuan, *Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach*, Comput. Optim. Appl. **59** (2014), no. 1-2, 135–161.

[13] B. He, F. Ma and X. Yuan, *Optimal proximal augmented Lagrangian method and its application to full Jacobian splitting for multi-block separable convex minimization problems*, IMA J. Numer. Anal. **40** (2020), no. 2, 1188–1216.

[14] B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective*, SIAM J. Imaging Sci. **5** (2012), no. 1, 119–149.

[15] ———, *On the $O(1/n)$ convergence rate of the Douglas–Rachford alternating direction method*, SIAM J. Numer. Anal. **50** (2012), no. 2, 700–709.

[16] ———, *Balanced augmented Lagrangian method for convex programming*, arXiv:2108.08554.

[17] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appl. **4** (1969), 303–320.

[18] K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, Advances in Design and Control **15**, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008.

[19] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, in: *Optimization (Sympos., Univ. Keele, Keele, 1968)*, 283–298, Academic Press, London, 1969.

[20] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res. **1** (1976), no. 2, 97–116.

[21] H. Sun, X.-C. Tai and J. Yuan, *Efficient and convergent preconditioned ADMM for the Potts models*, SIAM J. Sci. Comput. **43** (2021), no. 2, B455–B478.

[22] S. Xu, *A dual-primal balanced augmented Lagrangian method for linearly constrained convex programming*, J. Appl. Math. Comput. **69** (2023), no. 1, 1015–1035.

[23] J. Yuan, E. Bae, X.-C. Tai and Y. Boykov, *A continuous max-flow approach to Potts model*, in: *Computer Vision–ECCV 2010*, 379–392, Lecture Notes in Computer Science **6316**, Springer, Berlin, 2010.

Xiaoqing Ou
School of Mathematics and Information Science, North Minzu University, Yinchuan 750021, China
and
College of Management, Chongqing College of Humanities, Science & Technology, Chongqing 401524, China
E-mail address: ouxiaoqing413@163.com

Guolin Yu
School of Mathematics and Information Science, North Minzu University, Yinchuan 750021, China
E-mail address: guolin_yu@126.com

Jie Liu, Jiawei Chen and Zhaohan Liu
School of Mathematics and Statistics, Southwest University, Chongqing 400715, China
E-mail addresses: liujie66j@163.com, j.w.chen713@163.com, liuzhaohan6799@163.com