Proximal-Based Pre-correction Decomposition Methods for Structured Convex Minimization Problems

Yuan-Yuan Huang · San-Yang Liu

Received: 10 June 2012 / Revised: 4 February 2013 / Accepted: 20 December 2013 /
Published online: 7 May 2014
© Operations Research Society of China, Periodicals Agency of Shanghai University,
and Springer-Verlag Berlin Heidelberg 2014

Abstract This paper presents two proximal-based pre-correction decomposition methods for convex minimization problems with separable structures. The methods, derived from Chen and Teboulle's proximal-based decomposition method and He's parallel splitting augmented Lagrangian method, remain the nice convergence property of the proximal point method and could compute variables in parallel like He's method under the prediction-correction framework. Convergence results are established without additional assumptions. And the efficiency of the proposed methods is illustrated by some preliminary numerical experiments.

Keywords Structured convex programming · Parallel splitting · Proximal point method · Augmented Lagrangian · Prediction-correction method

Mathematics Subject Classification (GB/T13745-92) 110.74

1 Introduction

In this paper, we consider the following convex minimization program with separable structure of the form:

$$\begin{array}{ll} \min & f(x) + g(y) \\ \text{s.t.} & Ax = y, \end{array} \quad (1.1)$$

This work was supported by the National Natural Science Foundation of China (No. 61373174).

Y.-Y. Huang (✉) · S.-Y. Liu
School of Mathematics and Statistics, Xidian University, Xi'an 710071, China
e-mail: yyuanhuang@126.com

S.-Y. Liu
e-mail: liusanyang@126.com

where $f: \mathbb{R}^n \rightarrow (-\infty, +\infty]$ and $g: \mathbb{R}^m \rightarrow (-\infty, +\infty]$ are given closed proper convex functions, and $A \in \mathbb{R}^{m \times n}$ is a given matrix.

The Lagrangian for problem (1.1) is defined as

$$L(x, y, \lambda) = f(x) + g(y) + \langle \lambda, Ax - y \rangle,$$

and the dual program [11] of (1.1) is

$$\max_{\lambda} \{ \inf_{x, y} L(x, y, \lambda) \}.$$

(x^*, y^*) and λ^* are optimal solutions for problem (1.1) and its dual, respectively, if and only if the triple (x^*, y^*, λ^*) is a saddle point of the Lagrangian function $L(x, y, \lambda)$, i.e., $L(x^*, y^*, \lambda) \leq L(x^*, y^*, \lambda^*) \leq L(x, y, \lambda^*)$, for all $(x, y) \in \mathbb{R}^{n+m}$ and $\lambda \in \mathbb{R}^m$. Hence, based on the augmented Lagrange

$$L_{\alpha}(x, y, \lambda) = f(x) + g(y) + \langle \lambda, Ax - y \rangle + \frac{\alpha}{2} \|Ax - y\|^2, \quad \alpha > 0, \quad (1.2)$$

the augmented Lagrangian methods and the alternating direction methods for the convex programs in the form of problem (1.1) have been proposed, see e.g., [2–9] and references cited therein.

With the k -th iteration $w^k = (x^k, y^k, \lambda^k)$, $k > 0$, the augmented Lagrangian methods are applied to solve the new iteration (x^{k+1}, y^{k+1}) with λ^k fixed, then to generate λ^{k+1} . The good feature to do so is that the new iteration (x^{k+1}, y^{k+1}) can be generated at the same time, but the existence of $\|Ax - y\|^2$ in (1.2) takes some troubles for computing x^{k+1} and y^{k+1} simultaneously. The alternating direction method [3–5] overcomes the drawback of the augmented Lagrangian methods, but leads the fact that the new iterations x^{k+1} and y^{k+1} can not be obtained in parallel and do not fully exploit the desired separable property of f and g . When the proximal point algorithm of multipliers presented by Rockafellar [10] is applied to solve problem (1.1), the nice convergence property of proximal point algorithms is remained, while it has the same drawback with the augmented Lagrangian methods. Thereafter, Chen and Teboulle [1] presented a proximal-based decomposition method which generates (x^{k+1}, y^{k+1}) simultaneously via implementing proximal step before computing λ^{k+1} . It, however, is not one of pre-correction type methods though it has the nice convergence properties of proximal-type methods. The parallel splitting augmented Lagrangian method proposed in [7] is just implemented in prediction-correction framework. It generates the predictor $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ via certain criterions, then the new iteration $(x^{k+1}, y^{k+1}, \lambda^{k+1})$ via direction step. It satisfies the requirement of variables (x, y) 's parallel implementation and remains the nice feature of the prediction-correction methods. But the absence of the proximal terms in sub-problems (see 2.1 and 2.2) does not make the parallel splitting augmented Lagrangian method have proximal-type methods' nice convergence property.

One question is whether we can design improved versions of the methods in [1] and [7]. The answer, of course, is in the affirmative. We call the presenting methods

proximal-based pre-correction decomposition methods. At each iteration, the new iterating is generated by a direction step after the implementation of a proximal step.

The rest of this paper is organized as follows. In Sect. 2, the motivation to design the methods, and the descriptions of them are given. In Sect. 3, the convergence results are given under the same mild assumptions used in the proximal-based decomposition method. In Sect. 4, we verify the efficiency of the proposing methods by some preliminary numerical experiments. Finally, some conclusions are made in Sect. 5.

For convenience, we make an assumption that the primal-dual solution set $\mathcal{W}^* \in \mathbb{R}^{n+2m}$ of the problem (1.1) and its dual are nonempty. We denote $w = (x, y, \lambda) \in \mathbb{R}^{n+2m}$ and $\|w\|^2 = \|x\|^2 + \|y\|^2 + \|\lambda\|^2$.

Notation For any $x, y \in \mathbb{R}^n$, the usual inner product is defined as $\langle x, y \rangle = x^T y$, and the induced Euclidean norm is $\|x\| = \sqrt{\langle x, x \rangle}$. The M -norm is denoted by $\|x\|_M = \sqrt{\langle x, Mx \rangle}$ with $M \in \mathbb{R}^{n \times n}$ as a positive definite matrix (denoted by $M \succeq 0$). Obviously, it becomes the norm $\|x\|$ when M is the identity matrix. Let $T : \mathbb{R}^n \rightarrow 2^{\mathbb{R}^n}$ be a set-valued mapping, and it is monotone if $\langle x - y, u - v \rangle \geq 0$, for all $x, y \in \mathbb{R}^n, u \in T(x)$ and $v \in T(y)$. It is maximal monotone if it also satisfies the maximality, i.e., for any monotone operator $T' : \mathbb{R}^n \rightarrow 2^{\mathbb{R}^n}, T = T'$ if T 's graph $\{(x, y) \mid y \in T(x), \forall x \in \mathbb{R}^n\} := \text{Gph}(T) \subseteq \text{Gph}(T')$.

2 Motivation and Algorithms

With the given iteration $w^k = (x^k, y^k, \lambda^k), k > 0$, the parallel splitting augmented Lagrangian method (PSALM), proposed in [7] for solving variational inequality with separable structure, can be applied to solve problem (1.1) as follows:

firstly, the predictor $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ is got through the following sub-problems:

$$\tilde{x}^k = \arg \min_x \left\{ f(x) - \langle \lambda^k, Ax - y^k \rangle + \frac{1}{2} \|Ax - y^k\|_H^2 \right\}, \quad (2.1)$$

$$\begin{aligned} \tilde{y}^k &= \arg \min_y \left\{ g(y) - \langle \lambda^k, Ax^k - y \rangle + \frac{1}{2} \|Ax^k - y\|_H^2 \right\}, \text{ and} \\ \tilde{\lambda}^k &= \lambda^k - H(A\tilde{x}^k - \tilde{y}^k), \end{aligned} \quad (2.2)$$

secondly, the new iteration $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ is obtained via correction step:

$$w^{k+1} = w^k - \zeta_k G^{-1} M(w^k - \tilde{w}^k),$$

where $\zeta_k = \gamma \frac{\|w^k - \tilde{w}^k\|_M^2 + \langle \lambda^k - \tilde{\lambda}^k, (Ax^k - A\tilde{x}^k) - (y^k - \tilde{y}^k) \rangle}{\|G^{-1}M(w^k - \tilde{w}^k)\|_G^2}, \gamma \in [1, 2), G \in \mathbb{R}^{(n+2m) \times (n+2m)}$ is a symmetric and positive definite matrix, and

$$M = \begin{pmatrix} A^T H A & & \\ & H & \\ & & H^{-1} \end{pmatrix}$$

Obviously, the PSLAM can solve x -variables and y -variables in parallel and remains the advantage of the separation of f and g . Moreover, it combines the nice features of the prediction-correction methods.

The proximal-based decomposition method (PBDM) proposed by Chen and Teboulle [1] generates the new iteration by solving the following sub-problems:

$$\begin{aligned} x^{k+1} &= \arg \min_x \left\{ f(x) + \langle \lambda^k + \alpha_k (Ax^k - y^k), Ax \rangle + \frac{1}{2\alpha_k} \|x - x^k\|^2 \right\}, \\ y^{k+1} &= \arg \min_y \left\{ g(y) - \langle \lambda^k + \alpha_k (Ax^k - y^k), y \rangle + \frac{1}{2\alpha_k} \|y - y^k\|^2 \right\}, \text{ and} \\ \lambda^{k+1} &= \lambda^k + \alpha_k (Ax^{k+1} - y^{k+1}). \end{aligned}$$

It is clear that the PBDM solves problem (1.1) based on the proximal point method. In this way, the good convergence property of the proximal point algorithm can be used preferably. Hence, combining it with the PSLAM, we present proximal-based pre-correction decomposition methods.

Algorithm 2.1 Step 0: Choose $w^0 = (x^0, y^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$. Let $\{H_k \in \mathbb{R}^{n \times m}\}$, $\{R_k \in \mathbb{R}^{n \times n}\}$, $\{S_k \in \mathbb{R}^{m \times m}\}$ and $\{T_k \in \mathbb{R}^{m \times m}\}$ ($k = 0, 1, 2, \dots$) be symmetric and positive definite matrix sequences, $\gamma \in (0, 2)$, $\varepsilon > 0$, $k := 0$.

Step 1: (*proximal step*) Compute $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ by

$$\tilde{x}^k = \arg \min_x \left\{ f(x) + \langle \lambda^k, Ax - y^k \rangle + \frac{1}{2} \|Ax - y^k\|_{H_k}^2 + \frac{1}{2} \|x - x^k\|_{R_k}^2 \right\}, \quad (2.3)$$

$$\tilde{y}^k = \arg \min_y \left\{ g(y) + \langle \lambda^k, Ax^k - y \rangle + \frac{1}{2} \|Ax^k - y\|_{H_k}^2 + \frac{1}{2} \|y - y^k\|_{S_k}^2 \right\}, \quad (2.4)$$

$$\tilde{\lambda}^k = \arg \min_{\lambda} \left\{ -\langle \lambda, A\tilde{x}^k - \tilde{y}^k \rangle + \frac{1}{2} \|\lambda - \lambda^k\|_{T_k}^2 \right\}. \quad (2.5)$$

If $\max\{\|A(\tilde{x}^k - x^k)\|_{\infty}, \|\tilde{y}^k - y^k\|_{\infty}, \|\tilde{\lambda}^k - \lambda^k\|_{\infty}\} < \varepsilon$, stop, otherwise, continue.

Step 2: (*direction step*) The new iteration $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ is generated by

$$w^{k+1} = w^k - \zeta_k M_k (w^k - \tilde{w}^k), \quad (2.6)$$

with $M_k = N_k + P_k$, $\zeta_k = \gamma(\|w^k - \tilde{w}^k\|_{P_k}^2 + \frac{1}{2} \|A\tilde{x}^k - \tilde{y}^k\|_{H_k}^2 + \frac{1}{2} \|Ax^k - y^k\|_{H_k}^2) / \|M_k (w^k - \tilde{w}^k)\|^2$,

$$N_k = \begin{pmatrix} A^T H A & & \\ & H_k & \\ & & T_k \end{pmatrix}, \quad P_k = \begin{pmatrix} R_k & & \\ & S_k & \\ & & 0 \end{pmatrix}$$

$k:=k+1$, goto Step1.

The main difference between Algorithm 2.1 and the PSALM is that the sub-problems (2.3) and (2.4) include the proximal terms $\frac{1}{2}\|x - x^k\|_{R_k}^2$ and $\frac{1}{2}\|y - y^k\|_{S_k}^2$, respectively. If f is non-differentiable, the sub-problem (2.1) is difficult to solve, while the sub-problem (2.3) can be solved easily if $(I + \alpha \partial f)^{-1}$ ($\alpha > 0$), the resolvent of f which is single-valued [11], is expressible and R_k is specially chosen as $r_k I - A^T H_k A \succeq 0$.

Algorithm 2.2 Step 0: Choose $w^0 = (x^0, y^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$. Let $\{H_k \in \mathbb{R}^{n \times m}\}$, $\{R_k \in \mathbb{R}^{n \times n}\}$, $\{S_k \in \mathbb{R}^{m \times m}\}$, and $\{T_k \in \mathbb{R}^{m \times m}\}$ ($k = 0, 1, 2, \dots$) be symmetric and positive definite matrix sequences, $R_k - A^T H_k A \succeq 0$, $S_k - H_k \succeq 0$, $\gamma \in (0, 2)$, $\varepsilon > 0$, $k:=0$.

Step 1: (*proximal step*) Compute $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ by

$$\tilde{x}^k = \arg \min_x \left\{ f(x) + \langle \lambda^k + H_k(Ax^k - y^k), Ax \rangle + \frac{1}{2}\|x - x^k\|_{R_k}^2 \right\}, \quad (2.7)$$

$$\tilde{y}^k = \arg \min_y \left\{ g(y) - \langle \lambda^k + H_k(Ax^k - y^k), y \rangle + \frac{1}{2}\|y - y^k\|_{S_k}^2 \right\}, \quad (2.8)$$

$$\tilde{\lambda}^k = \arg \min_{\lambda} \left\{ -\langle \lambda, A\tilde{x}^k - \tilde{y}^k \rangle + \frac{1}{2}\|\lambda - \lambda^k\|_{T_k}^2 \right\}. \quad (2.9)$$

If $\max\{\|A(x^k - \tilde{x}^k)\|_{\infty}, \|y^k - \tilde{y}^k\|_{\infty}, \|\lambda^k - \tilde{\lambda}^k\|_{\infty}\} < \varepsilon$, stop, otherwise, continue.

Step 2: (*direction step*) The new iteration $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ is generated by

$$w^{k+1} = w^k - \xi_k M_k(w^k - \tilde{w}^k), \quad (2.10)$$

where $\xi_k = \gamma \varphi(w^k, \tilde{w}^k) / \|M_k(w^k - \tilde{w}^k)\|^2$ with

$$\varphi(w^k, \tilde{w}^k) = \|w^k - \tilde{w}^k\|_{M_k}^2 + \langle A\tilde{x}^k - \tilde{y}^k, H_k(Ax^k - y^k) \rangle, \quad (2.11)$$

and

$$M_k = \begin{pmatrix} R_k & & \\ & S_k & \\ & & T_k \end{pmatrix}$$

$k:=k+1$, goto Step 1.

Algorithm 2.2 requires $R_k - A^T H_k A \succeq 0$ and $S_k - H_k \succeq 0$ ($k > 0$) at each iterate.

If $R_k = S_k = T_k = r^{-1}I_n$ and $H_k = rI_n$ where I_n is an identity matrix with $n \times n$ dimensions, it is automatically satisfied that $r \leq \frac{1}{\sqrt{\|A^T A\|}}$, which is different from the restriction of r in the PBDM, see [1, Theorem 3.1].

Remark 2.1

- i) It is obvious that $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ is a primal-dual solution of problem (1.1) and its dual if and only if $Ax^k = A\tilde{x}^k$, $y^k = \tilde{y}^k$ and $\lambda^k = \tilde{\lambda}^k$, that is why we take $\max\{\|A(x^k - \tilde{x}^k)\|_\infty, \|y^k - \tilde{y}^k\|_\infty, \|\lambda^k - \tilde{\lambda}^k\|_\infty\} < \varepsilon$ as the stopping criterion in the two methods above.
- ii) Both of Algorithms 2.1 and 2.2 could solve x and y in parallel. As pointed in [7], if we solve the sub-problems (2.3) and (2.4) (2.7 and 2.8) separately in a determinate order, Algorithm 2.1 (Algorithm 2.2) leads to a variant of standard alternating direction.

3 Convergence Analysis

Suppose the primal-dual solution set, denoted by $\mathcal{W}^*$, is nonempty. For the convenience of convergence analysis, we would like to characterize problem (1.1) by a maximal formulation. Since $f: \mathbb{R}^n \rightarrow (-\infty, +\infty]$ and $g: \mathbb{R}^m \rightarrow (-\infty, +\infty]$ are closed proper convex functions, the set-valued sub-differential mapping

$$T(x, y, \lambda) := \partial_{x,y} L(x, y, \lambda) \times \partial_\lambda (-L(x, y, \lambda))$$

on $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ is maximal monotone [10]. Finding optimal solutions of problem (1.1) and its dual are equivalent to finding a triple $(x^*, y^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ which satisfies

$$0 \in T(x^*, y^*, \lambda^*) \Leftrightarrow \begin{cases} 0 & \in \partial f(x^*) + A^T \lambda^*, \\ 0 & \in \partial g(y^*) - \lambda^*, \\ 0 & = -(Ax^* - y^*). \end{cases} \quad (3.1)$$

Lemma 3.1 ([12, 13]) Consider any maximal monotone mapping $T: \mathbb{R}^n \rightarrow 2^{\mathbb{R}^n}$. If sequence $\{x^k\}$ in $\mathbb{R}^n$ converges to some x , sequence $\{u^k\}$ in $\mathbb{R}^n$ converges to some u and $u^k \in T(x^k)$ for all k , then $u \in T(x)$.

Lemma 3.2 Let $A \in \mathbb{R}^{m \times n}$ and $H \in \mathbb{R}^{m \times m}$ be symmetric and positive definite matrix. For any $x_1, x_2 \in \mathbb{R}^n$ and $y_1, y_2 \in \mathbb{R}^m$, the following inequality holds,

$$\begin{aligned} & \langle Ax_1 - y_1, H(Ax_2 - y_2) \rangle \\ & \geq \frac{1}{2} \|Ax_1 - y_1\|_H^2 + \frac{1}{2} \|Ax_2 - y_2\|_H^2 - \|x_1 - x_2\|_{A^T H A}^2 - \|y_1 - y_2\|_H^2. \end{aligned}$$

Proof Using

$$\langle Ax_1 - y_1, H(Ax_2 - y_2) \rangle = \frac{1}{2} (\|Ax_1 - y_1\|_H^2 + \|Ax_2 - y_2\|_H^2 - \|A(x_1 - x_2) - (y_1 - y_2)\|_H^2)$$

and $\|A(x_1 - x_2) - (y_1 - y_2)\|_H^2 \leq 2(\|x_1 - x_2\|_{A^T H A}^2 + \|y_1 - y_2\|_H^2)$, we have the desired inequality immediately.

3.1 Algorithm 2.1's convergence

Theorem 3.1 Suppose that the primal-dual solution set $\mathcal{W}^*$ is nonempty. Let $w^k = (x^k, y^k, \lambda^k)$, $k \geq 0$, be generated by Algorithm 2.1, $w^* = (x^*, y^*, \lambda^*) \in \mathcal{W}^*$, then, we have

$$\langle w^k - w^*, M_k(w^k - \tilde{w}^k) \rangle \geq \|w^k - \tilde{w}^k\|_{P_k}^2 + \frac{1}{2} \|A\tilde{x}^k - \tilde{y}^k\|_{H_k}^2 + \frac{1}{2} \|Ax^k - y^k\|_{H_k}^2. \quad (3.2)$$

Proof Since $(x^*, y^*, \lambda^*) \in \mathcal{W}^*$, (3.1a–3.1c) hold. From (2.3)–(2.5), we have

$$0 \in \partial f(\tilde{x}^k) + A^T[\lambda^k + H_k(A\tilde{x}^k - y^k)] + R_k(\tilde{x}^k - x^k), \quad (3.3)$$

$$0 \in \partial g(\tilde{y}^k) - [\lambda^k + H_k(Ax^k - \tilde{y}^k)] + S_k(\tilde{y}^k - y^k), \quad (3.4)$$

$$0 = -(A\tilde{x}^k - \tilde{y}^k) + T_k(\tilde{\lambda}^k - \lambda^k). \quad (3.5)$$

Using (3.1a), (3.3) and the monotonicity of ∂f , we obtain

$$\langle \tilde{x}^k - x^*, -A^T[\lambda^k + H_k(A\tilde{x}^k - y^k)] - R_k(\tilde{x}^k - x^k) + A^T\lambda^* \rangle \geq 0.$$

By rearranging the above terms, we obtain

$$\begin{aligned} & \langle \tilde{x}^k - x^*, (A^T H_k A + R_k)(x^k - \tilde{x}^k) \rangle - \langle \lambda^k - \lambda^*, A(\tilde{x}^k - x^*) \rangle \\ & \geq \langle A(\tilde{x}^k - x^*), H_k(Ax^k - y^k) \rangle. \end{aligned} \quad (3.6)$$

Using (3.1b), (3.4) and the monotonicity of ∂g , we have

$$\langle \tilde{y}^k - y^*, \lambda^k + H_k(Ax^k - \tilde{y}^k) - S_k(\tilde{y}^k - y^k) - \lambda^* \rangle \geq 0,$$

then

$$\langle \tilde{y}^k - y^*, \lambda^k - \lambda^* \rangle + \langle \tilde{y}^k - y^*, (H_k + S_k)(y^k - \tilde{y}^k) \rangle \geq -\langle \tilde{y}^k - y^*, H_k(Ax^k - y^k) \rangle. \quad (3.7)$$

Adding (3.6) and (3.7), we obtain

$$\begin{aligned} & \langle \tilde{x}^k - x^*, (A^T H_k A + R_k)(x^k - \tilde{x}^k) \rangle + \langle \tilde{y}^k - y^*, (H_k + S_k)(y^k - \tilde{y}^k) \rangle \\ & + \langle \lambda^k - \lambda^*, -(A\tilde{x}^k - \tilde{y}^k) \rangle \geq \langle A\tilde{x}^k - \tilde{y}^k, H_k(Ax^k - y^k) \rangle. \end{aligned} \quad (3.8)$$

Applying $-(A\tilde{x}^k - \tilde{y}^k) = T_k(\lambda^k - \tilde{\lambda}^k)$ to the last term of (3.8), we have

$$\begin{aligned} & \langle \tilde{x}^k - x^*, (A^T H_k A + R_k)(x^k - \tilde{x}^k) \rangle + \langle \tilde{y}^k - y^*, (H_k + S_k)(y^k - \tilde{y}^k) \rangle \\ & + \langle \tilde{\lambda}^k - \lambda^*, T_k(\tilde{\lambda}^k - \lambda^k) \rangle \geq \langle A\tilde{x}^k - \tilde{y}^k, H_k(Ax^k - y^k) \rangle. \end{aligned}$$

And by rearranging the terms, we get

$$\begin{aligned} & \langle w^k - w^*, M_k(w^k - \tilde{w}^k) \rangle \\ & \geq \|x^k - \tilde{x}^k\|_{A^T H_k A + R_k}^2 + \|y^k - \tilde{y}^k\|_{H_k + S_k}^2 + \langle A\tilde{x}^k - \tilde{y}^k, H_k(Ax^k - y^k) \rangle. \end{aligned} \quad (3.9)$$

It follows from Lemma 3.2 that

$$\begin{aligned} & \langle A\tilde{x}^k - \tilde{y}^k, H_k(Ax^k - y^k) \rangle \\ & \geq \frac{1}{2} \|A\tilde{x}^k - \tilde{y}^k\|_{H_k}^2 + \frac{1}{2} \|Ax^k - y^k\|_{H_k}^2 - \|x^k - x^k\|_{A^T H_k A}^2 - \|y^k - y^k\|_{H_k}^2. \end{aligned} \quad (3.10)$$

Replacing the last term in the right-hand side of (3.9) by (3.10), the desired result follows.

Remark 3.1 From Theorem 3.1, we get that $\langle w^k - w^*, -M_k(w^k - \tilde{w}^k) \rangle < 0$ when $w^k \neq \tilde{w}^k$, thus $-M_k(w^k - \tilde{w}^k)$ is a descent direction of $\frac{1}{2}\|w - w^*\|^2$ at $w^k \notin \mathcal{W}^*$.

Theorem 3.2 Let $\{w^k\}$ ($k = 1, 2, \dots$) be the sequence generated by Algorithm 2.1. If the inequality (3.2) holds, then the sequence $\{w^k\}$ is globally convergent to a pair of primal-dual solutions $w^* = (x^*, y^*, \lambda^*) \in \mathcal{W}^*$.

Proof Let (x^*, y^*) and λ^* be the optimal solutions of the primal problem (1.1) and its dual. Combining with (2.6) yields

$$\begin{aligned} & \|w^{k+1} - w^*\|^2 \\ & = \|w^k - \xi_k M_k(w^k - \tilde{w}^k) - w^*\|^2 \\ & = \|w^k - w^*\|^2 - 2\xi_k \langle w^k - w^*, M_k(w^k - \tilde{w}^k) \rangle + \xi_k^2 \|M_k(w^k - \tilde{w}^k)\|^2. \end{aligned} \quad (3.11)$$

Applying inequality (3.2) to (3.11), we get

$$\begin{aligned} \|w^{k+1} - w^*\|^2 & \leq \|w^k - w^*\|^2 - 2\xi_k \left(\|w^k - \tilde{w}^k\|_{P_k}^2 + \frac{1}{2} \|A\tilde{x}^k - \tilde{y}^k\|_{H_k}^2 + \frac{1}{2} \|Ax^k - y^k\|_{H_k}^2 \right) \\ & \quad + \xi_k^2 \|M_k(w^k - \tilde{w}^k)\|^2. \end{aligned} \quad (3.12)$$

Combining $\xi_k = \gamma(\|w^k - \tilde{w}^k\|_{P_k}^2 + \frac{1}{2} \|A\tilde{x}^k - \tilde{y}^k\|_{H_k}^2 + \frac{1}{2} \|Ax^k - y^k\|_{H_k}^2) / \|M_k(w^k - \tilde{w}^k)\|^2$ with (3.12) yields

$$\begin{aligned} & \|w^{k+1} - w^*\|^2 \leq \|w^k - w^*\|^2 \\ & \quad - \gamma(2 - \gamma) \frac{(\|w^k - \tilde{w}^k\|_{P_k}^2 + \frac{1}{2} \|A\tilde{x}^k - \tilde{y}^k\|_{H_k}^2 + \frac{1}{2} \|Ax^k - y^k\|_{H_k}^2)^2}{\|M_k(w^k - \tilde{w}^k)\|^2}. \end{aligned} \quad (3.13)$$

It follows from (3.13) that the sequence $\{\|w^k - w^*\|^2\}$ is nonincreasing and has a limit.

Hence, sequence $\{w^k\}$ is bounded and

$$\lim_{k \rightarrow +\infty} \frac{(\|w^k - \tilde{w}^k\|_{P_k}^2 + \frac{1}{2}\|A\tilde{x}^k - \tilde{y}^k\|_{H_k}^2 + \frac{1}{2}\|Ax^k - y^k\|_{H_k}^2)^2}{\|M_k(w^k - \tilde{w}^k)\|^2} = 0.$$

Consequently,

$$\lim_{k \rightarrow +\infty} \|w^k - \tilde{w}^k\|^2 = 0, \quad (3.14)$$

$$\lim_{k \rightarrow +\infty} \|Ax^k - y^k\|^2 = 0. \quad (3.15)$$

Sequence $\{w^k\}$ is bounded, then there exists at least one accumulation point $w^\infty = (x^\infty, y^\infty, \lambda^\infty)$.

It follows from (3.14) that the sequence $\{\tilde{w}^k\}$ has the same accumulation points with $\{w^k\}$.

Since $\tilde{x}^{k_i}$ and $\tilde{y}^{k_i}$ are the optimal solutions of (2.3) and (2.4), respectively, from (3.3) to (3.5), we have

$$\begin{cases} 0 = \tilde{u}^{k_i} + A^T[\lambda^{k_i} + H_k(A\tilde{x}^{k_i} - y^{k_i})] + R_k(\tilde{x}^{k_i} - x^{k_i}), \\ 0 = \tilde{v}^{k_i} - [\lambda^{k_i} + H_k(Ax^{k_i} - \tilde{y}^{k_i})] + S_k(\tilde{y}^{k_i} - y^{k_i}), \end{cases} \quad (3.16)$$

and

$$-(A\tilde{x}^{k_i} - \tilde{y}^{k_i}) + T_k(\tilde{\lambda}^{k_i} - \lambda^{k_i}) = 0,$$

with $\tilde{u}^{k_i} \in \partial f(\tilde{x}^{k_i})$, $\tilde{v}^{k_i} \in \partial g(\tilde{x}^{k_i})$. Obviously, $\{\tilde{u}^{k_i}\}$ and $\{\tilde{v}^{k_i}\}$ are bounded, therefore, there exist u and v which are two clusters of points of $\{\tilde{u}^{k_i}\}$ and $\{\tilde{v}^{k_i}\}$, respectively, i.e., there exist subsequence $\{\tilde{u}^{k_{ij}}\}$ of the sequence $\{\tilde{u}^{k_i}\}$ and $\{\tilde{v}^{k_{ij}}\}$ of $\{\tilde{v}^{k_i}\}$, which satisfy

$$\lim_{j \rightarrow +\infty} \tilde{u}^{k_{ij}} = u, \quad \lim_{j \rightarrow +\infty} \tilde{v}^{k_{ij}} = v.$$

Applying Lemma 3.1, we have $u \in \partial f(x^\infty)$ and $v \in \partial g(x^\infty)$. It follows from (3.16) that

$$\begin{cases} 0 = \lim_{j \rightarrow +\infty} (\tilde{u}^{k_{ij}} + A^T[\lambda^{k_{ij}} + H_k(A\tilde{x}^{k_{ij}} - y^{k_{ij}})] + R_k(\tilde{x}^{k_{ij}} - x^{k_{ij}})), \\ 0 = \lim_{j \rightarrow +\infty} (\tilde{v}^{k_{ij}} - [\lambda^{k_{ij}} + H_k(Ax^{k_{ij}} - \tilde{y}^{k_{ij}})] + S_k(\tilde{y}^{k_{ij}} - y^{k_{ij}})), \end{cases}$$

then

$$\begin{cases} 0 = u + A^T[\lambda^\infty + H_k(Ax^\infty - y^\infty)] + R_k(x^\infty - x^\infty), \\ 0 = v - [\lambda^\infty + H_k(Ax^\infty - y^\infty)] + S_k(y^\infty - y^\infty), \end{cases}$$

i.e.,

$$\begin{cases} 0 \in \partial f(x^\infty) + A^T\lambda^\infty, \\ 0 \in \partial g(y^\infty) - \lambda^\infty. \end{cases}$$

By (3.15), we have

$$0 = Ax^\infty - y^\infty.$$

Thus, $w^\infty = (x^\infty, y^\infty, \lambda^\infty)$ is a pair of primal-dual solutions of problem (1.1) and satisfies

$$\begin{aligned} \|w^{k+1} - w^\infty\|^2 &\leq \|w^k - w^\infty\|^2 \\ &\quad - \gamma(2 - \gamma) \frac{(\|w^k - \tilde{w}^k\|_{P_k}^2 + \frac{1}{2}\|A\tilde{x}^k - \tilde{y}^k\|_{H_k}^2 + \frac{1}{2}\|Ax^k - y^k\|_{H_k}^2)^2}{\|M_k(w^k - \tilde{w}^k)\|^2}. \end{aligned}$$

Accordingly, the sequence $\{w^k\}$ globally converges to $w^\infty \in \mathcal{W}^*$. The required result is obtained by taking $w^* = w^\infty$.

3.2 Algorithm 2.2's convergence

Theorem 3.3 Let $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ be generated by (2.7)–(2.9) with given $w^k = (x^k, y^k, \lambda^k)$ and $\varphi(w^k, \tilde{w}^k)$ be defined as (2.11). Then, we have

$$\begin{aligned} \varphi(w^k, \tilde{w}^k) &\geq \|x^k - \tilde{x}^k\|_{R_k - A^T H_k A}^2 + \|y^k - \tilde{y}^k\|_{S_k - H_k}^2 \\ &\quad + \|\lambda^k - \tilde{\lambda}^k\|_{T_k + \frac{1}{2}T_k H_k T_k}^2 + \frac{1}{2}\|Ax^k - y^k\|_{H_k}^2. \end{aligned}$$

Proof Using the definition of $\varphi(w^k, \tilde{w}^k)$ in (2.11), Lemma 3.2 and $T_k(\lambda^k - \tilde{\lambda}^k) = -(A\tilde{x}^k - \tilde{y}^k)$, the desired result is obtained immediately.

Theorem 3.4 Let $w^k = (x^k, y^k, \lambda^k)$, and $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ ($k \geq 0$) be generated by Algorithm 2.2, $w^* = (x^*, y^*, \lambda^*) \in \mathcal{W}^*$. Then, we have

$$\langle \tilde{w}^k - w^*, M_k(w^k - \tilde{w}^k) \rangle \geq \langle A\tilde{x}^k - \tilde{y}^k, H_k(Ax^k - y^k) \rangle, \quad (3.17)$$

and consequently

$$\langle w^k - w^*, M_k(w^k - \tilde{w}^k) \rangle \geq \varphi(w^k, \tilde{w}^k). \quad (3.18)$$

Proof From (2.7)–(2.9), we have

$$0 \in \partial f(\tilde{x}^k) + A^T(\lambda^k + H_k(Ax^k - y^k)) + R_k(\tilde{x}^k - x^k), \quad (3.19)$$

$$0 \in \partial g(\tilde{y}^k) - (\lambda^k + H_k(Ax^k - y^k)) + S_k(\tilde{y}^k - y^k), \quad (3.20)$$

$$0 = -(A\tilde{x}^k - \tilde{y}^k) + T_k(\tilde{\lambda}^k - \lambda^k). \quad (3.21)$$

Using (3.1a), (3.19) and the monotonicity of ∂f , we have

$$\langle \tilde{x}^k - x^*, -A^T(\lambda^k + H_k(Ax^k - y^k)) - R_k(\tilde{x}^k - x^k) + A^T\lambda^* \rangle \geq 0.$$

Accordingly,

$$\langle \tilde{x}^k - x^*, R_k(x^k - \tilde{x}^k) \rangle + \langle -A(\tilde{x}^k - x^*), \lambda^k - \lambda^* \rangle \geq \langle \tilde{x}^k - x^*, A^T H_k(Ax^k - y^k) \rangle. \quad (3.22)$$

Using (3.1b), (3.20) and the monotonicity of $\hat{\partial}g$, we obtain

$$\langle \tilde{y}^k - y^*, \lambda^k + H_k(Ax^k - y^k) - S_k(\tilde{y}^k - y^k) - \lambda^* \rangle \geq 0,$$

then

$$\langle \tilde{y}^k - y^*, S_k(y^k - \tilde{y}^k) \rangle + \langle \tilde{y}^k - y^*, \lambda^k - \lambda^* \rangle \geq -\langle \tilde{y}^k - y^*, H_k(Ax^k - y^k) \rangle. \quad (3.23)$$

Adding (3.22) to (3.23), and combining $-(A\tilde{x}^k - \tilde{y}^k) = T_k(\lambda^k - \tilde{\lambda}^k)$ and (3.1c), we get (3.17) directly. Furthermore, using the definition of $\varphi(w^k, \tilde{w}^k)$ and

$$\langle w^k - w^*, M_k(w^k - \tilde{w}^k) \rangle = \|w^k - \tilde{w}^k\|_{M_k}^2 + \langle \tilde{w}^k - w^*, M_k(w^k - \tilde{w}^k) \rangle,$$

(3.18) follows from (3.17).

Theorem 3.5 The sequence $\{w^k\}$ ($k = 1, 2, \dots$) generated by Algorithm 2.2 is globally convergent to a pair of primal-dual solutions $w^* = (x^*, y^*, \lambda^*) \in \mathcal{W}^*$.

Proof Similar as the proof procedure of Theorem 3.2, the desired convergence result can be obtained easily by Theorems 3.3 and 3.4.

4 Numerical Experiment

In this section, we consider the problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{1}{2} \|Kx - b\|^2 + \delta_{R_+^n}(y) \\ \text{s.t.} \quad & x = y \end{aligned} \quad (4.1)$$

to illustrate the effectiveness of the proximal-based pre-correction decomposition methods. In addition, $\delta_{R_+^n}(y) = 0$ if $y \geq 0$, otherwise, $\delta_{R_+^n}(y) = \infty$.

All codes were written by Matlab R2011a, and all the experiments were carried out on a notebook computer equipped with Windows 7, AMD Athlon(tm) II X2 260 Processor 3.20 GHz, and 2.00 GB of RAM.

Let $K \in \mathbb{R}^{l \times n}$ be generated as follows:

$$\begin{aligned} l &= 250; r = 200; \text{rng}('default'); \text{rng}(0); Kl = \text{sprandn}(m, r, 0.25); \\ &\text{rng}(1); Kr = \text{sprandn}(r, n, 0.1); K = Kl * Kr. \end{aligned}$$

Then, $b = K\bar{x} + z$, where $\bar{x} = \text{sprandn}(n, 1, 0.05)$ and $z = 0.01 * \text{randn}(l, 1)$ with seed 0 in MATLAB.

In the practical implementations, we set the starting point $w^0 = \text{rand}(2m + n, 1)$ and the stopping criterion as $\max\{\|x^k - \tilde{x}^k\|_\infty, \|y^k - \tilde{y}^k\|_\infty, \|\lambda^k - \tilde{\lambda}^k\|_\infty\} \leq 10^{-6}$. In Algorithms 2.1 and 2.2, $H_k = hI$, $R_k = r^{-1}I$, $S_k = s^{-1}I$, and $T_k = t^{-1}I$. The parameters of the Algorithm 2.1 were set as follows: $h = 1.1$, $r = 1.9$, $s = 1.9$, $t = 0.6$, $\gamma = 1.2$, which were set as 0.9, 0.7, 0.8, 0.8, and 0.9,

Table 1 Numerical results of the four methods

n	Algorithm 2.1		Algorithm 2.2		PSALM		PBDM	
	Iter	CPU	Iter	CPU	Iter	CPU	Iter	CPU
3 000	22	5.41	24	6.00	29	7.05	41	9.86
4 000	23	10.80	20	9.46	31	14.32	40	18.45
4 500	21	13.40	24	15.16	33	20.83	42	26.01

respectively, in the Algorithm 2.2. For the PSALM, we took $H = hI$ with $h = 1.2$, $\gamma = 1.0$. And for the PBDM, we chose $\alpha = 0.7$. The numerical performances were reported in Table 1, where Iter represents the number of iterations and CPU stands for the CPU time elapsed in seconds.

From Table 1, we can conclude that the proximal-based pre-correction decomposition methods are effective to solve problem (4.1), this coincides the theoretical analysis made above.

5 Conclusion

Two proximal-based pre-correction decomposition methods have been discussed for solving given convex minimization problem with separable structure. The replacement of the regularization parameters by the regularization matrixes in Algorithms 2.1 and 2.2 is one of the main differences from the PBDM, the direction steps are others. Comparing with the PSALM, they bring the proximal terms in proximal steps and have different descend directions. Like the PSALM and the PBDM, the proposed methods are variants of standard alternating direction when they are implemented sequentially, so their convergence properties are similar.

Acknowledgments The authors are grateful to the two anonymous referees for their constructive review and valuable comments and suggestions.

References

- [1] Chen, G., Teboulle, M.: A proximal-based decomposition method for convex minimization problems. *Math. Program.* **64**, 81–101 (1994)
- [2] Fukushima, M.: Application of the alternating direction method of multipliers to separable convex programming problems. *Comput. Optim. Appl.* **2**, 93–111 (1992)
- [3] Glowinski, R., Marocco, A.: Sur l'approximation par éléments finis d'ordre un, et la résolution par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. *Rev. Française d'Aut. Inf. Rech. Oper.* **9**, 41–76 (1975)
- [4] Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite-element approximations. *Comput. Math. Appl.* **2**, 17–40 (1976)
- [5] Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R. (eds.) *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Valued Problems*, pp. 299–331. Amsterdam, North-Holland (1983)
- [6] Goldfarb, D., Ma, S.Q., Scheinberg, K.: Fast alternating linearization methods for minimizing the sum of two convex functions. *Math. Program.* **141**, 349–382 (2013)

- [7] He, B.S.: Parallel splitting augmented Lagrangian methods for monotone structured variational inequalities. *Comput. Optim. Appl.* **42**, 195–212 (2009)
- [8] He, B.S., Tao, M., Xu, M., Yuan, X.M.: An alternating direction-based contraction method for linearly constrained separable convex programming problems. *Optimization* **62**, 573–596 (2013)
- [9] Nocedal, J., Wright, S.J.: *Numerical Optimization*, 2nd edn. Springer, Berlin (2006)
- [10] Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.* **1**, 97–116 (1976)
- [11] Rockafellar, R.T.: On the maximality of sums of nonlinear monotone operators. *Trans. Am. Math. Soc.* **149**, 75–88 (1970)
- [12] Brézis, H.: *Opérateurs Maximaux Monotones*. North-Holland, Amsterdam (1973)
- [13] Pascali, D., Sburlan, S.: *Nonlinear Mappings of Monotone Type*. Editura Academiei, Bucharest (1978)