Nonlinear Proximal Decomposition Method for Convex Programming [1]

Masahiro Kyono and Masao Fukushima

Department of Applied Mathematics and Physics

Graduate School of Informatics

Kyoto University

Kyoto 606-8501, Japan

July 27, 1999

Abstract. In this paper, we propose a new decomposition method for solving convex programming problems with separable structure. The proposed method is based on the decomposition method proposed by Chen and Teboulle and the nonlinear proximal point algorithm using Bregman function. An advantage of the proposed method is that, by a suitable choice of Bregman function, each subproblem essentially becomes unconstrained minimization of a finite-valued convex function. Under some appropriate assumptions, the method is globally convergent to a solution of the problem.

Key Words. Nonlinear proximal decomposition method, Bregman function, convex programming.

[1]This work was supported in part by the Scientific Research Grant-in-Aid from the Ministry of Education, Science, Sports, and Culture, Japan. The authors are grateful to Regina Burachik, Dan Butnariu, Alfredo Iusem, and Michael Solodov for their helpful comments.

1 Introduction

A point-to-set mapping T from $\Re^n$ into itself is called a monotone operator if it has the property that $\langle x - x', z - z' \rangle \geq 0,\ \forall x, x' \in \Re^n, z \in T(x), z' \in T(x')$, where $\langle \cdot, \cdot \rangle$ denotes the inner product. If there does not exist a monotone operator T' such that $\mathrm{Graph}(T) \subseteq \mathrm{Graph}(T')$ and $\mathrm{Graph}(T) \neq \mathrm{Graph}(T')$, where $\mathrm{Graph}(T) = \{(x, y) \in \Re^n \times \Re^n \mid y \in T(x)\}$, then T is said to be maximal monotone. The *proximal point algorithm* (PPA) proposed by Martinet [17] and developed by Rockafellar [20] is an iterative method for finding a root of a maximal monotone operator T.

The PPA generates a sequence $\{x^k\}$ with the following iterative scheme:

$$x^{k+1} \simeq (I + c_k T)^{-1}(x^k),$$

where I is the identity mapping on $\Re^n$ and $\{c_k\}$ is a nondecreasing sequence of positive scalars. If T is a maximal monotone operator, then $(I + c_k T)^{-1}$ is a point-to-point mapping and the sequence $\{x^k\}$ converges to a root of T. Recently, generalized PPAs using nonlinear functions such as Bregman function or φ divergence have been studied extensively [1, 2, 7, 8, 10, 11, 15, 22, 23]. The *nonlinear PPA* (NPPA) using Bregman function [10, 11] generates a sequence $\{x^k\}$ with the following iterative scheme:

$$x^{k+1} \simeq (\nabla h + c_k T)^{-1}(\nabla h(x^k)),$$

where h is a Bregman function whose definition will be given in Section 2. In particular, when applied to convex programming problems this method is called the *proximal minimization algorithm with D-function* (PMAD).

In this paper, we propose a new decomposition method based on the NPPA for solving the convex programming problem with separable structure:

$$\text{(P)} \quad \min_{x \in \Re^n, z \in \Re^m} \ f(x) + g(z) \quad \text{s.t.} \quad z = Ax,$$

where $f : \Re^n \to (-\infty, +\infty]$ and $g : \Re^m \to (-\infty, +\infty]$ are closed proper convex functions and A is an $m \times n$ matrix. The Fenchel's dual problem [19] for (P) is defined as

$$\text{(D)} \quad \min_{y \in \Re^m} \ f^*(-A^T y) + g^*(y),$$

where f^* and g^* are the conjugate functions of f and g, respectively, and y is a vector of Lagrange multipliers associated with the constraints in (P). The Lagrangian $L : \Re^n \times \Re^m \times \Re^m \to (-\infty, +\infty]$ for (P) is defined by

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

If (x^*, z^*) is an optimal solution of (P) and y^* is a vector of corresponding Lagrange multipliers, then y^* is an optimal solution of (D) and (x^*, z^*, y^*) is a saddle point of L [19], i.e., for all $x \in \mathrm{dom}\, f$, $z \in \mathrm{dom}\, g$ and $y \in \Re^m$,

$$L(x^*, z^*, y) \leq L(x^*, z^*, y^*) \leq L(x, z, y^*). \tag{1}$$

Classical methods for solving (P) based on the PPA are the method of multipliers and the proximal method of multipliers. The *method of multipliers* [5], which is obtained by applying the PPA to the dual (D), generates a sequence $\{(x^k, z^k, y^k)\}$ with the following iterative scheme:

$$\begin{aligned} &(x^{k+1}, z^{k+1}) \simeq \arg\min \left\{ L(x, z, y^k) + \frac{c_k}{2} \|Ax - z\|^2 \right\}, \\ &y^{k+1} = y^k + c_k(Ax^{k+1} - z^{k+1}). \end{aligned}$$

The *proximal method of multipliers* [21], which is obtained by applying the PPA to the primal-dual system, generates a sequence $\{(x^k, z^k, y^k)\}$ with the following iterative scheme:

$$\begin{aligned} &(x^{k+1}, z^{k+1}) \simeq \arg\min \left\{ L(x, z, y^k) + \frac{c_k}{2} \|Ax - z\|^2 + \frac{1}{2c_k}\|x - x^k\|^2 + \frac{1}{2c_k}\|z - z^k\| \right\}, \\ &y^{k+1} = y^k + c_k(Ax^{k+1} - z^{k+1}). \end{aligned}$$

However, these algorithms do not take full advantage of the separable structure the original problem (P) has. Various decomposition methods that exploit the special structure of the problem have been proposed. An example of such methods is the *alternating direction method of multipliers* (ADMM) [6, 14, 16], which generates a sequence $\{(x^k, z^k, y^k)\}$ with the following iterative scheme:

$$\begin{aligned} &x^{k+1} \simeq \arg\min \left\{ f(x) + \langle y, Ax \rangle + \frac{c_k}{2} \|Ax - z^k\|^2 \right\}, \\ &z^{k+1} \simeq \arg\min \left\{ g(z) - \langle y, z \rangle + \frac{c_k}{2} \|Ax^{k+1} - z\|^2 \right\}, \\ &y^{k+1} = y^k + c_k(Ax^{k+1} - z^{k+1}). \end{aligned}$$

This method has further been extended and modified in various ways [9, 12, 13, 24]. In this paper, we are particularly interested in the following decomposition algorithm proposed by Chen and Teboulle [9]:

$$p^{k+1} = y^k + c_k(Ax^k - z^k), \tag{2}$$

$$x^{k+1} \simeq \arg\min_{x\in\Re^n} \left\{ f(x) + \langle p^{k+1}, Ax\rangle + \frac{1}{2c_k}\|x - x^k\|^2 \right\}, \tag{3}$$

$$z^{k+1} \simeq \arg\min_{z\in\Re^m} \left\{ g(z) - \langle p^{k+1}, z\rangle + \frac{1}{2c_k}\|z - z^k\|^2 \right\}, \tag{4}$$

$$y^{k+1} = y^k + c_k(Ax^{k+1} - z^{k+1}). \tag{5}$$

This algorithm, called the *predictor corrector proximal multiplier method* (PCPMM) [9], is more amenable to parallel computation than ADMM, because the quadratic term in (3) does not involve matrix A.

The decomposition algorithm proposed in this paper is obtained by combining the PCPMM with the PMAD. An advantage of the proposed algorithm over the PCPMM is that, by a suitable choice of D-function that plays the role of an interior penalty function, each subproblem essentially becomes unconstrained minimization of a finite-valued convex function. To the authors' knowledge, this is the first attempt to develop a proximal-based decomposition method that utilizes Bregman D-function.

The organization of this paper is as follows. In Section 2, we state some fundamental properties of the NPPA using Bregman function. In Section 3, we describe the proposed algorithm, and in Section 4 we show its convergence. Finally, in Section 5 we give some concluding remarks.

2 Preliminaries

In this section, we state the definitions of D-function and Bregman function, and summarize some basic properties of the PMAD.

Definition 1 (D-function) *Let $S \subseteq \Re^n$ be a nonempty open convex set and $\bar{S}$ denote its closure. Let $\phi : \bar{S} \to \Re$ be a strictly convex function that is differentiable on S. The function $D_\phi : \bar{S} \times S \to \Re$ defined by*

$$D_\phi(u, v) = \phi(u) - \phi(v) - \langle \nabla\phi(v), u - v\rangle$$

is called D-function.

By the strict convexity of ϕ, the following relations are satisfied:

$$D_\phi(u,v) \geq 0,$$
$$D_\phi(u,v) = 0 \quad \text{iff} \quad u = v.$$

So we can interpret D-function as a generalized distance function. But, in general, the symmetry $D_\phi(u,v) = D_\phi(v,u)$ and the triangle inequality $D_\phi(u,v)+D_\phi(v,w) \geq D_\phi(u,w)$ do not hold.

In order to design effective algorithms that use D-function D_ϕ, we need to impose some additional requirements on the function ϕ. The conditions given below will turn out to be appropriate for the present purpose.

Definition 2 (Bregman function) *Let $S \subseteq \Re^n$ be a nonempty open convex set and $\bar{S}$ be its closure. A function $\phi : \bar{S} \to \Re$ is called a Bregman function with zone S if it satisfies the following conditions:*

(i) ϕ is strictly convex and continuous on $\bar{S}$.

(ii) ϕ is continuously differentiable on S.

(iii) Given any $u \in \bar{S}$ and $\alpha \in \Re$, the right partial level set $\mathcal{L}(u,\alpha) \equiv \{v \mid D_\phi(u,v) \leq \alpha\}$ is bounded.

(iv) If $\{v^k\} \subset S$ is a convergent sequence with limit v^∞, then $D_\phi(v^\infty, v^k) \to 0$.

The traditional definition of Bregman function requires two more conditions in the addition to the above conditions:

(a) The left partial level set $\mathcal{L}'(v,\alpha) \equiv \{u \mid D_\phi(u,v) \leq \alpha\}$ is bounded for every $u \in \bar{S}$ and $v \in S$.

(b) If $\{u^k\} \subset \bar{S}$ and $\{v^k\} \subset S$ are sequences such that $\{u^k\}$ is bounded and $\{v^k\}$ converges to $\bar{v}$, and $D_\phi(u^k, v^k) \to 0$, then $u^k \to \bar{v}$.

However, it has already been shown that these conditions are redundant because they are implied by the above conditions (i)-(iv) [4, 22].

Let us consider for the moment the problem of minimizing a closed proper convex function $h : \Re^n \to (-\infty, +\infty]$. When applied to this problem, the PMAD successively generates a next point u^+ from a current point $\bar{u} \in \mathrm{dom}\, h$ by the iterative scheme:

$$u^+ = \arg\min_{u \in \Re^n} \left\{ h(u) + \frac{1}{c} D_\phi(u, \hat{u}) \right\},$$

where $\phi : \bar{S} \to \Re$ is a Bregman function with $\mathrm{dom}\, h \subseteq \bar{S}$, and c is a positive scalar. An approximate version of the PMAD has been proposed [15], in which the above exact minimization is replaced with finding a pair (u^+, γ^+) satisfying the following conditions for some $\epsilon \geq 0$:

$$\begin{aligned} &\gamma^+ \in \partial_\epsilon h(u^+), \\ &\gamma^+ + \frac{1}{c} \partial_1 D_\phi(u^+, \hat{u}) \ni 0, \end{aligned} \tag{6}$$

where $\partial_\epsilon h$ is the ϵ-subdifferential of h defined by

$$\partial_\epsilon h(u) = \{p \in \Re^n \mid h(v) - h(u) \geq \langle p, v - u \rangle - \epsilon, \ \forall v \in \mathrm{dom}\, h\},$$

and $\partial_1 D_\phi$ denotes the subdifferential of D_ϕ with respect to the first argument.

From the definitions of D-function and ϵ-subdifferential, it is not difficult to show the following relation between u^+ and $\hat{u}$. The proof is omitted, since it is essentially proved in [15, Lemma 4.1].

Lemma 1 *Suppose that ϕ is a Bregman function with zone S and u^+ is satisfied (6). Then the following inequality is satisfied for all $u \in \mathrm{dom}\, \phi$:*

$$c\{h(u^+) - h(u)\} \leq D_\phi(u, \hat{u}) - D_\phi(u, u^+) - D_\phi(u^+, \hat{u}) + c\epsilon.$$

Proof See [15, Lemma 4.1]. ∎

The following concept [7] is crucial for the PMAD to be well-defined.

Definition 3 (Boundary coerciveness) *A Bregman function ϕ with zone S is said to be boundary coercive if it satisfies the following condition:*

(v) If $\{v^k\} \subset S$ is convergent to a point on the boundary of S and $u \in S$, then $D_\phi(u, v^k) \to +\infty$.

It has been shown [7, Lemma 1] that, if a Bregman function ϕ is boundary coercive, then for any fixed $v \in S$,

$$\partial_1 D_\phi(u, v) = \begin{cases} \nabla\phi(u) - \nabla\phi(v), & \text{if } u \in S \\ \emptyset & \text{otherwise.} \end{cases}$$

This property yields the following lemma, which ensures that the PMAD generates a sequence that lies in the open set S, whenever the Bregman function ϕ is boundary coercive.

Lemma 2 *If ϕ is a boundary coercive Bregman function with zone S, then condition (6) is rewritten as*

$$\begin{aligned} &\gamma^+ \in \partial_\epsilon h(u^+), \\ &\gamma^+ + \frac{1}{c}\{\nabla\phi(u^+) - \nabla\phi(\hat{u})\} = 0. \end{aligned}$$

In particular, for any $\hat{u} \in S$, u^+ is contained in S.

Proof Straightforward. ∎

3 Algorithm

In this section, we propose a generalized algorithm PCPMM that is obtained by replacing the quadratic terms in (3) and (4) with D-functions. In this algorithm, we will use the class of functions that are represented as

$$\phi(u) = \phi_0(u) + \frac{\mu}{2}\|u\|^2,$$

where ϕ_0 is a Bregman function and $\mu > 0$. It is easy to see that ϕ satisfies the conditions (i)-(iv) in Definition 1, so ϕ is also a Bregman function. We note that this class of Bregman functions has a similar flavor to the function used in the logarithmic-quadratic proximal algorithms of [1, 2].

The proposed algorithm, which we call the *nonlinear predictor corrector proximal multiplier method* (NPCPMM), is stated as follows:

Algorithm — NPCPMM

Step 0: Choose boundary coercive Bregman functions φ_0 and ψ_0 with zones $S_X \subseteq \Re^n$ and $S_Z \subseteq \Re^m$, respectively. Let functions $\varphi : \overline{S_X} \to \Re$ and $\psi : \overline{S_Z} \to \Re$ be defined by

$$\varphi(x) = \varphi_0(x) + \frac{\mu}{2}\|x\|^2, \tag{7}$$

and

$$\psi(z) = \psi_0(z) + \frac{\nu}{2}\|z\|^2, \tag{8}$$

respectively, where $\mu > 0$ and $\nu > 0$ are some constants. Choose an initial point $(x^0, z^0, y^0) \in S_X \times S_Z \times \Re^m$, and a constant $\varepsilon \in (0, \bar{c}/2)$, where

$$\bar{c} := \min\left\{\frac{\sqrt{\nu}}{2}, \frac{\sqrt{\mu}}{2\|A\|}\right\}.$$

Let $k := 0$ and go to Step 1.

Step 1: Terminate if a stopping criterion is met. Otherwise, choose scalars $a_k \geq 0$, $b_k \geq 0$ and c_k such that

$$\varepsilon < c_k < \bar{c} - \varepsilon \tag{9}$$

and calculate $p^{k+1} \in \Re^m$ by

$$p^{k+1} = y^k + c_k(Ax^k - z^k). \tag{10}$$

Step 2: Find $(x^{k+1}, \gamma^{k+1}) \in S_X \times \Re^n$ and $(z^{k+1}, \delta^{k+1}) \in S_Z \times \Re^m$ such that

$$\gamma^{k+1} \in \partial_{a_k} f^k(x^{k+1}), \quad \gamma^{k+1} + \frac{1}{c_k}\{\nabla\varphi(x^{k+1}) - \nabla\varphi(x^k)\} = 0, \tag{11}$$

$$\delta^{k+1} \in \partial_{b_k} g^k(z^{k+1}), \quad \delta^{k+1} + \frac{1}{c_k}\{\nabla\psi(z^{k+1}) - \nabla\psi(z^k)\} = 0, \tag{12}$$

where functions $f^k : \Re^n \to (-\infty, +\infty]$ and $g^k : \Re^m \to (-\infty, +\infty]$ are defined by $f^k(x) = f(x) + \langle p^{k+1}, Ax\rangle$ and $g^k(z) = g(z) - \langle p^{k+1}, z\rangle$, respectively.

Step 3: Let

$$y^{k+1} = y^k + c_k(Ax^{k+1} - z^{k+1}), \tag{13}$$

$k := k + 1$ and go to Step 1.

Note that Step 2 amounts to solving the following two convex minimization problems approximately:

$$\min_{x \in \Re^n} \; f(x) + \langle p^{k+1}, Ax \rangle + \frac{1}{c_k} D_\varphi(x, x^k), \tag{14}$$

$$\min_{z \in \Re^m} \; g(z) - \langle p^{k+1}, z \rangle + \frac{1}{c_k} D_\psi(z, z^k). \tag{15}$$

Of course, if $a_k = b_k = 0$, then x^{k+1} and z^{k+1} are the exact optimal solutions of these subproblems. Note that, by Lemma 2, the generated sequences $\{x^k\}$ and $\{z^k\}$ are guaranteed to lie in the open sets S_X and S_Z, respectively.

4 Convergence Analysis

In this section, we establish global convergence of the NPCPMM under the the following assumptions:

(A1) Problem (P) has an optimal solution (x^*, z^*) and a corresponding Lagrange multiplier y^*.

(A2) $\mathrm{dom}\, f \subseteq \overline{S_X}$ and $\mathrm{dom}\, g \subseteq \overline{S_Z}$.

(A3) $\sum_{k=0}^{\infty}(a_k + b_k) < \infty$.

Assumption (A1) implies that the Lagrangian L has a saddle point (x^*, z^*, y^*), and Assumption (A2) together with the boundary coerciveness of Bregman functions φ and ψ implies that the sequence $\{(x^k, z^k)\}$ generated by the NPCPMM is contained in $(\mathrm{dom}\, f \cap S_X) \times (\mathrm{dom}\, g \cap S_Z)$ by virtue of Lemma 2. Assumption (A3) requires approximate solutions (x^{k+1}, γ^{k+1}) and (z^{k+1}, δ^{k+1}) of subproblems (14) and (15), respectively, to be increasingly accurate as the iteration proceeds. The subsequent convergence analysis follows a line of argument to the one given in [9].

Lemma 3 *Suppose that $\{x^k\}$, $\{z^k\}$, $\{y^k\}$ and $\{p^k\}$ are sequences generated by the NPCPMM. Then, for all $x \in \mathrm{dom}\, f$, $z \in \mathrm{dom}\, g$ and $y \in \Re^m$, the following inequalities are satisfied:*

$$c_k\{L(x^{k+1}, z^{k+1}, p^{k+1}) - L(x, z, p^{k+1})\} \leq \{D_\varphi(x, x^k) - D_\varphi(x, x^{k+1})\}$$
$$+\{D_\psi(z, z^k) - D_\psi(z, z^{k+1})\} - \{D_\varphi(x^{k+1}, x^k) + D_\psi(z^{k+1}, z^k)\} + c_k(a_k + b_k), \quad (16)$$
$$2c_k\{L(x^k, z^k, y) - L(x^k, z^k, p^{k+1})\} \leq \|y^k - y\|^2 - \|p^{k+1} - y\|^2 - \|p^{k+1} - y^k\|^2, \quad (17)$$
$$2c_k\{L(x^{k+1}, z^{k+1}, y) - L(x^{k+1}, z^{k+1}, y^{k+1})\}$$
$$\leq \|y^k - y\|^2 - \|y^{k+1} - y\|^2 - \|y^{k+1} - y^k\|. \quad (18)$$

Proof From (11), (12) and Lemma 1, we have for all $x \in \mathrm{dom}\, f$ and $z \in \mathrm{dom}\, g$

$$c_k\{f^k(x^{k+1}) - f^k(x)\} \leq D_\varphi(x, x^k) - D_\varphi(x, x^{k+1}) - D_\varphi(x^{k+1}, x^k) + c_k a_k, \quad (19)$$
$$c_k\{g^k(z^{k+1}) - g^k(z)\} \leq D_\psi(z, z^k) - D_\psi(z, z^{k+1}) - D_\psi(z^{k+1}, z^k) + c_k b_k. \quad (20)$$

Then we obtain (16) simply by adding (19) and (20). Next observe that (10) and (13) are equivalent to the following two equations, respectively:

$$p^{k+1} = \operatorname*{arg\,min}_{y\in\Re^m} \left\{-L(x^k, z^k, y) + \frac{1}{2c_k}\|y - y^k\|^2\right\},$$
$$y^{k+1} = \operatorname*{arg\,min}_{y\in\Re^m} \left\{-L(x^{k+1}, z^{k+1}, y) + \frac{1}{2c_k}\|y - y^k\|^2\right\}.$$

Again by Lemma 1, for all $y \in \Re^m$, we obtain (17) and (18). ∎

Throughout the rest of this paper, we denote $w = (x, z, y)$ and define the function $\phi : \overline{S_X} \times \overline{S_Z} \times \Re^m \to \Re$ by

$$\phi(w) = \phi(x, z, y) = \varphi(x) + \psi(z) + \frac{1}{2}\|y\|^2.$$

Note that ϕ is also a Bregman function with D-function

$$D_\phi(w, w') = D_\varphi(x, x') + D_\psi(z, z') + \frac{1}{2}\|y - y'\|^2. \quad (21)$$

We will state a convergence theorem for the NPCPMM after showing some propositions.

Proposition 1 *Let $\{w^k\}$ and $\{p^k\}$ be sequences generated by the NPCPMM, and let (x^*, z^*) be an optimal solution of (P) and y^* be a corresponding Lagrange multiplier.*

Then we have for each k

$$D_\phi(w^*, w^{k+1}) \leq D_\phi(w^*, w^k) - \left\{D_{\varphi_0}(x^{k+1}, x^k) + \frac{1}{2}(\mu - 4c_k^2\|A\|^2)\|x^{k+1} - x^k\|^2\right\}$$
$$- \left\{D_{\psi_0}(z^{k+1}, z^k) + \frac{1}{2}(\nu - 4c_k^2)\|z^{k+1} - z^k\|^2\right\}$$
$$-\frac{1}{2}\{\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\} + c_k(a_k + b_k), \quad (22)$$

where $w^* = (x^*, z^*, y^*)$*. In particular, we have for all k*

$$D_\phi(w^*, w^{k+1}) \leq D_\phi(w^*, w^k) + c_k(a_k + b_k). \quad (23)$$

Proof By substituting (x^*, z^*) for (x, z) in (16), we have

$$c_k\{L(x^{k+1}, z^{k+1}, p^{k+1}) - L(x^*, z^*, p^{k+1})\} \leq \{D_\varphi(x^*, x^k) - D_\varphi(x^*, x^{k+1})\}$$
$$+\{D_\psi(z^*, z^k) - D_\psi(z^*, z^{k+1})\} - \{D_\varphi(x^{k+1}, x^k) + D_\psi(z^{k+1}, z^k)\} + c_k(a_k + b_k). \quad (24)$$

Since (x^*, z^*, y^*) is a saddle point of L, we obtain from (1)

$$c_k\{L(x^*, z^*, p^{k+1}) - L(x^{k+1}, z^{k+1}, y^*)\} \leq 0. \quad (25)$$

By adding (24) and (25) and rearranging terms, we get

$$D_\varphi(x^*, x^{k+1}) + D_\psi(z^*, z^{k+1}) \leq D_\varphi(x^*, x^k) + D_\psi(z^*, z^k)$$
$$-c_k\langle p^{k+1} - y^*, Ax^{k+1} - z^{k+1}\rangle - \{D_\varphi(x^{k+1}, x^k) + D_\psi(z^{k+1}, z^k)\} + c_k(a_k + b_k). \quad (26)$$

Now, by substituting y^{k+1} for y in (17) and y^* for y in (18), we have

$$2c_k\{L(x^k, z^k, y^{k+1}) - L(x^k, z^k, p^{k+1})\}$$
$$\leq \|y^k - y^{k+1}\|^2 - \|p^{k+1} - y^{k+1}\|^2 - \|p^{k+1} - y^k\|^2 \quad (27)$$

and

$$2c_k\{L(x^{k+1}, z^{k+1}, y^*) - L(x^{k+1}, z^{k+1}, y^{k+1})\}$$
$$\leq \|y^k - y^*\|^2 - \|y^{k+1} - y^*\|^2 - \|y^{k+1} - y^k\|, \quad (28)$$

respectively. Adding (27) and (28) and rearranging terms yield

$$\frac{1}{2}\|y^{k+1} - y^*\|^2 \leq \frac{1}{2}\|y^k - y^*\|^2 - \frac{1}{2}\{\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\}$$
$$-c_k\{\langle y^* - y^{k+1}, Ax^{k+1} - z^{k+1}\rangle + \langle y^{k+1} - p^{k+1}, Ax^k - z^k\rangle\}. \quad (29)$$

By adding (26) and (29), and taking into account (21), we have

$$
\begin{aligned}
D_\phi(w^*, w^{k+1}) &\le D_\phi(w^*, w^k) - \{D_\varphi(x^{k+1}, x^k) + D_\psi(z^{k+1}, z^k)\} \\
&\qquad -\frac{1}{2}\{\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\} \\
&\qquad +c_k\langle y^{k+1} - p^{k+1}, A(x^{k+1} - x^k) - (z^{k+1} - z^k)\rangle + c_k(a_k + b_k) \\
&= D_\phi(w^*, w^k) - \{D_\varphi(x^{k+1}, x^k) + D_\psi(z^{k+1}, z^k)\} \\
&\qquad -\frac{1}{2}\{\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\} \\
&\qquad +c_k^2\|A(x^{k+1} - x^k) - (z^{k+1} - z^k)\|^2 + c_k(a_k + b_k), \qquad (30)
\end{aligned}
$$

where the last equality follows from (10) and (13). Noting the inequality

$$
\|A(x^{k+1} - x^k) - (z^{k+1} - z^k)\|^2 \le 2\{\|A\|^2\|x^{k+1} - x^k\|^2 + \|z^{k+1} - z^k\|^2\},
$$

we obtain from (30)

$$
\begin{aligned}
D_\phi(w^*, w^{k+1}) &\le D_\phi(w^*, w^k) - \{D_\varphi(x^{k+1}, x^k) - 2c_k^2\|A\|^2\|x^{k+1} - x^k\|^2\} \\
&\qquad -\{D_\psi(z^{k+1}, z^k) - 2c_k^2\|z^{k+1} - z^k\|^2\} \\
&\qquad -\frac{1}{2}\{\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\} + c_k(a_k + b_k). \qquad (31)
\end{aligned}
$$

Since (7), (8) and the definition of D-function yield

$$
D_\varphi(x^{k+1}, x^k) = D_{\varphi_0}(x^{k+1}, x^k) + \frac{\mu}{2}\|x^{k+1} - x^k\|^2
$$

and

$$
D_\psi(z^{k+1}, z^k) = D_{\psi_0}(z^{k+1}, z^k) + \frac{\nu}{2}\|z^{k+1} - z^k\|^2,
$$

the desired inequality (22) follows from (31). Then the inequality (23) immediately follows from (22) and (9). ∎

The following lemma is well-known.

Lemma 4 *[18] Let $\{\alpha_k\}$ and $\{\beta_k\}$ be nonnegative sequences such that*

$$
\alpha_{k+1} \le \alpha_k + \beta_k \quad \textit{for all } k
$$

and

$$
\sum_{k=0}^{\infty} \beta_k < \infty.
$$

Then $\{\alpha_k\}$ is convergent.

Proposition 2 *Let $\{w^k\}$ be a sequence generated by the NPCPMM. Then $\{w^k\}$ is bounded and every accumulation point of $\{w^k\}$ is a saddle point of the Lagrangian L.*

Proof First note that Assumption (A3) together with (9) ensures

$$\sum_{k=0}^{\infty} c_k(a_k + b_k) < \infty. \tag{32}$$

Then, from (23), we have $w^k \in \mathcal{L}(w^*, \bar{\alpha}) = \{w \mid D_\phi(w^*, w) \leq \bar{\alpha}\}$ for all k, where $\bar{\alpha} = D_\phi(w^*, w^0) + \sum_{k=0}^{\infty} c_k(a_k + b_k)$. This implies $\{w^k\}$ is bounded because of condition (iii) in Definition 2.

Now let (x^*, z^*) be an optimal solution of (P), y^* be a corresponding Lagrange multiplier, and $w^* = (x^*, z^*, y^*)$. Then, from (23) in Proposition 1, (32), Lemma 4, and Assumption (A3), there exists $\beta \geq 0$ such that $\lim_{k\to\infty} D_\phi(w^*, w^k) = \beta$. Therefore, by taking the limits on both sides of (22) and rearranging terms, we obtain

$$\begin{aligned}\limsup_{k\to\infty} \Bigg[\left\{ D_{\varphi_0}(x^{k+1}, x^k) + \frac{1}{2}(\mu - 4c_k^2\|A\|^2)\|x^{k+1} - x^k\|^2 \right\} \\ + \left\{ D_{\psi_0}(z^{k+1}, z^k) + \frac{1}{2}(\nu - 4c_k^2)\|z^{k+1} - z^k\|^2 \right\} \\ + \frac{1}{2}\{\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\} \Bigg] \leq 0.\end{aligned} \tag{33}$$

It then follows from (33) and (9) that

$$\begin{aligned} &D_{\varphi_0}(x^{k+1}, x^k) \to 0, && \|x^{k+1} - x^k\| \to 0, \\ &D_{\psi_0}(z^{k+1}, z^k) \to 0, && \|z^{k+1} - z^k\| \to 0, \\ &\|p^{k+1} - y^{k+1}\| \to 0, && \|p^{k+1} - y^k\| \to 0. \end{aligned}$$

Now let w^∞ be an arbitrary accumulation point of $\{w^k\}$, and $\{w^k\}_{k\in K}$ a subsequence converging to w^∞. Then, from condition (iv) in Definition 2, we obtain

$$\lim_{k\to\infty,\, k\in K} D_\phi(w^\infty, w^k) = 0.$$

By taking the limit in (16) on the subsequence, we have for all $x \in \mathrm{dom}\, f$ and $z \in \mathrm{dom}\, g$

$$L(x^\infty, z^\infty, y^\infty) - L(x, z, y^\infty) \leq 0. \tag{34}$$

Similarly, by taking the limit in (18), we have for all $y \in \Re^m$

$$L(x^\infty, z^\infty, y) - L(x^\infty, z^\infty, y^\infty) \leq 0. \tag{35}$$

From (34) and (35), we obtain for all $x \in \mathrm{dom}\, f$, $z \in \mathrm{dom}\, g$ and $y \in \Re^m$

$$L(x^\infty, z^\infty, y) \leq L(x^\infty, z^\infty, y^\infty) \leq L(x, z, y^\infty).$$

Therefore, w^∞ is a saddle point of L. ∎

Proposition 3 *Let $\{w^k\}$ be a sequence generated by the NPCPMM. Then $\{w^k\}$ is convergent.*

Proof The proof is based on a standard argument. Suppose $\{w^k\}$ has two accumulation points w_1^∞ and w_2^∞ such that $w_1^\infty \neq w_2^\infty$, and let $\{w^k\}_{k \in K_1}$ and $\{w^k\}_{k \in K_2}$ be subsequences that converge to w_1^∞ and w_2^∞, respectively. Then, from condition (iv) in Definition 2, we have

$$\lim_{k \to \infty,\, k \in K_1} D_\phi(w_1^\infty, w^k) = \lim_{k \to \infty,\, k \in K_2} D_\phi(w_2^\infty, w^k) = 0. \tag{36}$$

Since (23) yields $0 \leq D_\phi(w_1^\infty, w^{k+1}) \leq D_\phi(w_1^\infty, w^k) + c_k(a_k + b_k)$, it follows from (A3) and Lemma 4 that $D_\phi(w_1^\infty, w^k)$ converges and its limit is equal to zero by (36). Therefore we have $\lim_{k \to \infty\, k \in K_2} D_\phi(w_1^\infty, w^k) = D_\phi(w_1^\infty, w_2^\infty) = 0$, which implies $w_1^\infty = w_2^\infty$. But this is a contradiction and hence $\{w^k\}$ has a unique limit point. ∎

From the above propositions, we obtain the following convergence theorem for the NPCPMM.

Theorem 1 *Let $\{w^k\}$ be a sequence generated by the NPCPMM. Then $\{w^k\}$ converges to a solution of (P).*

Proof Immediate from Propositions 2 and 3. ∎

5 Conclusion

In this paper, we have proposed the new decomposition algorithm called NPCPMM for solving a class of convex programming problems. The algorithm is an extension of the PCPMM of Chen and Teboulle [9], combined with the Bregman-based PMAD. We have established global convergence of the NPCPMM under appropriate assumptions. Further research topics may include convergence rate analysis, computational experiments with the NPCPMM, and extension of the NPCPMM to solve more general problems such as variational inequalities.

During the Workshop on Continuous Optimization held in Rio de Janeiro, June 21-25, 1999, we have learned from Prof. Auslender that Auslender and Teboulle [3] have developed a similar decomposition method independently and almost simultaneously. Their method is designed to solve variational inequality problems with separable structure and is based on the logarithmic-quadratic proximal method in studied in [2].

References

[1] A. Auslender, M. Teboulle and S. Ben-Tiba, Interior proximal and multiplier methods based on second order homogeneous kernels, *Mathematics of Operations Research*, to appear.

[2] A. Auslender, M. Teboulle and S. Ben-Tiba, A logarithmic-quadratic proximal method for variational inequalities, *Computational Optimization and Applications*, Vol. 12, pp. 31-40, 1999.

[3] A. Auslender and M. Teboulle, Entropic proximal decomposition methods for convex programs and variational inequalities, manuscript, Laboratoire d'Econometrie de L'Ecole Polytechnique, Paris, June 1999.

[4] H.H. Bauschke and J.M. Borwein, Legendre functions and the method of random Bregman projections, *Journal of Convex Analysis*, Vol. 4, pp. 27-67, 1997.

[5] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, NY, 1982.

[6] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Englewood Cliffs, NJ, 1989.

[7] R.S. Burachik and A.N. Iusem, A generalized proximal point algorithm for the variational inequality problem in a Hilbert space, *SIAM Journal on Optimization*, Vol. 8, pp. 197-216, 1998.

[8] Y. Censor and S.A. Zenios, *Parallel Optimization*, Oxford University Press, New York, NY, 1997.

[9] G. Chen and M. Teboulle, A proximal-based decomposition method for convex minimization problems, *Mathematical Programming*, Vol. 64, pp. 81-101, 1994.

[10] J. Eckstein, Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming, *Mathematics of Operations Research*, Vol.18, pp. 202-226, 1993.

[11] J. Eckstein, Approximate iterations in Bregman-function-based proximal algorithms, *Mathematical Programming*, Vol. 83, pp. 113-123, 1998.

[12] J. Eckstein and M. Fukushima, Some reformulations and applications of the alternating direction method of multipliers, W.W. Hager, D.W. Hearn and P.M. Pardalos, eds., *Large Scale Optimization: State of the Art*, Kluwer, Boston, MA, pp. 115-134, 1994.

[13] M. Fukushima, Application of the alternating direction method of multipliers to separable convex programming problems, *Computational Optimization and Applications*, Vol. 1, pp. 93-112, 1992.

[14] D. Gabay, Applications of the method of multipliers to variational inequalities, M. Fortin and R. Glowinski, eds., *Augmented Lagrangian Method: Applications to the Numerical Solution of Boundary-Value Problems*, North-Holland, Amsterdam, pp. 299-331, 1983.

[15] K.C. Kiwiel, Proximal minimization methods with generalized Bregman functions, *SIAM Journal on Control and Optimization*, Vol. 35, pp. 1142-1168, 1997.

[16] P.L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM Journal on Numerical Analysis*, Vol. 16, pp. 964-979, 1979.

[17] B. Martinet, Regularisation d'inéquations variationnelles par approximations successive, *Revue Francaise d'Automatique et Informatique Recherche Opérationnelle*, Vol. 4, pp. 154-159, 1970.

[18] B.T. Polyak, *Introduction to Optimization*, Optimization Software Inc., New York, NY, 1987.

[19] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[20] R.T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM Journal on Control and Optimization*, Vol. 14, pp. 877-898, 1976.

[21] R.T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, *Mathematics of Operations Research*, Vol. 1, pp. 97-116, 1976.

[22] M.V. Solodov and B.F. Svaiter, An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions, manuscript, Instituto de Mathemática Pura e Aplicada, Rio de Janeiro, May 1998.

[23] M. Teboulle, Convergence of proximal-like algorithms, *SIAM Journal on Optimization*, Vol. 7, pp. 1069-1083, 1997.

[24] P. Tseng, Applications of a splitting algorithm to decomposition in convex programming and variational inequalities, *SIAM Journal on Control and Optimization*, Vol. 29, pp. 119-138, 1991.