J. Nonlinear Var. Anal. 6 (2022), No. 2, pp. 5-14
Available online at http://jnva.biemdas.com
https://doi.org/10.23952/jnva.6.2022.2.02

ON THE LINEAR CONVERGENCE OF A BREGMAN PROXIMAL POINT ALGORITHM

KE GUO*, CHUNRONG ZHU

School of Mathematics and Information, China West Normal University, Nanchong 637002, China

Abstract. In this paper, we study a Bregman proximal point algorithm (BPPA) for convex optimization problems. Though the convergence and sublinear convergence rate for BPPA are well-understand, the linear convergence rate for BPPA has yet been thoroughly studied in the literature. In this paper, we analyze the linear convergence rate of BPPA. Under the assumption that the objective function is strongly convex relative to a Legendre function, we establish the linear convergence for the function values sequence. Moreover, if the Legendre function is strongly convex and smooth, the linear convergence for the iterative sequence of BPPA is obtained.

Keywords. Bregman proximal point algorithm; Linear convergence; Relatively strongly convex.

1. Introduction

Consider the following optimization problem

$$\min_{x} f(x), \tag{1.1}$$

where $f:\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is a proper, lower semicontinuous, and convex function. We assume that the solution set of problem (1.1), denoted by X^*, is nonempty. Proximal point algorithm (PPA) is a very fundamental algorithm for optimization problems. Based on the early work of Minty [1] and Moreau [2], the PPA was promoted to the optimization community [3]. Among different algorithms for solving problem (1.1), PPA is the most fundament one and its iteration reads as follows:

$$x_{k+1} = \arg\min_{x} \left\{ f(x) + \frac{1}{2\lambda_k}\|x - x_k\|^2 \right\}, \tag{1.2}$$

where $\lambda_k > 0$ is the proximal parameter. Under the conditions that $X^* \neq \emptyset$ and the proximal parameter is selected such that $\{\lambda_k\} \subset [\lambda, \infty)$ with $\lambda > 0$, the convergence and convergence rate of PPA (1.2) were analyzed in the literature [4–8]. In 1992, Censor and Zenios [9] first proposed a PPA with Bregman regularization (Bregman PPA) and the concrete iterative scheme is listed as follows:

$$x_{k+1} = \arg\min_{x} \left\{ f(x) + \frac{1}{\lambda_k} D_h(x, x_k) \right\}, \tag{1.3}$$

*Corresponding author.
E-mail addresses: keguo2014@126.com (K. Guo), chunrongz1997@163.com (C. Zhu).
Received September 10, 2021; Accepted February 14, 2022.

where $D_h(\cdot,\cdot)$ is the Bregman distance function defined by

$$D_h(x,y) = h(x) - h(y) - \langle \nabla h(x), x - y \rangle$$

for some given Legendre function h. Note that, the Bregman distance function is reduced to Euclidean distance if the Legendre function $h(x) = \frac{1}{2}\|x\|^2$. Thus, BPPA (1.3) can be viewed as an extension of the classic PPA (1.2). As we mentioned before, the convergence of PPA (1.2) has been extensively studied in the literature. However, the convergence of BPPA (1.3) has only been studied in few works [10–14]. To the best of our knowledge, the linear convergence of BPPA (1.3) is still absent in the literature. In this paper, we aim to fill this gap by establishing its linear convergence in terms of function value sequences and iterative sequences.

The remainder of this paper is organized as follows. In Section 2, we recall some important definitions and some known results for further analysis. Then the linear convergence for Bregman PPA under relatively strongly convex is established in Section 3. Finally, some conclusions are drawn in Section 4.

2. Preliminary

This section contains some definitions and basic results that will be used in our subsequent analysis.

Definition 2.1. [15] For an extended real-valued function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, the effective domain or just the domain is the set

$$\mathrm{dom}(f) = \{x \in \mathbb{R}^n : f(x) < \infty\}.$$

Definition 2.2. [15] A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be proper if there exists at least one $x \in \mathbb{R}^n$ such that $f(x) < \infty$.

Definition 2.3. [15] A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be lower semicontinuous at $x \in \mathbb{R}^n$ if

$$f(x) \le \liminf_{k \to \infty} f(x_k)$$

for any sequence $\{x_k\} \subseteq \mathbb{R}^n$ for which $x_k \to x$ as $k \to \infty$. Moreover, f is said to be lower semicontinuous if it is lower semicontinuous at each point in $\mathbb{R}^n$.

Definition 2.4. [16] An operator $S : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous with constant $\beta \in \mathbb{R}_+$ if, for every $x, y \in \mathbb{R}^n$, $\|S(x) - S(y)\| \le \beta \|x - y\|$.

Definition 2.5 (Legendre Function). [16] Let $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper, lower semicontinuous, and convex function. It is said to be:

(i) essentially smooth if h is differentiable on $\mathrm{int}(\mathrm{dom}(h))$ with $\|\nabla h(x_k)\| \to \infty$ for every sequence $\{x_k\}_{k\in\mathbb{N}} \subset \mathrm{int}(\mathrm{dom}(h))$ converging to a boundary point of $\mathrm{dom}(h)$ as $k \to +\infty$;

(ii) of Legendre type if h is essentially smooth and strictly convex on $\mathrm{int}(\mathrm{dom}(h))$.

It is known from [16] that h is of Legendre type if and only if its conjugate h^* is of Legendre type. Moreover, the gradient of a Legendre function h is a bijection from $\mathrm{int}(\mathrm{dom}(h))$ to $\mathrm{int}(\mathrm{dom}(h^*))$ and its inverse is the gradient of the conjugate, that is,

$$(\nabla h)^{-1} = \nabla h^*, \text{ and } h^*(\nabla h(x)) = \langle x, \nabla h(x) \rangle - h(x).$$

Moreover,

$$\mathrm{dom}(\partial h) = \mathrm{int}(\mathrm{dom}(h)), \text{ with } \partial h(x) = \{\nabla h(x)\}, \; \forall x \in \mathrm{int}(\mathrm{dom}(h)).$$

Now, we introduce the fundamental proximity measure associated to any given Legendre function h.

Definition 2.6. [17] Let $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a Legendre function. The Bregman distance associated with h is the function $D_h(\cdot,\cdot) : \mathrm{dom}(h) \times \mathrm{int}(\mathrm{dom}(h)) \to \mathbb{R}$ given by

$$D_h(x,y) = h(x) - h(y) - \langle \nabla h(y), x-y \rangle, \; \forall (x,y) \in \mathrm{dom}(h) \times \mathrm{int}(\mathrm{dom}(h)). \tag{2.1}$$

The Bregman distance is a proximity measure in the sense that $D_h(x,y) \geq 0$. From the strict convexity of $D_h(\cdot,y)$, we have

$$D_h(x,y) = 0 \quad \Leftrightarrow \quad x = y, \; \forall (x,y) \in \mathrm{dom}(h) \times \mathrm{int}(\mathrm{dom}(h)).$$

Indeed, the proximity measure defined through the Bregman distance $D_h(\cdot,\cdot)$ has been extensively studied. For early developments, the examples of Bregman distances, as well and many other useful properties, we refer to, e.g., [9, 12, 18–23] and the references therein.

Lemma 2.1. *[24, Proposition 2.3] Let $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a Legendre function. The following properties follow directly from (2.1). Let $x, y \in int(dom(h))$ and $z \in dom(h)$. Then*

(i) $D_h(x,y) + D_h(y,x) = \langle \nabla h(x) - \nabla h(y), x-y \rangle$;

(ii) *the three-point identity holds:*

$$D_h(z,x) - D_h(z,y) - D_h(y,x) = \langle \nabla h(x) - \nabla h(y), y-z \rangle.$$

Definition 2.7. Let $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a Legendre function, and let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper lower semicontinuous. We say that $f(\cdot)$ is μ-strongly convex relative to $h(\cdot)$ if, for any $x \in \mathrm{dom}(\partial f) \cap \mathrm{int}(\mathrm{dom}(h)), y \in \mathrm{dom}(f)$ and $g_x \in \partial f(x)$, there is a scalar $\mu \geq 0$ such that

$$f(y) \geq f(x) + \langle g_x, y-x \rangle + \mu D_h(y,x).$$

Remark 2.1. Definition 2.7 is a slight extension of the definition of the relatively strong convexity for smooth functions [25, Definition 1.2] to the nonsmooth functions. Note that, a relatively strongly convex function is always convex, and especially, f is strongly convex when $h(\cdot) = \frac{1}{2}\|\cdot\|^2$.

Similarly, we can define the relatively strong monotonicity as follows.

Definition 2.8. Let $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be an operator, and let $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a Legendre function. Then A is μ-strongly monotone relative to $h(\cdot)$ if, for any $x, y \in \mathrm{dom}(A) \cap \mathrm{int}(\mathrm{dom}(h)), g_x \in A(x), g_y \in A(y)$,

$$\langle g_x - g_y, x-y \rangle \geq \mu \langle \nabla h(x) - \nabla h(y), x-y \rangle.$$

Especially, when $h(\cdot) = \frac{1}{2}\|\cdot\|^2$, we say that A is μ-strongly monotone.

The following proposition establishes the relation between the relatively strong monotonicity and the relatively strong convexity.

Proposition 2.1. *Let $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a Legendre function, and let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper, lower semicontinuous, and μ-strongly convex relative to $h(\cdot)$. Then ∂f is μ-strongly monotone relative to $h(\cdot)$.*

Proof. Since f is μ-strongly convex relative to $h(\cdot)$, so for any $x \in \mathrm{dom}(\partial f) \cap \mathrm{int}(\mathrm{dom}(h))$, $y \in \mathrm{dom}(f)$ and $g_x \in \partial f(x)$ we have

$$f(y) \geq f(x) + \langle g_x, y - x \rangle + \mu D_h(y, x), \tag{2.2}$$

Changing the roles of x and y yields the inequality

$$f(x) \geq f(y) + \langle g_y, x - y \rangle + \mu D_h(x, y), \tag{2.3}$$

where $y \in \mathrm{dom}(\partial f) \cap \mathrm{int}(\mathrm{dom}(h))$, $x \in \mathrm{dom}(f)$ and $g_y \in \partial f(y)$. Adding inequalities (2.2) and (2.3), we can finally conclude that

$$\langle g_x - g_y, x - y \rangle \geq \mu (D_h(y, x) + D_h(x, y)) = \mu \langle \nabla h(x) - \nabla h(y), x - y \rangle,$$

where the equation follows from (i) of Lemma 2.1. The proof is complete. □

Finally, we list some following Lemmas that will be used in our analysis.

Lemma 2.2. [26, Proposition 3.9] *Let D be a nonempty subset of $\mathbb{R}^n$, let $K : D \to \mathbb{R}^n$, and let $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. If $K + M$ is maximally monotone and strongly monotone, then ran$(K) \subset$ ran$(K + M)$ and $K + M$ is injective.*

Lemma 2.3. [26, Proposition 3.8] *Let D be a nonempty subset of $\mathbb{R}^n$, let $K : D \to \mathbb{R}^n$, and let $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be such that ran$(K) \subset$ ran$(K + M)$ and $K + M$ is injective. Then $J_M^K = (K + M)^{-1} \circ K : D \to D$.*

Lemma 2.4. [27, Example 20.31] *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be cocoereive with constant $\beta > 0$. Then T is maximally monotone.*

Lemma 2.5. [27, Theorem 20.25] *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper lower semicontinuous convex. Then ∂f is maximally monotone.*

Lemma 2.6. [27, Corollary 25.5] *Let $T_1 : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and $T_2 : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be maximally monotone operators such that dom$(T_1) = \mathbb{R}^n$. Then $T_1 + T_2$ is maximally monotone.*

3. Linear Convergence for the Bregman PPA

Recall that the Bregman PPA has the following update formula

$$x_{k+1} = \arg\min_x \left\{ f(x) + \frac{1}{\lambda_k} D_h(x, x_k) \right\}, \tag{3.1}$$

where h is a Legendre function, and D_h is the Bregman distance. Since h is strictly convex, one has that $f(x) + \frac{1}{\lambda_k} D_h(x, x_k)$ is also strictly convex. Thus, the solution set of the righthand side of (3.1) is at most single-valued. Without loss of generality, we may assume that it is nonempty to ensure the well-definedness of the sequence $\{x_k\}$. Actually, this assumption can be relaxed under stronger assumptions on h, which will be elaborated later on.

Theorem 3.1. *Suppose that $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is proper, lower semicontinuous, and μ-strongly convex relative to a Legendre function $h(\cdot)$. Let $\{x_k\}$ be the sequence generated by the Bregman PPA (3.1), and let the proximal parameter sequence $\{\lambda_k\}$ is bounded from zero ($\lambda_k \geq \lambda > 0$). Then, for any $x^* \in X^*$, $\{f(x_k)\}$ is linearly convergent to $f(x^*)$ and $D_h(x^*, x_k)$ is linearly convergent to 0.*

Proof. By the optimality condition of Bregman PPA (3.1), we have

$$\frac{1}{\lambda_k}(\nabla h(x_k) - \nabla h(x_{k+1})) \in \partial f(x_{k+1}). \tag{3.2}$$

Since f is strongly convex relative to $h(\cdot)$, we obtain

$$f(x) \geq f(x_{k+1}) + \frac{1}{\lambda_k}\langle \nabla h(x_k) - \nabla h(x_{k+1}), x - x_{k+1}\rangle + \mu D_h(x, x_{k+1}). \tag{3.3}$$

By setting $x = x_k$ and $x = x^*$ in (3.3), respectively, we obtain the following two inequalities

$$\begin{aligned} f(x_k) &\geq f(x_{k+1}) + \frac{1}{\lambda_k}\langle \nabla h(x_k) - \nabla h(x_{k+1}), x_k - x_{k+1}\rangle + \mu D_h(x_k, x_{k+1}) \\ &= f(x_{k+1}) + \frac{1}{\lambda_k}(D_h(x_k, x_{k+1}) + D_h(x_{k+1}, x_k)) + \mu D_h(x_k, x_{k+1}) \\ &= f(x_{k+1}) + \left(\frac{1}{\lambda_k} + \mu\right) D_h(x_k, x_{k+1}) + \frac{1}{\lambda_k} D_h(x_{k+1}, x_k) \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} f(x^*) &\geq f(x_{k+1}) - \frac{1}{\lambda_k}\langle \nabla h(x_k) - \nabla h(x_{k+1}), x_{k+1} - x^*\rangle + \mu D_h(x^*, x_{k+1}) \\ &= f(x_{k+1}) - \frac{1}{\lambda_k}(D_h(x^*, x_k) - D_h(x^*, x_{k+1}) - D_h(x_{k+1}, x_k)) + \mu D_h(x^*, x_{k+1}) \\ &= f(x_{k+1}) - \frac{1}{\lambda_k} D_h(x^*, x_k) + \left(\frac{1}{\lambda_k} + \mu\right) D_h(x^*, x_{k+1}) + \frac{1}{\lambda_k} D_h(x_{k+1}, x_k), \end{aligned} \tag{3.5}$$

where the first equation in (3.4) follows from (i) of Lemma 2.1 and the first equation in (3.5) follows from (ii) of Lemma 2.1. For any $\theta_k \in [0,1]$, multiplying (3.4) by $(1-\theta_k)$ and (3.5) by θ_k and adding the two resulting inequalities, we obtain that

$$\begin{aligned} & f(x_{k+1}) - f(x^*) - (1-\theta_k)(f(x_k) - f(x^*)) \\ & \leq -(1-\theta_k)\left(\frac{1}{\lambda_k} + \mu\right) D_h(x_k, x_{k+1}) + \frac{\theta_k}{\lambda_k} D_h(x^*, x_k) \\ & \quad - \theta_k \left(\frac{1}{\lambda_k} + \mu\right) D_h(x^*, x_{k+1}) - \frac{1}{\lambda_k} D_h(x_{k+1}, x_k). \end{aligned}$$

Since both $D_h(x_k, x_{k+1})$ and $D_h(x_{k+1}, x_k)$ are nonnegative, the above inequality implies that

$$\begin{aligned} & f(x_{k+1}) - f(x^*) + \theta_k \left(\frac{1}{\lambda_k} + \mu\right) D_h(x^*, x_{k+1}) \\ & \leq (1-\theta_k)(f(x_k) - f(x^*)) + \frac{\theta_k}{\lambda_k} D_h(x^*, x_k). \end{aligned} \tag{3.6}$$

Now, if we set $\theta_k = \frac{\lambda_k \mu}{1+\lambda_k \mu}$, then (3.6) yields

$$\begin{aligned} f(x_{k+1}) - f(x^*) + \mu D_h(x^*, x_{k+1}) &\leq \frac{1}{1+\lambda_k \mu} \cdot \{f(x_k) - f(x^*) + \mu D_h(x^*, x_k)\} \\ &\leq \frac{1}{1+\lambda \mu} \cdot \{f(x_k) - f(x^*) + \mu D_h(x^*, x_k)\}, \end{aligned}$$

where the second inequality is due to $\lambda_k \geq \lambda > 0$. Thus, we obtain

$$f(x_k) - f(x^*) + \mu D_h(x^*, x_k) \leq \left(\frac{1}{1+\lambda\mu}\right)^k \cdot \{f(x_0) - f(x^*) + \mu D_h(x^*, x_0)\}. \tag{3.7}$$

The proof is complete. □

Remember that, when $h(x) = \frac{1}{2}\|x\|^2$, Bregman PPA (3.1) reduces to the classic PPA (1.2), while the relatively strong convexity is nothing but the strong convexity. Then we obtain the following result.

Corollary 3.1. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}^n \cup \{+\infty\}$ is proper, lower semicontinuous, and μ-strongly convex relative to the Legendre function $h(x) = \frac{1}{2}\|x\|^2$. Let $\{x_k\}$ be the sequence generated by the classic PPA (1.2), and let the proximal parameter sequence $\{\lambda_k\}$ be bounded from zero ($\lambda_k \geq \lambda > 0$). Then, for any $x^* \in X^*$, $\{f(x_k)\}$ is linearly convergent to $f(x^*)$ and $\{x_k\}$ is linearly convergent to x^*.*

Proof. Note that $D_h(x,y) = \frac{1}{2}\|x-y\|^2$ when $h(x) = \frac{1}{2}\|x\|^2$. We know from (3.7) that

$$f(x_k) - f(x^*) + \frac{\mu}{2}\|x_k - x^*\|^2 \leq \left(\frac{1}{1+\lambda\mu}\right)^k \cdot \{f(x_0) - f(x^*) + \frac{\mu}{2}\|x_0 - x^*\|^2\}.$$

Since f is strongly convex, for any $x^* \in X^*$, we have $f(x_0) - f(x^*) \geq \frac{\mu}{2}\|x_0 - x^*\|^2$. Thus, the conclusion immediately follows from the above two inequalities. □

Next, we focus on the R-linear convergence for the sequence $\{D_h(x^*, x_k)\}$.

Theorem 3.2. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}^n \cup \{+\infty\}$ is proper lower semicontinuous and μ-strongly convex relative to a Legendre function $h(\cdot)$. Let $\{x_k\}$ be the sequence generated by the Bregman PPA (3.1), and let the proximal parameter sequence $\{\lambda_k\}$ be bounded from zero ($\lambda_k \geq \lambda > 0$). Then, for any $x^* \in X^*$, $\{D_h(x^*, x_k)\}$ is R-linearly convergent to 0.*

Proof. For any $x^* \in X^*$, we have $0 \in \partial f(x^*)$. Then it follows from (3.2) and Proposition 2.1 that

$$\left\langle \frac{1}{\lambda_k}(\nabla h(x_k) - \nabla h(x_{k+1})), x_{k+1} - x^* \right\rangle \geq \mu \langle \nabla h(x_{k+1}) - \nabla h(x^*), x_{k+1} - x^* \rangle.$$

In view of Lemma 2.1, the above inequality can be rewritten as

$$\frac{1}{\lambda_k}(D_h(x^*, x_k) - D_h(x^*, x_{k+1}) - D_h(x_{k+1}, x_k)) \geq \mu(D_h(x_{k+1}, x^*) + D_h(x^*, x_{k+1})).$$

Since both $D_h(x_{k+1}, x^*)$ and $D_h(x_{k+1}, x_k)$ are nonnegative, the above inequality yields

$$D_h(x^*, x_{k+1}) \leq \frac{1}{1+\lambda_k\mu} \cdot D_h(x^*, x_k) \leq \frac{1}{1+\lambda\mu} \cdot D_h(x^*, x_k),$$

where the second inequality follows from $\lambda_k \geq \lambda > 0$. the proof is complete. □

Remark 3.1. When $h(\cdot) = \frac{1}{2}\|\cdot\|^2$ (that is f is strongly convex), we deduce from Theorem 3.2 that

$$\|x_k - x^*\| \leq \left(\frac{1}{\sqrt{1+\lambda\mu}}\right)^k \cdot \|x_0 - x^*\|. \tag{3.8}$$

This means the rate of linear convergence of $\{x_k\}$ to x^* is $\frac{1}{\sqrt{1+\lambda\mu}}$, which is weaker than classic result whose rate is $\frac{1}{1+\lambda\mu}$. More precisely, for the case that f is μ-strongly convex, it is known from [8] that

$$\|x_k - x^*\| \leq \left(\frac{1}{1+\lambda\mu}\right)^k \cdot \|x_0 - x^*\|. \tag{3.9}$$

Next, we analyze the possible reasons. Indeed, by the optimality condition, the Bregman PPA (3.1) can be rewritten as

$$x_{k+1} = (\nabla h + \lambda_k \partial f)^{-1}(\nabla h(x_k)). \tag{3.10}$$

For simplicity, let us denote $J_{\lambda f}^{\nabla h} = (\nabla h + \lambda \partial f)^{-1} \circ \nabla h$, which is called the D-resolvent in the literature [24]. When $h(x) = \frac{1}{2}\|x\|^2$, $J_{\lambda f}^{\nabla h}$ reduces to the proximal mapping of f:

$$\text{prox}_{\lambda f} = (I + \lambda \partial f)^{-1},$$

which is single-valued everywhere. Moreover, when f is μ-strongly convex, $\text{prox}_{\lambda f}$ is contractive with constant $\frac{1}{1+\lambda\mu}$. From this viewpoint, inequality (3.9) can be obtained immediately. However, when f is μ-strongly convex relative to a Legendre function h, we can not derive the similar property of $J_{\lambda f}^{\nabla h}$. This is because Legendre function h is only strictly convex and essentially smooth. If we additionally assume that h is strongly convex and smooth (which holds trivially when $h(x) = \frac{1}{2}\|x\|^2$), the rate $\frac{1}{\sqrt{1+\lambda\mu}}$ established in (3.8) can be improved to $\frac{1}{1+\lambda\mu}$ naturally.

Next, we prove a key result, which can be viewed as an extension of the classic contraction property of the proximal mapping of strongly convex functions to the relatively strongly convex setting. In addition, the following proposition guarantees that (3.1) is well-defined without assuming that its solution set is nonempty.

Proposition 3.1. *Suppose that $h : \mathbb{R}^n \to \mathbb{R}$ is strongly convex with constant $\alpha > 0$ and smooth with constant $\beta > 0$, and $f : \mathbb{R}^n \to \mathbb{R}^n \cup \{+\infty\}$ is proper lower semicontinuous and μ-strongly convex relative to $h(\cdot)$. Then $J_{\lambda f}^{\nabla h} = (\nabla h + \lambda \partial f)^{-1} \circ \nabla h : \mathbb{R}^n \to \mathbb{R}^n$ is single-valued. Moreover, $J_{\lambda f}^{\nabla h}$ is Lipschitz continuous with $\frac{\beta}{\alpha(1+\mu\lambda)}$. Furthermore, if $\lambda > \frac{\beta-\alpha}{\alpha\mu}$, then $J_{\lambda f}^{\nabla h}$ is contractive.*

Proof. Denote $K = \nabla h$ and $M = \partial f$. We first prove that $J_{\lambda f}^{\nabla h}$ is single-valued. On one hand, by the convexity and smoothness of h, it follows from the Baillon-Haddad theorem [27, Corollary 18.17] that K is cocoercive with constant $1/\beta$, which implies that K is maximally monotone in view of Lemma 2.4. Note that, a relatively strongly convex function is also a convex function, Lemma 2.5 yields that M is maximally monotone. Now, let us further denote $T_1 = K$ and $T_2 = M$. Since $\text{dom}(T_1) = \mathbb{R}^n$, it follows from Lemma 2.6 that $K + M$ is maximally monotone. On the other hand, the strong convexity of h implies K is strongly monotone. Since M is monotone, one has that $K + M$ is strongly monotone. Therefore, the single-valuedness of $J_{\lambda f}^{\nabla h}$ follows from Lemmas 2.2 and 2.3 immediately.

Now, for any $x, y \in \mathbb{R}^n$, by denoting $p = J_{\lambda f}^{\nabla h}(x)$ and $q = J_{\lambda f}^{\nabla h}(y)$, we have that

$$\frac{1}{\lambda}(\nabla h(x) - \nabla h(p)) \in \partial f(p), \quad \frac{1}{\lambda}(\nabla h(y) - \nabla h(q)) \in \partial f(q).$$

Again since f is μ-strongly convex relative to $h(\cdot)$, it follows from Proposition 2.1 that ∂f is μ-strongly monotone relative to $h(\cdot)$, that is, for any $x,y\in\mathbb{R}^n$,

$$\langle \nabla h(x)-\nabla h(y)-(\nabla h(p)-\nabla h(q)),p-q\rangle \geq \mu\lambda\langle \nabla h(p)-\nabla h(q),p-q\rangle,$$

which is equivalent to

$$\langle \nabla h(x)-\nabla h(y),p-q\rangle \geq (1+\mu\lambda)\langle \nabla h(p)-\nabla h(q),p-q\rangle. \tag{3.11}$$

Then, we deduce that

$$\begin{aligned}\alpha\|p-q\|^2 \leq \langle p-q,\nabla h(p)-\nabla h(q)\rangle &\leq \frac{1}{1+\mu\lambda}\cdot\langle p-q,\nabla h(x)-\nabla h(y)\rangle\\ &\leq \frac{1}{1+\mu\lambda}\cdot\|p-q\|\cdot\|\nabla h(x)-\nabla h(y)\|\\ &\leq \frac{\beta}{1+\mu\lambda}\cdot\|p-q\|\cdot\|x-y\|,\end{aligned}$$

where the first inequality follows from the strong monotonicity of ∇h, the second inequality follows from inequality (3.11), and the last inequality follows from the smoothness of h. Thus

$$\|J_{\lambda f}^{\nabla h}(x)-J_{\lambda f}^{\nabla h}(y)\| = \|p-q\| \leq \frac{\beta}{\alpha(1+\mu\lambda)}\cdot\|x-y\|, \quad \forall x,y\in\mathbb{R}^n.$$

This proves the second part. For the remaining part, if $\lambda > \frac{\beta-\alpha}{\alpha\mu}$, we know that $\frac{\beta}{\alpha(1+\mu\lambda)}\in(0,1)$. Hence $J_{\lambda f}^{\nabla h}$ is contractive. The proof is complete. □

Remark 3.2. If $h(x)=\frac{1}{2}\|x\|^2$, then $\alpha=\beta=1$ and $J_{\lambda f}^{\nabla h}=\mathrm{prox}_{\lambda f}$. In this case, we recover the classic $\frac{1}{1+\mu\lambda}$-contraction property of $\mathrm{prox}_{\lambda f}$ for the strongly convex function h. See [27, Proposition 23.13] for more details. Remember that, as we discussed in Remark 3.1, this contraction constant is vital to characterize the linear rate of classic PPA.

Compared to Theorem 3.2, armed with the contraction of $J_{\lambda f}^{\nabla h}$, we can prove the linear convergence of the sequence $\{x_k\}$ to its solution set. This also leads to an improved rate in comparison to (3.8) for the case $h(x)=\frac{1}{2}\|x\|^2$.

Theorem 3.3. *Let $\{x_k\}$ be the sequence generated by the Bregman PPA (3.1). Suppose that f is μ-strongly convex relative to a Legendre function $h(\cdot)$ with constant $\mu>0$, and $h(\cdot)$ is strongly convex with constant $\alpha>0$ and smooth with constant $\beta>0$. Assume that the proximal parameter sequence $\{\lambda_k\}$ satisfies $\lambda_k\geq\lambda>\frac{\beta-\alpha}{\alpha\mu}$. Then, for any $x^*\in X^*$,*

$$\|x_k-x^*\| \leq \left(\frac{\beta}{\alpha(1+\lambda\mu)}\right)^k\cdot\|x_0-x^*\|.$$

More precisely, when $h(\cdot)=\frac{1}{2}\|\cdot\|^2$, we have

$$\|x_k-x^*\| \leq \left(\frac{1}{1+\lambda\mu}\right)^k\cdot\|x_0-x^*\|.$$

Proof. In view of (3.10), it follows from Proposition 3.1 that the Bregman PPA (3.1) can be written as

$$x_{k+1} = J_{\lambda_k f}^{\nabla h}(x_k).$$

For any $x^* \in X^*$ and $\lambda_k > 0$, we have

$$0 \in \partial f(x^*) \Leftrightarrow \nabla h(x^*) \in \nabla h(x^*) + \lambda_k \partial f(x^*) \Leftrightarrow x^* = J_{\lambda_k f}^{\nabla h}(x^*).$$

Thus,

$$\|x_k - x^*\| = \|J_{\lambda_k f}^{\nabla h}(x_{k-1}) - J_{\lambda_k f}^{\nabla h}(x^*)\| \leq \frac{\beta}{\alpha(1+\lambda_k \mu)} \cdot \|x_{k-1} - x^*\| \leq \frac{\beta}{\alpha(1+\lambda \mu)} \cdot \|x_{k-1} - x^*\|,$$

which implies that

$$\|x_k - x^*\| \leq \left(\frac{\beta}{\alpha(1+\lambda \mu)}\right)^k \cdot \|x_0 - x^*\|.$$

More precisely, when $h(x) = \frac{1}{2}\|x\|^2$, we obtain $\alpha = \beta = 1$. The proof is complete. □

4. CONCLUSIONS

In this paper, we established the linear convergence of the Bregman proximal point algorithm (BPPA) for convex optimization problems under the relatively strong convexity assumption. For the general Legendre function, we proved the linear convergence for the function value sequence. If the Legendre function was further assumed to be strongly convex and smooth, the linear convergence for the iterative sequence of BPPA was obtained. Since the proximal point algorithm can be relaxed, we will study the linear convergence of the generalized BPPA in the future.

Acknowledgments

This first author was supported by the Natural Science Foundation of China (Grant Nos. 11801455, 11871059, 11971238) and the Open Project of Key Laboratory (Grant No. CSSXKFKTM202004), School of Mathematical Sciences, Chongqing Normal University, China.

REFERENCES

[1] G. Minty, Monotone operators in Hilbert space, Duke Math. J. 29 (1962), 341-346.
[2] J. Moreau, Proximité et dualité dans un espace hilbertien, Bulletin de la Socit Chimique de France, 93 (1965), 273-299.
[3] B. Martinet, Regularization d'inequations variationelles par approximations successives, Revue Francaise d'Informatique et de Recherche Opérationelle, 4 (1970), 154-159.
[4] X.J. Cai, K. Guo, F. Jiang, K. Wang, Z.M. Wu, D.R. Han, The developments of proximal point algorithms, J. Oper. Res. Soc. China, doi: https://doi.org/10.1007/s40305-021-00352-x.
[5] M. Fukushima, H. Mine, A generalized proximal point algorithm for certain non-convex minimization problems, Int. J. Sys. Sci. 12 (1981), 989-1000.
[6] O. Güler, On the convergence of the proximal point algorithm for convex minimization, SIAM J. Control Optim. 29 (1991), 403-419.
[7] O. Güler, New proximal point algorithms for convex minimization, SIAM J. Optim. 2 (1992), 649-664.
[8] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Control Optim. 14 (1986), 877-898.

[9] Y. Censor, S.A. Zenios, Proximal minimization algorithm with D-functions, J. Optim. Theory Appl. 73 (1992), 451-464.
[10] A. Auslender, M. Haddou, An interior-proximal method for convex linearly constrained problems and its extension to variational inequalities, Math. Program. 71 (1995), 77-100.
[11] G. Chen, M. Teboulle, Convergence analysis of a proximal-like minimization algorithm using Bregman functions, SIAM J. Optim. 3 (1993), 538-543.
[12] J. Eckstein, Nonlinear proximal point algorithms using bregman functions, with applications to convex programming, Math. Oper. Res. 18 (1993), 202-226.
[13] A.N. Iusem, B.F. Svaiter, M. Teboulle, Entropy-like proximal methods in convex programming, Math. Oper. Res. 19 (1994), 790-814.
[14] K.C. Kiwiel, Proximal minimization methods with generalized Bregman functions, SIAM J. Control Optim. 35 (1997), 1142-1168.
[15] A. Beck, First Order Methods in Optimization, SIAM, Philadephia, 2017.
[16] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, 1970.
[17] L.M. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, USSR Comput. Math. Math. phys. 7 (1967), 200-217.
[18] H.H. Bauschke, J.M. Borwein, Legendre functions and the method of random Bregman projections, J. Convex Anal. 4 (1997), 27-67.
[19] H.H. Bauschke, J.M. Borwein, Joint and separate convexity of the Bregman distance, Stud. Comput. Math. 8 (2001), 23-36.
[20] J. Bolte, M. Teboulle, Barrier operators and associated gradient like dynamical systems for constrained minimization problems, SIAM J. Control Optim. 42 (2003), 1266-1292.
[21] R. Burachik, A. Iusem, A generalized proximal point algorithm for the variational inequality problem in a Hilbert space, SIAM J. Optim. 8 (1998), 197-216.
[22] A.N. Iusem, On dual convergence and the rate of primal convergence of Bregman convex programming method, SIAM J. Optim. 1 (1991), 401-423.
[23] M. Teboulle, Entropic proximal mappings with application to nonlinear programming, Math. Oper. Res. 17 (1992), 670-690.
[24] H.H. Bauschke, J.M. Borwein, P.L. Combettes, Bregman monotone optimization algorithms, SIAM J. Control Optim. 42 (2003), 596-636.
[25] H.H. Lu, R.M. Freund, Y. Nesterov, Relatively smooth convex optimization by first-order methods, and applications, SIAM J. Optim. 28 (2018), 333-354.
[26] M.B. Bùi, P.L. Combettes, Warped proximal iterations for monotone inclusions, J. Math. Anal. Appl. 491 (2020), 124-135.
[27] H.H. Bauschke, P.L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Space, Springer-Verlag, New York, 2011.