On the Convergence and O(1/N) Complexity of a Class of Nonlinear Proximal Point Algorithms for Monotonic Variational Inequalities

Jian Wu [1,] and Gaohang Yu [1, *]

[1] *Key laboratory of numerical simulation technology of Jiangxi Province, School of Mathematics and Computer Sciences, Gannan Normal University, China.*

Abstract This paper presents a class of proximal point algorithms (PPA) with nonlinear proximal terms. Proximal minimization algorithm using Bregman distance for convex minimization is extended for solving monotonic variational inequalities. Under suitable conditions, the convergence and $O(1/N)$ computing complexity/convergence rate of the proposed algorithm is obtained. Further more, connections to some existing popular methods are given, which shows that the presented algorithm can include metric proximal point algorithm and projection method within a general form.

1. Introduction

Variational inequality (VI) has received a lot of attention due to its various applications in operation research, economic equilibrium, engineering design, and other fields [5, 6].

In this paper, we will study iterative algorithms for monotonic VI problems, which can be summarized in a form as follows: find a point $\boldsymbol{u}^*$ in Ω such that

$$\langle \boldsymbol{u} - \boldsymbol{u}^*, \mathcal{F}(\boldsymbol{u}^*) \rangle \geq 0, \forall \boldsymbol{u} \in \Omega, \tag{1}$$

---

*Correspondence to: School of Mathematics and Computer Sciences, Gannan Normal University, Ganzhou, 341000, China. email: maghyu@163.com

where $\mathcal{F} : \mathbb{R}^d \to \mathbb{R}^d$ is a mapping from a Euclidian space $\mathbb{R}^d$ to itself, and Ω a convex subset in $\mathbb{R}^d$.

VI has an important application in optimization. Let $F : \mathbb{R}^d \to \mathbb{R}$ be a differentiable convex function on Ω. Then any minimum point $\boldsymbol{x}^*$ of F on Ω satisfies the inequality

$$\langle \boldsymbol{x} - \boldsymbol{x}^*, \nabla F(\boldsymbol{x}^*) \rangle \geq 0, \forall \boldsymbol{x} \in \Omega, \tag{2}$$

which means that the VI problem (1) includes as a subproblem the following optimization problem

$$\min_{\boldsymbol{x} \in \Omega} F(\boldsymbol{x}).$$

In this paper we are interested in a class of so-called monotonic VIs in the sense of

$$\langle \boldsymbol{u} - \boldsymbol{v}, \mathcal{F}(\boldsymbol{u}) - \mathcal{F}(\boldsymbol{v}) \rangle \geq 0, \forall \boldsymbol{u}, \boldsymbol{v} \in \Omega, \tag{3}$$

for two considerations. Firstly this class of VIs has a nice property that its solution set is nonempty; secondly, it responds to the convex program in the optimization field.

There is extensive literature on numerical algorithms for VIs [1, 2, 4, 7, 10]. Among these solvers, a classical algorithm widely known as proximal point algorithm (PPA) (or proximal minimization algorithm specially for minimization problems [3]) is first proposed in [8] and then developed in [9]. PPA solves the main problem (1) by successive approximation, or in another word, solves in every iteration a subproblem proximal to the main problem. Say

$$\langle \boldsymbol{u} - \boldsymbol{u}^{n+1}, \mathcal{F}(\boldsymbol{u}^{n+1}) + M(\boldsymbol{u}^{n+1} - \boldsymbol{u}^n) \rangle \geq 0, \forall \boldsymbol{u} \in \Omega, \tag{4}$$

or

$$\boldsymbol{x}^{n+1} = \arg\min_{\boldsymbol{x} \in \Omega} F(\boldsymbol{x}) + \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{x}^n\|_M^2, \tag{5}$$

where M is a symmetric and positive-definite matrix, and the proximal term in (5) is known as Mahalanobis distance defined as

$$\frac{1}{2} \|\boldsymbol{x} - \boldsymbol{y}\|_M^2 := \frac{1}{2} \langle \boldsymbol{x} - \boldsymbol{y}, M(\boldsymbol{x} - \boldsymbol{y}) \rangle.$$

All aforementioned PPAs so far are using a linear proximal term, and the main aim of this paper is to present a class of PPAs using nonlinear proximal terms, and to provide a theoretic justification of its convergence and computing complexity.

The rest of this paper is organized as follow: Section 2 is devoted to present the algorithm, and the proof of its convergence and complexity analysis are given in Section 3. Section 4 makes several connections between the proposed algorithm and other popular methods. Finally, we have a concluding remark.

2. Motivation and Proposed Algorithm

Suppose f is a strongly function on X, the Bregman distance induced by f is defined as below:

$$D_f(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{x}) - f(\boldsymbol{y}) - \langle \boldsymbol{x} - \boldsymbol{y}, \nabla f(\boldsymbol{y}) \rangle, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in X,$$

where $\nabla f(\boldsymbol{y})$ is a subgradient of f at $\boldsymbol{x}$.

Bregman distance is an extension of traditional Euclidian or Mahalanobis distance, this point can be seen more clearly thorough some special cases of Bregman distance:

1. $D_f(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|^2$, when $f(\cdot) = \frac{1}{2}\|\cdot\|^2$;
2. $D_f(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_M^2$, when $f(\cdot) = \frac{1}{2}\|\cdot\|_M^2$;
3. $D_f(\boldsymbol{x}, \boldsymbol{y}) = -\sum_{j=1}^d x_j \left(\log \frac{x_j}{y_j} - 1 \right) + \sum_{j=1}^d y_j,$ when $f(\boldsymbol{x}) = \sum_{j=1}^d x_j \log \frac{x_j}{a_j}.$

It is a natural idea to replace the Mahalanobis distance in (5) with a Bregman distance, which is the main contribution presented in [3]. Then the proximal minimization algorithm with a D-function, so called by authors of [3], is proposed as

$$\boldsymbol{x}^{n+1} = \arg \min_{\boldsymbol{x} \in \Omega} F(\boldsymbol{x}) + D_f(\boldsymbol{x}, \boldsymbol{x}^n). \tag{6}$$

According the optimality condition of (6), we have

$$\langle \boldsymbol{x} - \boldsymbol{x}^{n+1}, \nabla F(\boldsymbol{x}^{n+1}) + \nabla f(\boldsymbol{x}^{n+1}) - \nabla f(\boldsymbol{x}^n) \rangle \geq 0, \forall \boldsymbol{x} \in \Omega,$$

which can be viewed as a PPA for (2). It is worthy of notice that the proximal term in the inequality above is not necessarily linear. Enlighten by this, we propose a nonlinear PPA for the general monotonic VI (1) as follow

$$\langle \boldsymbol{u} - \boldsymbol{u}^{n+1}, \mathcal{F}(\boldsymbol{u}^{n+1}) + \mathcal{G}(\boldsymbol{u}^{n+1}) - \mathcal{G}(\boldsymbol{u}^n) \rangle \geq 0, \forall \boldsymbol{u} \in \Omega. \tag{7}$$

A remarkable difference between (6) and (7) is that in (6), ∇F needs to be the gradient of a convex function, while in (7) $\mathcal{F}$ needn't, which means the later is an extension of the former.

In another word, (7) can solve variational problems which are not derived from minimization problems, such as those derived from saddle point problems and complementary problems. For instance, suppose one is going to solve a saddle point problem

$$\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y}} \Phi(\boldsymbol{x}, \boldsymbol{y}). \tag{8}$$

By transforming it as a variational inequality problem

$$\left\langle \begin{bmatrix} \boldsymbol{x} - \boldsymbol{x}^* \\ \boldsymbol{y} - \boldsymbol{y}^* \end{bmatrix}, \begin{bmatrix} \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}^*, \boldsymbol{y}^*) \\ -\nabla_{\boldsymbol{y}} \Phi(\boldsymbol{x}^*, \boldsymbol{y}^*) \end{bmatrix} \right\rangle \geq 0, \forall \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \in \mathcal{X} \times \mathcal{Y}, \tag{9}$$

the proposed algorithm (7) can be applied, while application of algorithm (6) does not cover (8) or (9).

$\mathcal{G}$ in (7) still needs to be gradient/subgradient of a strongly convex function f, thus two assumptions upon $\mathcal{G}$ are necessary:

A1 $\mathcal{G}$ is strongly monotonic (from the strongly convexity of f);

A2 $\mathcal{G}$ is continuous.

This work looks similar to that in [4], but some differences are essential. At first, algorithm presented in [4] takes approximate iterations as

$$\left\langle \boldsymbol{u} - \boldsymbol{u}^{n+1}, \mathcal{F}(\boldsymbol{u}^{n+1}) + \mathcal{G}(\boldsymbol{u}^{n+1}) - \mathcal{G}(\boldsymbol{u}^n) + \boldsymbol{\epsilon}^{n+1} \right\rangle \geq 0, \forall \boldsymbol{u} \in \Omega,$$

while in this work, proximal subproblem is solved exactly. Secondly, as we will see in the coming section, this work provides a complexity analysis result, which is absent in [4].

3. Convergence and $O(1/N)$ Complexity

In this section, we will thoroughly study the convergence of the proposed algorithm. The analytic tool used here is similar with that in [8] but is more general in this paper.

Lemma 1
Let $\{\boldsymbol{u}^n\}$ be a sequence generated by the algorithm (7), and $\boldsymbol{u}^*$ a solution of the VI (1), then we have

$$\left\langle \boldsymbol{u}^* - \boldsymbol{u}^{n+1}, \mathcal{G}(\boldsymbol{u}^{n+1}) - \mathcal{G}(\boldsymbol{u}^n) \right\rangle \geq 0. \tag{10}$$

Proof. By setting $\boldsymbol{u} = \boldsymbol{u}^*$ in (7), we have

$$\left\langle \boldsymbol{u}^* - \boldsymbol{u}^{n+1}, \mathcal{G}(\boldsymbol{u}^{n+1}) - \mathcal{G}(\boldsymbol{u}^n) \right\rangle \geq \left\langle \boldsymbol{u}^{n+1} - \boldsymbol{u}^*, \mathcal{F}(\boldsymbol{u}^{n+1}) \right\rangle$$
$$\text{monotonicity of } \mathcal{F} \text{ (3)} \quad \geq \left\langle \boldsymbol{u}^{n+1} - \boldsymbol{u}^*, \mathcal{F}(\boldsymbol{u}^*) \right\rangle$$
$$\text{VI (1)} \quad \geq 0 \quad \square$$

Lemma 2
Let $\{\boldsymbol{u}^n\}$ be a sequence generated by the algorithm (7), and $\boldsymbol{u}^*$ a solution of the VI (1), then we have

$$D_f(\boldsymbol{u}^{n+1}, \boldsymbol{u}^n) \leq D_f(\boldsymbol{u}^*, \boldsymbol{u}^n) - D_f(\boldsymbol{u}^*, \boldsymbol{u}^{n+1}) \tag{11}$$

and

$$\left\langle \boldsymbol{u} - \boldsymbol{u}^{n+1}, \mathcal{G}(\boldsymbol{u}^{n+1}) - \mathcal{G}(\boldsymbol{u}^n) \right\rangle \leq D_f(\boldsymbol{u}, \boldsymbol{u}^n) - D_f(\boldsymbol{u}, \boldsymbol{u}^{n+1}), \forall \boldsymbol{u} \in \Omega. \tag{12}$$

Proof. By direct calculation we have

$$D_f(\boldsymbol{u}, \boldsymbol{u}^n) - D_f(\boldsymbol{u}, \boldsymbol{u}^{n+1})$$
$$= f(\boldsymbol{u}) - f(\boldsymbol{u}^n) - \langle \boldsymbol{u} - \boldsymbol{u}^n, \mathcal{G}(\boldsymbol{u}^n) \rangle - f(\boldsymbol{u}) + f(\boldsymbol{u}^{n+1}) + \langle \boldsymbol{u} - \boldsymbol{u}^{n+1}, \mathcal{G}(\boldsymbol{u}^{n+1}) \rangle$$
$$= f(\boldsymbol{u}^{n+1}) - f(\boldsymbol{u}^n) - \langle \boldsymbol{u}^{n+1} - \boldsymbol{u}^n, \mathcal{G}(\boldsymbol{u}^n) \rangle + \langle \boldsymbol{u} - \boldsymbol{u}^{n+1}, \mathcal{G}(\boldsymbol{u}^{n+1}) - \mathcal{G}(\boldsymbol{u}^n) \rangle$$
$$= D_f(\boldsymbol{u}^{n+1}, \boldsymbol{u}^n) + \langle \boldsymbol{u} - \boldsymbol{u}^{n+1}, \mathcal{G}(\boldsymbol{u}^{n+1}) - \mathcal{G}(\boldsymbol{u}^n) \rangle,$$

from which immediately we obtain (12). By setting $\boldsymbol{u} = \boldsymbol{u}^*$ in the equality above, and using (10) we can easily obtain (11) as well. $\quad\square$

Lemma 3
Let $\{\boldsymbol{u}^n\}$ be a sequence generated by the algorithm (7), then we have

$$\lim_{n\to\infty} \left\| \mathcal{G}(\boldsymbol{u}^{n+1}) - \mathcal{G}(\boldsymbol{u}^n) \right\| = 0 \tag{13}$$

Proof. From (11) we know that $\left\{ D_f(\boldsymbol{u}^*, \boldsymbol{u}^{n+1}) \right\}$ is monotonically non-increase and bounded below, thus has a limit, say d.

$$\lim_{n\to\infty} D_f(\boldsymbol{u}^{n+1}, \boldsymbol{u}^n) \leq \lim_{n\to\infty} \left(D_f(\boldsymbol{u}^*, \boldsymbol{u}^n) - D_f(\boldsymbol{u}^*, \boldsymbol{u}^{n+1}) \right) = d - d = 0.$$

Since f is strongly convex (**A1**), $\lim_{n\to\infty} D_f(\boldsymbol{u}^{n+1}, \boldsymbol{u}^n) = 0$ implies

$$\lim_{n\to\infty} \left\| \boldsymbol{u}^{n+1} - \boldsymbol{u}^n \right\| = 0.$$

According to the assumption **A2**, the assertion (13) is obtained. $\quad\square$

Theorem 1
Let $\{\boldsymbol{u}^n\}$ be a sequence generated by the algorithm (7), then it is bounded and any cluster point of $\{\boldsymbol{u}^n\}$ is a solution point of (1). Further more

$$\sup_{\boldsymbol{u}\in\Omega} \left\langle \bar{\boldsymbol{u}}^N - \boldsymbol{u}, \mathcal{F}(\boldsymbol{u}) \right\rangle \leq \frac{\alpha}{N} \tag{14}$$

where $\alpha = \sup \left\{ D_f(\boldsymbol{u}, \boldsymbol{u}^0) | \boldsymbol{u} \in \Omega \right\}$ and $\bar{\boldsymbol{u}}^N = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{u}^n$.

Proof. At first, the boundedness of $\{\boldsymbol{u}^n\}$ can be deduced from (11), hence a cluster point $\boldsymbol{u}^\infty$ of $\{\boldsymbol{u}^n\}$ exists. Secondly from (7) we have

$$\left\langle \boldsymbol{u} - \boldsymbol{u}^{n+1}, \mathcal{F}(\boldsymbol{u}^{n+1}) \right\rangle \geq - \left\langle \boldsymbol{u} - \boldsymbol{u}^{n+1}, \mathcal{G}(\boldsymbol{u}^{n+1}) - \mathcal{G}(\boldsymbol{u}^n) \right\rangle$$
$$\geq - \left\| \boldsymbol{u} - \boldsymbol{u}^{n+1} \right\| \left\| \mathcal{G}(\boldsymbol{u}^{n+1}) - \mathcal{G}(\boldsymbol{u}^n) \right\|.$$

Computing lower limits of both sides of the inequality above, we have

$$\lim_{n\to\infty} \inf \left\langle \boldsymbol{u} - \boldsymbol{u}^{n+1}, \mathcal{F}(\boldsymbol{u}^{n+1}) \right\rangle \geq 0.$$

Subsequently we have

$$\langle \boldsymbol{u} - \boldsymbol{u}^\infty, \mathcal{F}(\boldsymbol{u}^\infty) \rangle \geq 0, \forall \boldsymbol{u} \in \Omega.$$

Hence $\boldsymbol{u}^\infty$ is a solution of (1).

Now we move to prove the last part of the theorem. From the monotonicity of $\mathcal{F}$ we have

$$
\begin{aligned}
\langle \boldsymbol{u}^{n+1} - \boldsymbol{u}, \mathcal{F}(\boldsymbol{u}) \rangle &\leq \langle \boldsymbol{u}^{n+1} - \boldsymbol{u}, \mathcal{F}(\boldsymbol{u}^{n+1}) \rangle \\
(7) \quad &\leq \langle \boldsymbol{u} - \boldsymbol{u}^{n+1}, \mathcal{G}(\boldsymbol{u}^{n+1}) - \mathcal{G}(\boldsymbol{u}^n) \rangle \\
(12) \quad &\leq D_f(\boldsymbol{u}, \boldsymbol{u}^n) - D_f(\boldsymbol{u}, \boldsymbol{u}^{n+1}).
\end{aligned}
$$

Summarize the inequality above from 0 to $N - 1$,

$$\left\langle \frac{1}{N} \sum_{n=1}^N \boldsymbol{u}^n - \boldsymbol{u}, \mathcal{F}(\boldsymbol{u}) \right\rangle \leq \frac{1}{N} \left(D_f(\boldsymbol{u}, \boldsymbol{u}^0) - D_f(\boldsymbol{u}, \boldsymbol{u}^N) \right) \leq \frac{D_f(\boldsymbol{u}, \boldsymbol{u}^0)}{N}.$$

Finally we get the result (14), which completes the proof. $\square$

The last assertion of the theorem means that the proposed algorithm has a complexity of $O(1/N)$.

4. Connection to Other Methods

This section is devoted to make some connections to some existing methods, which shows that our algorithm could include these methods in a uniformed way.

4.1. Linear Case

It is very obvious that when $\mathcal{G}$ is a linear mapping, say a matrix M, then algorithm (7) become a linear PPA (4). While what needs to be pointed out is that M must fulfill the assumptions **A1**, **A2**. Specially M must be positive-definite and bounded.

Now we verify that linear mapping $\mathcal{G}$ so defined fulfills the assumptions **A1**, **A2**. Assume the maximum and minimum eigenvalue of M respectively be $\lambda_{\max}$ and $\lambda_{\min}$, then we have

$$\langle \boldsymbol{u} - \boldsymbol{v}, \mathcal{G}(\boldsymbol{u}) - \mathcal{G}(\boldsymbol{v}) \rangle = \langle \boldsymbol{u} - \boldsymbol{v}, M(\boldsymbol{u} - \boldsymbol{v}) \rangle \geq \lambda_{\min} \|\boldsymbol{u} - \boldsymbol{v}\|^2$$

and

$$\|\mathcal{G}(\boldsymbol{u}) - \mathcal{G}(\boldsymbol{v})\| = \|M(\boldsymbol{u} - \boldsymbol{v})\| \leq \lambda_{\max} \|\boldsymbol{u} - \boldsymbol{v}\|$$

Hence we see that $\mathcal{G}$ is both strongly monotonic and Lipschitz continuous.

4.2. Entropy Case

It is very natural to present an entropy-like version of the algorithm (7) motivated by the third case of the D-function listed in Section 2,

$$D_f(\boldsymbol{u}, \boldsymbol{v}) = -\sum_{j=1}^{d} u_j \left(\log \frac{u_j}{v_j} - 1 \right) + \sum_{j=1}^{d} v_j.$$

By differentiate it over the variable $\boldsymbol{u}$, we obtain a corresponding nonlinear proximal term for our algorithm

$$\mathcal{G}(\boldsymbol{u}) - \mathcal{G}(\boldsymbol{v}) = \left(\log \frac{u_1}{v_1}, \cdots, \log \frac{u_d}{v_d} \right)^T. \tag{15}$$

However in this case, we need to restrict the domain to a positive orthant, say

$$\Omega \subset \left\{ (u_1, \cdots, u_d)^T \in \mathbb{R}^d : 0 < u_j \leq \frac{1}{c}, j = 1, \ldots, d \right\},$$

where c is a sufficiently small positive number.

The continuity of $\mathcal{G}$ is derived from its definition, and strongly monotonicity of $\mathcal{G}$ can be seen from following arguments:

$$\langle \boldsymbol{u} - \boldsymbol{v}, \mathcal{G}(\boldsymbol{u}) - \mathcal{G}(\boldsymbol{v}) \rangle = \sum_{j=1}^{d} (u_j - v_j)(\log u_j - \log v_j) = \sum_{j=1}^{d} \frac{1}{\eta_j} |u_j - v_j|^2$$

$$\geq c \sum_{j=1}^{d} |u_j - v_j|^2 = c \|\boldsymbol{u} - \boldsymbol{v}\|^2.$$

In the argument, a mean value theorem

$$|\log u_j - \log v_j| = \frac{1}{\eta_j} |u_j - v_j|, \quad \text{for some } \eta_j \in (u_j, v_j)$$

on each component is applied.

4.3. Projection Method: a Special Case

VI problem is closely related to the fixed-point problem. The fixed-point theory has played an important role in the development of various algorithms for solving VIs. In fact we have the following well-known result

Lemma 4
$\boldsymbol{u}^*$ is a solution of VI (1) if and only if

$$\boldsymbol{u}^* = \mathcal{P}_\Omega \left[\boldsymbol{u}^* - c\mathcal{F}(\boldsymbol{u}^*) \right], \forall c > 0,$$

where the operator $\mathcal{P}_\Omega(\cdot)$, call projection, is defined as

$$\mathcal{P}_\Omega(\boldsymbol{v}) = \arg\min \{\|\boldsymbol{u} - \boldsymbol{v}\| : \boldsymbol{u} \in \Omega\} . \tag{16}$$

The fixed-point formulation in the above lemma suggests the simple iterative algorithm solving (1)

$$\boldsymbol{u}^{n+1} = \mathcal{P}_\Omega \left[\boldsymbol{u}^n - c_n \mathcal{F}(\boldsymbol{u}^n)\right] . \tag{17}$$

This algorithm is widely known as projection method, and converges to a solution point of (1) provided

1. $\mathcal{F}$ is Lipschitz continuous with constant L;
2. $\mathcal{F}$ is strongly monotonic;
3. stepsize fulfills $c_n < 1/L$.

We now demonstrate how these convergence conditions coincide those of the proposed algorithm **A1**, **A2**. By applying the principle of (2) in the minimization problem in (16), one has

$$\langle \boldsymbol{u} - \mathcal{P}_\Omega(\boldsymbol{v}), \mathcal{P}_\Omega(\boldsymbol{v}) - \boldsymbol{v} \rangle \geq 0, \forall \boldsymbol{u} \in \Omega,$$

which leads to an important fact that

$$\left\langle \boldsymbol{u} - \boldsymbol{u}^{n+1}, \mathcal{F}(\boldsymbol{u}^n) + \frac{1}{c_n}(\boldsymbol{u}^{n+1} - \boldsymbol{u}^n) \right\rangle \geq 0, \forall \boldsymbol{u} \in \Omega,$$

or

$$\left\langle \boldsymbol{u} - \boldsymbol{u}^{n+1}, \mathcal{F}(\boldsymbol{u}^{n+1}) + \left(\frac{1}{c_n}\mathcal{I} - \mathcal{F}\right)(\boldsymbol{u}^{n+1}) - \left(\frac{1}{c_n}\mathcal{I} - \mathcal{F}\right)(\boldsymbol{u}^n) \right\rangle \geq 0, \forall \boldsymbol{u} \in \Omega.$$

From this point of view, the projection method (17) can be seen as a special case of (7) where $\mathcal{G} = \frac{1}{c_n}\mathcal{I} - \mathcal{F}$.

Now we move to justify that $\mathcal{G} = \frac{1}{c_n}\mathcal{I} - \mathcal{F}$ is continuous and strongly monotonic. Assume $\mathcal{F}$ is Lipschitz continuous with constant L, immediately we know the continuity of $\mathcal{G}$ from its definition. On the other hand, one has

$$\begin{aligned}
\langle \boldsymbol{u} - \boldsymbol{v}, \mathcal{G}(\boldsymbol{u}) - \mathcal{G}(\boldsymbol{v}) \rangle &= \left\langle \boldsymbol{u} - \boldsymbol{v}, \left(\frac{1}{c_n}\mathcal{I} - \mathcal{F}\right)(\boldsymbol{u}) - \left(\frac{1}{c_n}\mathcal{I} - \mathcal{F}\right)(\boldsymbol{v}) \right\rangle \\
&= \frac{1}{c_n}\|\boldsymbol{u} - \boldsymbol{v}\|^2 - \langle \boldsymbol{u} - \boldsymbol{v}, \mathcal{F}(\boldsymbol{u}) - \mathcal{F}(\boldsymbol{v}) \rangle \\
&\geq \frac{1}{c_n}\|\boldsymbol{u} - \boldsymbol{v}\|^2 - \|\boldsymbol{u} - \boldsymbol{v}\|\, \|\mathcal{F}(\boldsymbol{u}) - \mathcal{F}(\boldsymbol{v})\| \\
&\geq \frac{1}{c_n}\|\boldsymbol{u} - \boldsymbol{v}\|^2 - L\|\boldsymbol{u} - \boldsymbol{v}\|^2 \\
&= \left(\frac{1}{c_n} - L\right)\|\boldsymbol{u} - \boldsymbol{v}\|^2 .
\end{aligned}$$

Hence we know that $\mathcal{G}$ is strongly monotonic on the condition of $c_n < 1/L$.

5. Conclusion

In this work, we proposed a general nonlinear proximal point algorithm which can unify some popular methods, including metric proximal point algorithm and projection method. We gave the theoretic proof the convergence and convergence rate as well.

Although we have more choices to determine the proximal term $\mathcal{G}(\boldsymbol{u}) - \mathcal{G}(\boldsymbol{v})$ according to the special structure of the problem than the existing metric proximal point algorithm and projection method could do, the choice of $\mathcal{G}$ is still subjected to a strict restriction: it must be gradient/subgradient of a strongly convex function.

One strategy to overcome this shortcoming may be to replace Bregman distance by other type of distance, for example

$$D(\boldsymbol{u}, \boldsymbol{v}) = \langle \boldsymbol{u} - \boldsymbol{v}, \mathcal{G}(\boldsymbol{u}) - \mathcal{G}(\boldsymbol{v}) \rangle,$$

where $\mathcal{G}$ is strongly monotonic so that the restriction on $\mathcal{G}$ being a gradinet/subgradient can be loosen. A second issue worthy further consideration is that since in our argument only the difference $\mathcal{G}(\boldsymbol{u}) - \mathcal{G}(\boldsymbol{v})$ rather than $\mathcal{G}(\cdot)$ is involved, so can we consider a more general form $\mathcal{K}(\boldsymbol{u}, \boldsymbol{v})$ as the proximal term? Whether these ideas works and if so how they are realized are questions open to answers.

REFERENCES

1. Auslender A, Teboulle M. Lagrangian duality and related multiplier methods for variational inequality problems, *SIAM Journal on Optimization*, 2000, 10(4): 1097-1115.
2. Burke J V, Qian M. On the superlinear convergence of the variable metric proximal point algorithm using Broyden and BFGS matrix secant updating, *Mathematical Programming*, 2000, 88(1): 157-181.
3. Censor Y, Zenios S A. Proximal minimization algorithm withD-functions, *Journal of Optimization Theory and Applications*, 1992, 73(3): 451-464.
4. Eckstein J. Approximate iterations in Bregman-function-based proximal algorithms, *Mathematical Programming*, 1998, 83(1-3): 113-123.
5. Ferris M C, Pang J S. Engineering and economic applications of complementarity problems, *SIAM Review*, 1997, 39(4): 669-713.
6. He B. Inexact implicit methods for monotone general variational inequalities, *Mathematical Programming*, 1999, 86(1): 199-217.
7. He B., Fu X., Jiang Z. Proximal-point algorithm using a linear proximal term, *Journal of Optimization Theory and Applications*, 2009, 141(2): 299-319.
8. Martinet, B. Regularisation d'inequations variationnelles par approximations successives, *Revue francaise d'informatique et de recherche operationnelle*, serie rouge, 1970, 4(3): 154-158.
9. Rockafellar R T. Monotone operators and the proximal point algorithm, *SIAM Journal on Control and Optimization*, 1976, 14(5): 877-898.
10. Teboulle M. Convergence of proximal-like algorithms, *SIAM Journal on Optimization*, 1997, 7(4): 1069-1083.