Journal of Global Optimization **26:** 125–140, 2003.

A New Hybrid Generalized Proximal Point Algorithm for Variational Inequality Problems

DEREN HAN
School of Mathematics and Computer Science, Nanjing Normal University, Nanjing 210097, P.R. China
(e-mail: handroo@hotmail.com)

Abstract. In this paper, we propose a modified Bregman-function-based proximal point algorithm for solving variational inequality problems. The algorithm adopts a similar constructive approximate criterion as the one developed by Solodov and Svaiter (Set Valued Analysis 7 (1999) 323) for solving the classical proximal subproblems. Under some suitable conditions, we can get an approximate solution satisfying the accuracy criterion via a single Newton-type step. We obtain the Fejér monotonicity to solutions of VIP for paramonotone operators. Some preliminary computational results are also reported to illustrate the method.

AMS subject classifications: 90C30, 90C33, 65K05

Key words: Bregman functions; Inexact methods; Monotone operators; Proximal point algorithms; Variational inequalities

1. Introduction

A classical variational inequality problem, denoted by VIP(F, Ω) for simplicity, is to find a vector $x^* \in \Omega$, such that

$$\langle x - x^*, F(x^*)\rangle \geqslant 0\,, \quad \forall x \in \Omega\,, \tag{1}$$

where Ω is a nonempty closed convex subset of R^n, F is a mapping from R^n into itself, and $\langle \cdot, \cdot \rangle$ stands for the usual inner product in R^n.

It is well known that VIP(F, Ω) is closely related to the problem of finding a zero of a maximal monotone operator $T: R^n \to 2^{R^n}$

$$0 \in T(x)\,, \quad x \in R^n\,. \tag{2}$$

If the feasible set $\Omega = \mathrm{R}^n$, then VIP(F, Ω) reduces to (2) with $T = F$. On the other hand, (1) is a special case of (2) with $T(x) = F(x) + N_\Omega(x)$ and

$$N_\Omega(x) = \begin{cases} \{y : \langle x' - x, y\rangle \leqslant 0,\ \forall x' \in \Omega\}\,, & \text{if } x \in \Omega \\ \emptyset & \text{otherwise} \end{cases}$$

is the normal cone operator.

The proximal algorithm introduced by Martinet [20] for solving problem (2) generates a sequence $\{x^k\}$ by solving a series of proximal subproblems

$$0 \in c_k T(x) + x - x^k \,, \tag{3}$$

where $\{c_k\} \subset [c, \infty)$ with $c > 0$ is a sequence of regularization parameters. Since solving the subproblem (3) exactly can be quite difficult or even impossible in practice, it is essential to use approximate solutions in devising implementable algorithms. In [22], the inexact version of the method (3) was introduced

$$e^{k+1} + x^k \in c_k T(x^{k+1}) + x^{k+1} \,, \tag{4}$$

where e^{k+1} is the associated error term. Under the condition that $\Sigma_{k=0}^{\infty} \|e^k\| < \infty$, the method was proved convergent globally.

Much recent research has focused on the ‘‘nonlinear’’ generalization of the proximal point method [5, 9, 11, 13, 16–18]. That is, x^{k+1} is obtained by solving the generalized proximal point subproblem

$$0 \in c_k T(x) + \nabla h(x) - \nabla h(x^k) \,, \tag{5}$$

where $h: R^n \to R$ is a Bregman function [4], which is strictly convex, differentiable in the interior of Ω. When applied to VIP(F, Ω), the subproblems of the generalized proximal point algorithm are essentially systems of equations, for all information about the feasible set Ω is embedded in the function h. By contrast, subproblems in the classical proximal point algorithm are themselves nonlinear variational inequality problems, which are structurally considerable more difficult to solve than systems of equations. See [5] for detailed examples.

To make the generalized proximal algorithm more implementable, Eckstein [12] gave an inexact version of the method (5)

$$e^{k+1} + \nabla h(x^k) \in c_k T(x^{k+1}) + \nabla h(x^{k+1}) \,, \tag{6}$$

and proved that the generated sequence $\{x^k\}$ converges to a zero of T under the condition

$$\begin{aligned} &\sum_{k=0}^{\infty} \|e^k\| < \infty \quad \text{and} \\ &\sum_{k=0}^{\infty} \langle e^k, x^k \rangle \quad \text{exists and is finite} \,. \end{aligned} \tag{7}$$

Other inexact generalized proximal point algorithm are [6, 18, 26]. However, as discussed in [12] and [23], the approach of [12] is the simplest and easiest to use in practice. Still, the approximate criterion (7) is more restrictive than that for classical proximal point algorithm. Recently, Solodov and Svaiter [23] proposed a new generalized proximal point algorithm. At the kth step, they get a proximal solution $x^{k+1} \in \Omega$ by solving

$$0 \in c_k T(y^k) + \nabla h(y^k) - \nabla h(x^k) \tag{8}$$

and

$$x^{k+1} = \nabla h^{-1}(\nabla h(x^k) - c_k T(y^k)) \tag{9}$$

satisfying

$$D_h(y^k, x^{k+1}) \leqslant \sigma^2 D_h(y^k, x^k) , \tag{10}$$

where $D_h(\cdot, \cdot)$ is ''D-function'' which will be defined in the sequel and $\sigma \in (0, 1)$ is a constant. Note that the error tolerance (10) is more constructive than (7), because $\sigma \in (0, 1)$ is a constant. However, to verify if y^k is an acceptable proximal solution, we have first to solve the problem

$$\nabla h(x) = \nabla h(x^k) - c_k T(y^k) ,$$

to get a trial point x^{k+1}, which may be computational expensive in many cases when ∇h is difficult to invert. Their method therefore, has its advantage only for the case that ∇h^{-1} is easy to get.

In this paper, we propose a new Bregman function-based proximal point algorithm for solving variational inequality problems. The approximate criterion we adopt here is similar to that in [24] for classical proximal point algorithm. An advantage of this accuracy criterion is that, under some suitable conditions, the approximate solution y^k of the proximal subproblem can be obtained via only one Newton-type step. The proximal subproblem therefore, is not computational expensive. To ensure the convergence of the algorithm, we make a projection step to generate the next iteration. This is what our method ''hybrid'' is named after. Note that the projection step is not time consuming too, at least for the cases that the projection to Ω is easy to get, such as when Ω is the nonnegative orthant of R^n, a box or a ball. The method has both the advantages of classical proximal point algorithms and generalized proximal point algorithms. That is, the approximate criterion is as constructive as in [24] and the subproblems are essentially systems of equations.

The rest of the paper is organized as follows. In Section 2, we summarize some mathematical preliminaries about the underlying mapping F, the projection operator on Ω, and Bregman functions. In Section 3, the new algorithm is described and the global convergence is proved. In Section 4, we discuss a Newton-type method for the generalized proximal subproblem and prove that under some suitable conditions, we can get y^k via only one step. Some preliminary computational results are reported in Section 5 and some concluding remarks are given in Section 6.

Throughout this paper, for a given nonempty closed convex subset S of R^n, $\bar{S}$ will denote the closure, $\mathrm{int}(S)$ will denote the interior, and ∂S will denote the boundary of S. Furthermore, we assume that the solution set of VIP(F, Ω), denoted by Ω^*, is nonempty.

2. Preliminaries

In this section, we give some properties of the mappings $F(\cdot)$ and $h(\cdot)$ which will be used in the sequel.

Let $F(\cdot)$ be a mapping from R^n to itself. $F(\cdot)$ is said to be monotone, if

$$\langle F(x) - F(y), x - y\rangle \geqslant 0\,, \quad \forall x, y \in R^n\,.$$

Let $P_\Omega(\cdot)$ denote the projection from R^n onto Ω. That is, $P_\Omega(x)$ is the solution of the problem

$$\min_{y\in\Omega}\|x - y\|\,.$$

The projection operator $P_\Omega(\cdot)$ is nonexpansive, i.e.,

$$\|P_\Omega(x) - P_\Omega(y)\| \leqslant \|x - y\|\,, \quad \forall x, y \in R^n\,. \tag{11}$$

Let S be an open and convex subset of R^n and $\bar{S}$ be its closure. Given a strictly convex function h, finite at x, y and differentiable at y, we can measure a ''distance'' of sorts between x and y via the ''D-function'' $D_h(x, y)$: $\bar{S} \times S \to R$:

$$D_h(x, y) := h(x) - h(y) - \langle \nabla h(y), x - y\rangle\,. \tag{12}$$

It follows from the strict convexity of the function h that $D_h(x, y) \geqslant 0$ and $D_h(x, y) = 0$ if and only if $x = y$. We have the following property of the function $D_h(x, y)$.

LEMMA 2.1 ([9], Lemma 3.1). *Given* h: $R^n \to (-\infty, +\infty]$, $u \in \bar{S}$, *and* $s, t \in S$, *then*

$$D_h(u, s) = D_h(t, s) + D_h(u, t) + \langle \nabla h(t) - \nabla h(s), u - t\rangle\,. \tag{13}$$

Proof. It follows from the definition of the ''D-function''.

DEFINITION 2.2. Let S be a convex open subset of R^n, we say that h: $\bar{S} \to R$ is a Bregman function with zone S if the following conditions hold:

B1: h is continuous, and strictly convex in $\bar{S}$.
B2: h is continuously differentiable in S.
B3: Given any $x \in \bar{S}$ and $\alpha \in R$, the right partial level set

$$L(x, \alpha) := \{y \mid D_h(x, y) \leqslant \alpha\}$$

is bounded.
B4: If $\{y^k\} \subset S$ is a convergent sequence with limit y^*, then $D_h(y^*, y^k) \to 0$.

The original definition of a Bregman function also requires the left partial set

$$L'(\alpha, y) = \{x \in \bar{S} \mid D_h(x, y) \leqslant \alpha\}$$

to be bounded for any $y \in S$ and any $\alpha \in R$. It is known that this boundedness condition is extraneous regardless, since it is a consequence of Definition 2.2 (e.g., see [3]). In addition, Solodov and Svaiter [23] proved that the *convergence consistency* of h, required in the original definition of a Bregman function, is also a consequence of Definition 2.2.

We need the following assumptions to guarantee the generalized proximal subproblem solutions exist and belong to the interior of Ω.

A1. For any $x \in \mathrm{int}(\Omega)$ and $c > 0$, the generalized proximal subproblem

$$0 \in c(T(\cdot) + \nabla h(\cdot) - \nabla h(x)) + (\cdot - x)$$

has a solution.

A2. If $\{y^k\} \subset S$ converges to a point on the boundary of S and $x \in S$, then

$$\lim_{k\to\infty} \langle \nabla h(y^k), y^k - x\rangle = +\infty .$$

Assumption A2 is called boundary coerciveness. A simple sufficient condition for A1 and A2 is the zone coerciveness of h.

DEFINITION 2.3. The Bregman function h is called zone coercive if, for any $y \in R^n$, there exists an $x \in S$, such that $\nabla h(x) = y$.

It follows from boundary coerciveness of h and the definition of the ''*D*-function'' that

LEMMA 2.4. *For all $u \in S$ and all sequence $\{x^k\} \subset S$ converging to a point $\bar{x} \in \partial S$, we have*

$$\lim_{k\to\infty} D_h(u, x^k) = \infty .$$

3. The algorithm and its convergence

In this section, we first describe our generalized proximal point algorithm and then analyze its convergence.

Algorithm: inexact generalized proximal point algorithm.

S0. Choose a Bregman function h satisfying Assumptions A1 and A2, which zone is the interior of Ω. Choose some $0 < \underline{c} < \bar{c} < +\infty$, $t \in (0, 1)$, the error tolerance parameter $\sigma \in (0, 1)$, and $x^0 \in \mathrm{int}(\Omega)$. Set $k := 0$.

S1. Choose the regularization parameter $c_k \in [\underline{c}, \bar{c}]$, find the inexact solution $y^k \in \mathrm{int}(\Omega)$ of the proximal subproblem

$$c_k(F(\cdot) + \nabla h(\cdot) - \nabla h(x^k)) + (\cdot - x^k) = r^k \tag{14}$$

satisfying

$$\|r^k\| \leqslant \sigma \|x^k - y^k\| . \tag{15}$$

S2. Compute x^{k+1} via

$$x^{k+1} = (1-t)P_\Omega[x^k - c_k\alpha_k F(y^k)] + tx^k\,, \tag{16}$$

where

$$\alpha_k = \frac{\langle F(y^k), x^k - y^k\rangle}{c_k\|F(y^k)\|^2}\,. \tag{17}$$

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

Assumptions A1 and A2 guarantee that the generalized proximal subproblem (14) always has an exact solution in $\text{int}(\Omega)$ ($r^k = 0$). This problem will certainly always have inexact solution y satisfying $y \in \text{int}(\Omega)$, and the algorithm is thus well-defined. Moreover, note that the generated sequence $\{x^k\}$ is contained in $\text{int}(\Omega)$, since $P_\Omega(\cdot) \in \Omega$, $x^0 \in \Omega$, and $t \in (0, 1)$.

We now begin to analyze the convergence of the method, starting with a series of lemmas.

LEMMA 3.1. *Suppose that F is continuous and monotone and $x^* \in \Omega$ is a solution of $VIP(F, \Omega)$ (1). Then*

1. *The generated sequence $\{x^k\}$ is bounded*;
2. *The sequence $\{y^k\}$ is bounded*;
3. *$\lim_{k\to\infty}(x^k - y^k) = 0$*;
4. *$\lim_{k\to\infty}(x^k - x^{k+1}) = 0$.*

Proof. From (16), it follows that

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &= \|(1-t)(P_\Omega[x^k - \alpha_k c_k F(y^k)] - x^*) + t(x^k - x^*)\|^2 \\ &= (1-t)^2\|P_\Omega[x^k - \alpha_k c_k F(y^k)] - x^*\|^2 + t^2\|x^k - x^*\|^2 \\ &\quad + 2t(1-t)\langle P_\Omega[x^k - \alpha_k c_k F(y^k)] - x^*, x^k - x^*\rangle\,.\end{aligned}$$

Since $2\langle a, b\rangle \leqslant \|a\|^2 + \|b\|^2$, it follows that

$$\begin{aligned}&2t(1-t)\langle P_\Omega[x^k - \alpha_k c_k F(y^k)] - x^*, x^k - x^*\rangle \\ &\quad \leqslant t(1-t)(\|P_\Omega[x^k - \alpha_k c_k F(y^k)] - x^*\|^2 + \|x^k - x^*\|^2)\,.\end{aligned}$$

From the above two inequalities, we have

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &\leqslant (1-t)^2\|P_\Omega[x^k - \alpha_k c_k F(y^k)] - x^*\|^2 + t^2\|x^k - x^*\|^2 \\ &\quad + t(1-t)(\|P_\Omega[x^k - \alpha_k c_k F(y^k)] - x^*\|^2 + \|x^k - x^*\|^2) \\ &= (1-t)\|P_\Omega[x^k - \alpha_k c_k F(y^k)] - x^*\|^2 + t\|x^k - x^*\|^2\,.\end{aligned}$$

Since the projection operator $P_\Omega[\cdot]$ is nonexpansive and $x^* = P_\Omega[x^*]$, we have

$$\begin{aligned}
&\|P_\Omega[x^k - \alpha_k c_k F(y^k)] - x^*\|^2 \\
&\quad \leqslant \|x^k - x^* - \alpha_k c_k F(y^k)\|^2 \\
&\quad = \|x^k - x^*\|^2 - 2\alpha_k c_k \langle F(y^k), x^k - x^* \rangle + \alpha_k^2 \|c_k F(y^k)\|^2 \\
&\quad \leqslant \|x^k - x^*\|^2 - 2\alpha_k c_k \langle F(y^k), x^k - y^k \rangle + \alpha_k^2 \|c_k F(y^k)\|^2,
\end{aligned}$$

where the last inequality follows from the fact that

$$\langle F(y^k), y^k - x^* \rangle \geqslant 0 .$$

Recall that

$$\alpha_k = \frac{\langle F(y^k), x^k - y^k \rangle}{c_k \|F(y^k)\|^2} ,$$

it follows that

$$\|P_\Omega[x^k - \alpha_k c_k F(y^k)] - x^*\|^2 \leqslant \|x^k - x^*\|^2 - c_k \alpha_k \langle F(y^k), x^k - y^k \rangle .$$

Thus,

$$\|x^{k+1} - x^*\|^2 \leqslant \|x^k - x^*\|^2 - (1-t) c_k \alpha_k \langle F(y^k), x^k - y^k \rangle . \tag{18}$$

From (15) and the strict convexity of h, we have

$$\begin{aligned}
\langle c_k F(y^k), x^k - y^k \rangle &= \langle c_k(\nabla h(x^k) - \nabla h(y^k)), x^k - y^k \rangle + \|x^k - y^k\|^2 + \langle r^k, x^k - y^k \rangle \\
&\geqslant \|x^k - y^k\|^2 - \|r^k\| \, \|x^k - y^k\| \\
&\geqslant (1-\sigma) \|x^k - y^k\|^2 .
\end{aligned} \tag{19}$$

It follows from inequalities (18) and (19) that

$$\|x^{k+1} - x^*\|^2 \leqslant \|x^k - x^*\|^2 - (1-t)(1-\sigma)^2 \frac{\|x^k - y^k\|^4}{c_k^2 \|F(y^k)\|^2} , \tag{20}$$

which means

$$\|x^{k+1} - x^*\|^2 \leqslant \|x^k - x^*\|^2 \leqslant \cdots \leqslant \|x^0 - x^*\|^2 .$$

Thus, the sequence $\{x^k\}$ is bounded and the first assertion is obtained.

From the monotonicity of F and (19), we have that

$$\langle c_k F(x^k), x^k - y^k \rangle \geqslant \langle c_k F(y^k), x^k - y^k \rangle \geqslant (1-\sigma) \|x^k - y^k\|^2 .$$

Using the Cauchy-Schwartz inequality, it follows from the above inequality that

$$\|c_k F(x^k)\| \geqslant (1-\sigma) \|x^k - y^k\| .$$

The second assertion then follows from the boundedness of $\{x^k\}$ and $\{c_k\}$, the continuity of F and the above inequality.

Since $\{y^k\}$ is bounded and F is continuous on Ω, there exists a constant M, such that $\|F(y^k)\| < M$, for all $k > 0$. Then, it follows from (20) that

$$\sum_{k=0}^{\infty} \frac{(1-t)(1-\sigma)^2\|x^k - y^k\|^4}{c_k^2 M^2} < +\infty .$$

The third assertion follows immediately.

From the nonexpansiveness of the projection operator and the fact that $x^k \in \Omega$, we have

$$\begin{aligned}\|x^k - x^{k+1}\| &= (1-t)\|P_\Omega[x^k - c_k\alpha_k F(y^k)] - x^k\| \\ &\leqslant (1-t)c_k\alpha_k\|F(y^k)\| \\ &\leqslant (1-t)\|x^k - y^k\| ,\end{aligned}$$

where the second inequality follows from the definition of α_k and the Cauchy-Schwartz inequality. The conclusion thus follows from the third assertion. This completes the proof. □

LEMMA 3.2. *If F is continuous and monotone on Ω and $\{x^k\}$ has a cluster point $x^\infty \in \operatorname{int}\Omega$, then $F(x^\infty) = 0$ and x^∞ solves the variational inequality problem (1).*

Proof. It follows from (15) and the last assertion in Lemma 3.1 that $r^k \to 0$. Let $\{x^{k_j}\}$ be the corresponding subsequence converging to x^∞. Then from Lemma 3.1 we have $y^{k_j} \to x^\infty$. From (14), we have

$$F(y^{k_j}) = (\nabla h(x^{k_j}) - \nabla h(y^{k_j})) + \frac{1}{c_{k_j}}(x^{k_j} - y^{k_k} + r^{k_j}) .$$

Since $x^\infty \in \operatorname{int}(\Omega)$, ∇h is smooth at x^∞. And since c_k is bounded away from zero, taking limit on the both sides of the above equation and using the continuity of F, we conclude that $F(x^\infty) = 0$ and thus x^∞ solves the variational inequality problem (1). □

To prove the convergence of the algorithm, we have to impose a condition on F, which is, however, weaker than strong or strict monotonicity.

DEFINITION 3.3. A mapping $F: R^n \to R^n$ is paramonotone in a convex set Ω if and only if F is monotone and

$$\langle F(x) - F(y), x - y\rangle = 0 , \quad \forall x, y \in \Omega$$

implies $F(x) = F(y)$.

The following lemmas are taken from [8], the reader can refer to Propositions 23 and 24 in [8] for the proof.

LEMMA 3.4. *If $\{x^k\}$ is bounded and has no cluster points in $\operatorname{int}\Omega$, and*

$$\lim_{k\to\infty}(x^k - x^{k+1}) = 0 ,$$

then for each $u \in \operatorname{int}\Omega$ *there exists a cluster point* $\bar{x}(u)$ *of* $\{x^k\}$ *in* $\partial\Omega$ *such that*

$$\langle F(\bar{x}(u)), u - \bar{x}(u)\rangle \geqslant 0 . \qquad \square$$

LEMMA 3.5. *Suppose that F is continuous and paramonotone. If the solution set* $\Omega^* \neq \emptyset$ *and* $\{x^k\}$ *has no cluster points in* $\operatorname{int}\Omega$, *then there exists a cluster point* $\tilde{x}$ *of* $\{x^k\}$ *in* $\partial\Omega$ *that solves* $VIP(F, \Omega)$. □

We are now in the position to present our main result.

THEOREM 3.6. *If F is continuous and paramonotone,* Ω *is closed and convex with nonempty interior, and h is a Bregman function satisfying Assumptions A*1 *and A*2 *whose zone is the interior of* Ω. *Then the sequence* $\{x^k\}$ *generated by the inexact algorithm converges to a solution of* $VIP(F, \Omega)$ *whenever the solution set* Ω^* *of the problem is nonempty*.

Proof. By Lemma 3.1 $\{x^k\}$ conforming to the algorithm is bounded. So it has at least one cluster point $\bar{x} \in \Omega$. It follows from Lemma 3.2 and Lemma 3.4 that $\bar{x}$ solves VIP(F, Ω). Then substitute x^* by $\bar{x}$ in Lemma 3.1, we have

$$\|x^{k+1} - \bar{x}\|^2 \leqslant \|x^k - \bar{x}\|^2 .$$

The whole sequence $\{x^k\}$ therefore, converges to $\bar{x}$, a solution of VIP(F, Ω). □

4. Solving the proximal subproblem

Note that the proximal subproblem (14) is essentially a system of equations

$$E_k(x) = 0 , \tag{21}$$

with

$$E_k(x) = F(x) + \nabla h(x) - \nabla h(x^k) + \frac{1}{c_k}(x - x^k) .$$

In an alternative way, it can also be viewed as the classical proximal point algorithm to solve the system

$$F(x) + \nabla h(x) - \nabla h(x^k) = 0 .$$

For this problem, Solodov and Svaiter [24] gave a regularized Newton-type algorithm. In the following, we give a similar analysis, which shows that we can get an approximate solution satisfying (15) via a single Newton step.

Suppose that $F\colon R^n \to R^n$ is monotone and continuously differentiable, $h\colon R^n \to R$ is twice continuously differentiable in int(Ω), with $\nabla F + \nabla^2 h$ being Lipschitz continuous. That is, there exist a constant $L > 0$ such that

$$\|(\nabla F(x) + \nabla^2 h(x)) - (\nabla F(y) + \nabla^2 h(y))\| \leqslant L\|x - y\| , \quad \forall x, y \in \operatorname{int}(\Omega) .$$

The fundamental method to solve the subproblem (14) is the Newton method. Next, we will show that if x^k is not a solution of VIP(F, Ω), then a single Newton-type step will make us in the situation of (15) in the kth iterate, provided we choose

$$c_k \in [0.1(L\|F(x^k)\|)^{-1/2}, (L\|F(x^k)\|)^{-1/2}] ,$$

$\sigma \in [1/2, 1)$ and

$$y^k := x^k + s^k ,$$

where

$$s^k := -(c_k(\nabla F(x^k) + \nabla^2 h(x^k)) + I)^{-1} c_k F(x^k) ,$$

and $y^k \in \Omega$. We now prove this fact.

By the Lipschitz continuity of the operator $\nabla F + \nabla^2 h$, we have

$$\begin{aligned} &\|(F(y^k) + \nabla h(y^k)) - (F(x^k) + \nabla h(x^k)) - (\nabla F(x^k) + \nabla^2 h(x^k))(y^k - x^k)\| \\ &\quad \leqslant \frac{L}{2}\|y^k - x^k\|^2 . \end{aligned} \tag{22}$$

Since

$$y^k - x^k = s^k = -c_k F(x^k) - c_k(\nabla F(x^k) + \nabla^2 h(x^k))s^k ,$$

combining the above equation with (22), we get

$$\frac{1}{c_k}\|c_k(F(y^k) + \nabla h(y^k) - \nabla h(x^k)) + y^k - x^k\| \leqslant \frac{L}{2}\|s^k\|\,\|x^k - y^k\| .$$

Since $\nabla F(x)$ and $\nabla^2 h(x)$ are positive semidefinite, we have

$$\|s^k\| \leqslant c_k\|F(x^k)\| .$$

Thus, it follows that

$$\|c_k(F(y^k) + \nabla h(y^k) - \nabla h(x^k)) + (y^k - x^k)\| \leqslant \frac{c_k^2 L\|F(x^k)\|}{2}\|y^k - x^k\| .$$

Then, (15) follows immediately from the choice of σ and c_k.

5. Computational results

In this section, we implement the proposed generalized proximal point method to solve some variational inequalities, and give some preliminary computational results.

In the case that $\Omega = R^n_+$, VIP(F, Ω) becomes the nonlinear complementarity problem (NCP): Find a vector $x^* \in R^n$, such that

$$x^* \geqslant 0 , \qquad F(x^*) \geqslant 0 , \qquad \langle x^*, F(x^*)\rangle = 0 .$$

A Bregman function for this case is

$$h(x) = \sum_{j=1}^{n} x_j \log x_j$$

with the convention that $0 \log 0 = 0$. We define the diagonal matrix $\nabla^2 h(x)$ as

$$\nabla^2 h(x)(i,i) = \begin{cases} 1.0\,, & \text{if } x(i) = 0 \\ 1.0/x(i)\,, & \text{otherwise}\,. \end{cases}$$

As we have discussed in the last section, we use just one Newton-type step to solve the subproblem (14). We take

$$c_k = 1.0/\sqrt{\|F(x^k)\|}\,,$$

$t = 0.01$ and the stop criterion in the following examples is

$$\|x^{k+1} - x^k\| \leqslant 10^{-6}\,.$$

All codes were written in Matlab 5.3 and run on a PIII 600 personal computer.

The first problem is a linear complementarity problem (LCP), in which

$$F(x) = Mx + q\,,$$

where

$$M = \begin{bmatrix} 1 & 2 & \cdots & \cdots & 2 \\ 0 & 1 & 2 & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 2 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}, \quad q = (-1, -1, \ldots, -1)^T\,.$$

It is known that the Lemke's algorithm for this problem run in exponential time [21]. Hanrker and Pang [14] used the damped-Newton algorithm to solve this problem up to $n = 128$ and He and Yang used a projection and contraction algorithm to solve it with n up to 10000. Table 1 quotes their results.

For this problem we take the initial point $x^0 = (1, 1, \ldots, 1)$. The computational results are reported in Table 2.

Table 1. Results by damped-Newton and projection methods

Dimension	8	16	32	64	128
Damped-Newton method	9	20	72	208	>300
Projection method	24	25	27	29	32

Table 2. Computational results of LCP with $n = 8$ to 256

Dimension	8	16	32	64	128	256
Iter. num	7	9	10	11	13	16
CPU (s)	0.06	0.11	0.21	0.28	1.92	16.81

Table 3. Computational results of LCP with $n = 64$ for different initial point

Trial	1	2	3	4	5	6	7	8	9	10
Iter. num	27	25	28	27	26	29	24	27	28	27
CPU (s)	0.31	0.22	0.33	0.32	0.27	0.33	0.27	0.32	0.33	0.27

Tables 1 and 2 show that, compared with the damped-Newton algorithm and the projection and contraction algorithm, the proposed method is substantial computational efficiency. The iterative number of the proposed algorithm is much less than those algorithms. The subproblem in our proposed algorithm is nonlinear equation, which is generally difficult to solve than that in the damped-Newton algorithm (linear equation) and the projection and contraction algorithm (projection to the feasible set). However, since we just need to get an approximately solution and a Newton step is enough, the subproblem is not complex in this sense.

To show that our method converges globally, we solve the above problem with $n = 64$ and the initial point generated uniformly in $(0, 10)$. Table 3 reports the computational results.

From Tables 2 and 3, we can see that the iterative number and the CPU time are quite insensitive to the starting point. The above two tables show that our method is efficient and converges globally.

In our second test problem we take

$$F(x) = D(x) + Mx + q\,,$$

where $D(x)$ and $Mx + q$ are the nonlinear part and the linear part of $F(x)$, respectively. We form the linear part $Mx + q$ similarly as in [14].[1] The matrix $M = A^\top A + B$, where A is an $n \times n$ matrix whose entries are randomly generated in the interval $(-1, +1)$ and a skew-symmetric matrix B is generated in the same way. The vector q is generated from a uniform distribution in the interval $(-500, +500)$. In $D(x)$, the nonlinear part of $F(x)$, the components are $D_j(x) = a_j * \arctan(x_j)$ and a_j is a random variable in $(0, 100)$. A similar type of the problem was tested in [19, 25].[2] Table 4 reports our computational results for n from 10 to 100 with the

Table 4. Computational results of NCP

n	10	20	30	40	50	60	70	80	90	100
Iter. num.	11	12	15	16	18	15	25	36	40	37
CPU (s)	0.05	0.05	0.11	0.17	0.26	0.47	0.82	1.02	0.99	1.27

[1] In the paper by Harker and Pang [14], the matrix $M = A^\top A + B + D$, where A and B are the same matrices as here, and D is a diagonal matrix with uniformly distributed random variable $d_{jj} \in (0.0, 0.3)$.

[2] In [19, 25], the components of nonlinear mapping $D(u)$ are $D_j(x) = \rho * \arctan(x_j)$ and $\rho > 0$ is a constant.

initial point generated randomly in (0, 1). It seems that the iterations number varies just slightly.

To give some comparison, we solve the same problem as in [19, 25] where the constraint set S is

$$S = \left\{ x \in R^5 \,\middle|\, \sum_{i=1}^{5} x_i \geqslant 10,\ x_i \geqslant 0,\ i = 1, 2, \ldots, 5 \right\}$$

but M is a 5×5 asymmetric positive definite matrix whose entries are randomly generated in $(-5, 5)$, the vector q is generated from a uniform distribution in the interval $(-10, 10)$, and $D_i(x) = \arctan(x_i - 2),\ i = 1, 2, \ldots, 5$.

To solve this problem, we first introduce a Lagrange multiplier to convert the problem to an equivalent nonlinear complementarity problem. We also code the globally convergent Newton method (GCNM) of Taji et al. [25]. We use the quadratic-program solver quadprog.m from the MATLAB optimization toolbox to perform the projection. We rewrite the subproblem in [25] as a linear complementarity problem (LCP) and solve it by Lemke's complementarity pivoting method [10], which finds a solution of LCP in a finite number of steps. The parameters used in their algorithm are set the same as those in [25]. Tables 5 and 6 give the numerical results for $\rho = 100$ and $\rho = 200$, respectively. For simplicity, in these tables, we denote our method by GPPA.

From Tables 5 and 6, we can see though the iterative number is larger than Newton-type method [25], the total CPU time is smaller. Especially, the computational cost at each iteration is much smaller. There are two reasons. The first one is that, at each iteration, the Newton-type method [25] needs to make some projections to the feasible set S, which is more difficult than to make projections to the simple set Ω (In the test problem, $\Omega = R^5 \times R^5_+$, where R^5_+ denotes the nonnegative orthant of R^5). The other one is that, the Newton-type method [25] needs to solve a linear variational inequality problem at each iteration, while the proposed algorithm needs only to solve a system of nonlinear equations approximately.

Table 5. Numerical results for $\rho = 100$

Starting point	Algorithm	Num. of iter.	CPU (s)
(25, 0, 0, 0, 0)	GCNM	6	0.29
	GPPA	13	0.06
(10, 0, 10, 0, 10)	GCNM	6	0.34
	GPPA	11	0.05
(10, 0, 0, 0, 0)	GCNM	7	0.33
	GPPA	11	0.05
(0, 2.5, 2.5, 2.5, 2.5)	GCNM	5	0.23
	GPPA	10	0.05

Table 6. Numerical results for $\rho = 200$

Starting point	Algorithm	Num. of iter.	CPU (s)
(25, 0, 0, 0, 0)	GCNM	7	0.38
	GPPA	15	0.06
(10, 0, 10, 0, 10)	GCNM	6	0.29
	GPPA	13	0.05
(10, 0, 0, 0, 0)	GCNM	7	0.44
	GPPA	14	0.05
(0, 2.5, 2.5, 2.5, 2.5)	GCNM	5	0.28
	GPPA	11	0.05

6. Conclusions

In this paper, we give a new Bregman function-based proximal point algorithm for the variational inequality problem with monotone operators. We allow to solve the subproblems approximately, with the constructive accuracy criterion (15). Under some suitable conditions, we prove the global convergence of the algorithm. A Newton-type method is discussed to solve the generalized proximal subproblem and a single step is enough to ensure the approximate solution satisfies the criterion. Furthermore, we also report some preliminary computational results to show the efficiency of the method.

The Bregman function we used in this paper is of the form

$$h_k(x) = h_0(x) + \frac{c_k}{2}\|x\|^2 ,$$

where $h_0(\cdot)$ is a Bregman function. The idea that using the Bregman function with this structure is natural: the Bregman function h_0 is used to restrict the generated sequence $\{y^k\}$ exists in Ω and $c_k/2\|x\|^2$ is used to control the error tolerance [1, 2]. Moreover, we allow to use different Bregman function per iteration, which enable us to choose c_k self-adaptively such that the trial point generated by a single Newton step satisfies (15).

Acknowledgment

The author would like to thank the two anonymous referees for their careful reading of an earlier version of this paper and constructive suggestions which led to improvements of the paper.

References

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

[2] Auslander, A., Teboulle, M. and Ben-Tiba, S. (1999), Interior proximal and multiplier methods based on second order homogeneous kernels, *Mathematics of Operations Research* 24, 645–668.

[3] Bauschke, H.H. and Borwein, J.M. (1997), Legendre functions and the method of random Bregman projections, *Journal of Convex Analysis* 4, 27–67.

[4] Bregman, L.M. (1967), The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming, *USSR Computational Mathematics and Mathematical Physics* 7, 200–217.

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

[6] Burachik, R.S., Iusem, A.N. and Svaiter, B.F. (1997), Enlargement of monotone operators with applications to variational inequalities, *Set-Valued Analysis* 5, 159–180.

[7] Censor, Y. and Zenios, S.A. (1992), The proximal minimization algorithm with *D*-functions, *Journal of Optimization Theory and Applications* 73, 451–464.

[8] Censor, Y., Iusem, A.N. and Zenios, S.A. (1998), An interior point method with Bregman functions for variational inequality problem with paramonotone operators, *Mathematical Programming* 81, 373–400.

[9] Chen, G. and Teboulle, M. (1993), Convergence analysis of proximal-like optimization algorithm using Bregman functions, *SIAM Journal on Optimization* 3, 538–543.

[10] Eaves, B.C. (1978), Computing stationary points, *Mathematical Programming Study* 7, 1–14.

[11] Eckstein, J. (1993), Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming, *Mathematics of Operations Research* 18, 202–226.

[12] Eckstein, J. (1998), Approximate iterations in Bregman function-based proximal algorithms, *Mathematical Programming* 83, 113–123.

[13] Eckstein, J. and Ferris, M.C. (1999), Smooth methods of multipliers for complementarity problems, *Mathematical Programming* 86, 65–90.

[14] Harker, P.T. and Pang, J.S. (1990), A damped-Newton method for the linear complementarity problem, *Lectures in Applied Mathematics* 26, 265–284.

[15] He, B.S. and Yang, H. (2000), A neural-network model for monotone linear asymmetric variational inequalities, *IEEE Transaction on Neural Networks* 11, 3–16.

[16] Iusem, A.N. (1998), On some properties of generalized proximal point methods for the variational inequality problem, *Journal of Optimization Theory and Applications* 96, 337–362.

[17] Iusem, A.N. (1995), On some properties of generalized proximal point methods for quadratic and linear programming, *Journal of Optimization Theory and Applications* 85, 596–612.

[18] Kiwiel, K.C. (1997), Proximal minimization methods with generalized Bregman functions, *SIAM Journal on Control and Optimization* 35, 1142–1168.

[19] Marcotte, P. and Dussault, J.P. (1987), A note on a globally convergent Newton method for solving variational inequalities, *Operations Research Letters* 6, 35–42.

[20] Martinet, B. (1970), Regularization d'inequations variationelles par approximations successives, *Revue Francaise d'Informatique et de Recherche Opérationelle* 4, 154–159.

[21] Murty, K.G. (1988), *Linear Complementarity*, *Linear and Nonlinear Programming*, Helderman-Verlag, Berlin.

[22] Rockafellar, R.T. (1976), Monotone operators and the proximal point algorithm, *SIAM Journal on Control and Optimization* 14, 877–898.

[23] Solodov, M.V. and Svaiter, B.F. (2000), An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions, *Mathematics of Operations Research* 25, 214–230.

[24] Solodov, M.V. and Svaiter, B.F. (1999), A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator, *Set Valued Analysis* 7, 323–345.

[25] Taji, K., Fukushima, M. and Ibaraki, T. (1993), A globally convergent Newton method for solving strongly monotone variational inequalities, *Mathematical Programming* 58, 369–383.

[26] Teboulle, M. (1997), Convergence of proximal-like algorithms, *SIAM Journal on Optimization* 7, 1069–1083.