Projections Onto Super-Half-Spaces for Monotone Variational Inequality Problems in Finite-Dimensional Space

Yair Censor and Aviv Gibali

Department of Mathematics, University of Haifa,

Mt. Carmel, Haifa 31905, Israel.

(yair@math.haifa.ac.il, aviv_gibali@yahoo.com)

July 2, 2008. Revised: August 19, 2008.

Abstract

The variational inequality problem (VIP) is considered here. We present a general algorithmic scheme which employs projections onto hyperplanes that separate balls from the feasible set of the VIP instead of projections onto the feasible set itself. Our algorithmic scheme includes the classical projection method and Fukushima's subgradient projection method as special cases.

1   Introduction

We consider the variational inequality problem (VIP) in the Euclidean space R^n. Given a nonempty closed convex set $X \subseteq R^n$ and a function $f : R^n \to R^n$, the VIP is to find a point x^* such that

$$x^* \in X \text{ and } \langle f(x^*), x - x^* \rangle \geq 0, \text{ for all } x \in X. \tag{1}$$

This problem was well-studied in the last decades, see, e.g., the treatise of Facchinei and Pang [7] and the review papers by Noor [12] and by Xiu and Zhang [13]. In particular, algorithmic approaches were investigated, using projections of different types, in order to generate a sequence of iterates

1

that converges to a solution. See, e.g., Yang [15], Yamada and Ogura [14], Auslender and Teboulle [3] or Censor, Iusem and Zenios [5], to name but (very) few out of the existing vast literature. The importance of VIPs stems from the fact that some fundamental problems can be cast in this form, see, e.g., [7, Volume I, Subsection 1.4].

Some algorithms for solving the VIP fit into the framework of the following general iterative scheme.

Algorithm 1
Initialization: *Let $\{\tau_k\}_{k=0}^{\infty}$ be a user-chosen positive real sequence, select an arbitrary starting point $x^0 \in R^n$ and set the iteration index $k = 0$.*
Iterative step: *Given the current iterate x^k, calculate the next iterate*

$$x^{k+1} = P_X(x^k - \tau_k G^{-1} f(x^k)),$$

where G is a symmetric positive definite matrix, P_X is the projection operator onto X with respect to the G-norm $\|z\|_G = \langle z, Gz \rangle^{1/2}$.

See, Auslender [2] and consult [7, Volume 2, Subsection 12.1] for more details. Such methods are particularly useful when the set X is simple enough to project on. However, in general, one has to solve at each iterative step the minimization problem

$$\min\{\|x - (x^k - \tau_k G^{-1} f(x^k))\|_G \mid \text{for all } x \in X\}.$$

The efficiency of such a projection method may be seriously affected by the need to solve such optimization problems at each iterative step.

An orthogonal projection of a point z onto a set X can be viewed as an orthogonal projection of z onto the hyperplane H which separates z from X, and supports X at the closest point to z in X. But, of course, at the time of performing such an orthogonal projection, neither the closest point to z in X, nor the separating and supporting hyperplane H are available. In view of the simplicity of an orthogonal projection onto a hyperplane, it is natural to ask whether one could use other separating supporting hyperplanes instead of that particular hyperplane H through the closest point to z. Aside from theoretical interest, this may lead to algorithms useful in practice, provided that the computational effort of finding such other hyperplanes favorably competes with the work involved in performing orthogonal projections directly onto the given sets.

To circumvent the difficulties associated with the orthogonal projections onto the feasible set of (1) Fukushima [10] developed a method that utilizes outer approximations of X. His method replaces the orthogonal projection onto the set X by a projection onto a half-space containing X, which is easier to calculate. Letting $X := \{x \in R^n \mid g(x) \le 0\}$ where $g : R^n \to R$ is convex, Fukushima's algorithm is as follows.

Algorithm 2
Initialization: *Let $\{\rho_k\}_{k=0}^{\infty}$ be a user-chosen positive real sequence, select an arbitrary starting point $x^0 \in R^n$ and set the iteration index $k = 0$.*
Iterative step: *Given the current iterate x^k,*
_(1)_ *choose a subgradient $\xi^k \in \partial g(x^k)$ of g at x^k and let*

$$T^k := \{x \in R^n \mid g(x^k) + \langle \xi^k, x - x^k \rangle \le 0\}.$$

_(2)_ *Calculate the "shifted point"*

$$z^k := x^k - \rho_k G^{-1} f(x^k) / \|f(x^k)\|_2,$$

and then the next iterate x^{k+1} is the projection of z^k onto the half-space T^k with respect to the G-norm, namely,

$$x^{k+1} = P_{T^k}(z^k).$$

_(3)_ *If $x^{k+1} = x^k$ then stop, otherwise, set $k = k + 1$ and return to **(1)**.*

Since the bounding hyperplanes of the subgradiental half-spaces T^k, used by Fukushima, separate the current point z from the set X, the question again arises whether or not any other separating hyperplanes can be used in the algorithm while retaining the overall convergence to the solution.

This question presents a theoretical challenge and we are able to offer here answers that hold under some not too restrictive conditions. Under these conditions, we are able to show that, as a matter of fact, the hyperplanes need to separate not just the point z from the feasible set of (1), but rather separate a "small" ball around z from X. This is inspired by our earlier work [1] on the convex feasibility problem. Whether or not our current restrictions can be relaxed or removed still remains to be seen.

3

Our goal is to present the structural algorithmic discovery that both Algorithm 1 and Algorithm 2 are realizations of a more general algorithmic principle. Once we achieve this we are less concerned at the moment about the minimal strength of the conditions under which our results hold. This and generalizations of the new algorithmic structure to algorithms which use two projections per iteration for the VIP are currently under investigation. Our work is admittedly a theoretical development and no numerical advantages are claimed at this point. The large "degree of freedom" of choosing the super-sets, onto which the projections of the algorithm are performed, from a wide family of half-spaces may include specific algorithms that have not yet been explored. In Section 2 we present the algorithmic scheme and in Section 3 we give our convergence analysis. Section 4 discusses special cases.

2 The Algorithmic Scheme

2.1 Assumptions

Let $X := \{x \in R^n \mid g(x) \leq 0\}$ where $g : R^n \to R$ is convex. Let G be a symmetric positive definite matrix and denote the distance from a point $x \in R^n$ to the set X with respect to the norm $\|\cdot\|_G$ by

$$\text{dist}(x, X) := \min\{\|x - z\|_G \mid z \in X\}.$$

For any $\varepsilon > 0$, we denote

$$X_\varepsilon := \{x \in R^n \mid g(x) < \varepsilon\}.$$

Following [10], we assume that the following conditions are satisfied.

Condition 3 *f is continuous on X_ε for some $\varepsilon > 0$.*

Condition 4 *f is strongly monotone with constant α on X_ε for some $\varepsilon > 0$, i.e.,*

$$\langle f(x) - f(y), x - y \rangle \geq \alpha \|x-y\|_2^2, \quad \text{for all} \ \ x, y \in X_\varepsilon \ \ \text{and some} \ \alpha > 0. \quad (2)$$

Condition 5 *For some $y \in X$, there exist a $\beta > 0$ and a bounded set $D \subset R^n$ such that*

$$\langle f(x), x - y \rangle \geq \beta \|f(x)\|_2, \ \text{for all} \ x \notin D.$$

It is well-known that under Conditions 3 and 4, the problem (1) has a unique solution, see, e.g., Kinderlehrer and Stampacchia [11, Corollary 4.3, p. 14].

4

2.2 The algorithmic scheme

In order to present the algorithmic scheme a few definitions are needed.

Definition 6 *Let G be a symmetric positive definite matrix. Given a $0 \leq \delta \leq 1$, a closed convex set $X \subseteq R^n$ and a point $x \in R^n$,*

(i) $B(x, X, \delta) := B(x, \delta \operatorname{dist}(x, X)) = \{z \in R^n \mid \|x - z\|_G \leq \delta \operatorname{dist}(x, X)\}$ is the G-ball centered at x with radius $\delta \operatorname{dist}(x, X)$,

(ii) for any $x \notin \operatorname{int} X$, denote by $\mathcal{H}(x, X, \delta)$ the set of all hyperplanes which separate $B(x, X, \delta)$ from X,

(iii) for $x, y \in R^n$, define the mapping

$$\mathcal{A}_{X,\delta}(x, y) := \begin{cases} \{x\}, & \text{if } x \in \operatorname{int} X, \\ \{P_{H^-}(y) \mid H \in \mathcal{H}(x, X, \delta)\}, & \text{if } x \notin \operatorname{int} X, \end{cases}$$

where P_{H^-} is the projection operator, with respect to the G-norm, onto the half-space whose bounding hyperplane is H and such that $X \subseteq H^-$.

The mapping $\mathcal{A}$ defined above maps a quadruple (x, y, X, δ) onto a set. A *selection* from $\mathcal{A}_{X,\delta}(x, y)$ means that if $x \notin \operatorname{int} X$ a specific hyperplane $H \in \mathcal{H}(x, X, \delta)$ is chosen and $P_{H^-}(y)$ is selected. If $x \in \operatorname{int} X$ then x is selected.

Let $\{\rho_k\}_{k=0}^{\infty}$ be a sequence of positive numbers satisfying

$$\lim_{k \to \infty} \rho_k = 0 \quad \text{and} \quad \sum_{k=0}^{\infty} \rho_k = +\infty. \tag{3}$$

Our algorithmic scheme for the VIP is as follows.

Algorithm 7
Initialization*: Let $\{\rho_k\}_{k=0}^{\infty}$ be a user-chosen positive real sequence that fulfills (3). Choose a constant δ such that $0 < \delta \leq 1$, select an arbitrary starting point $x^0 \in R^n$ and set $k = 0$.*
Iterative step*: Given the current iterate x^k,*
(1) *calculate the "shifted point"*

$$z^k := \begin{cases} x^k - \rho_k G^{-1} f(x^k) / \|f(x^k)\|_2, & \text{if } f(x^k) \neq 0, \\ x^k, & \text{if } f(x^k) = 0. \end{cases} \tag{4}$$

(2) *Calculate the next iterate by*

$$x^{k+1} \quad \text{is a selection from} \quad \mathcal{A}_{X,\delta}(x^k, z^k). \tag{5}$$

(3) *If $x^{k+1} = x^k$, stop, otherwise, set $k = k+1$ and return to **(1)**.*

In what follows, we shall denote by P_k the projection operator onto H_k^- where H_k is the selected hyperplane $H_k \in \mathcal{H}(x^k, X, \delta)$ Thus, in (5) we have

$$x^{k+1} = \begin{cases} x^k, & \text{if } x^k \in \text{int } X, \\ P_k(z^k), & \text{if } x^k \notin \text{int } X. \end{cases} \tag{6}$$

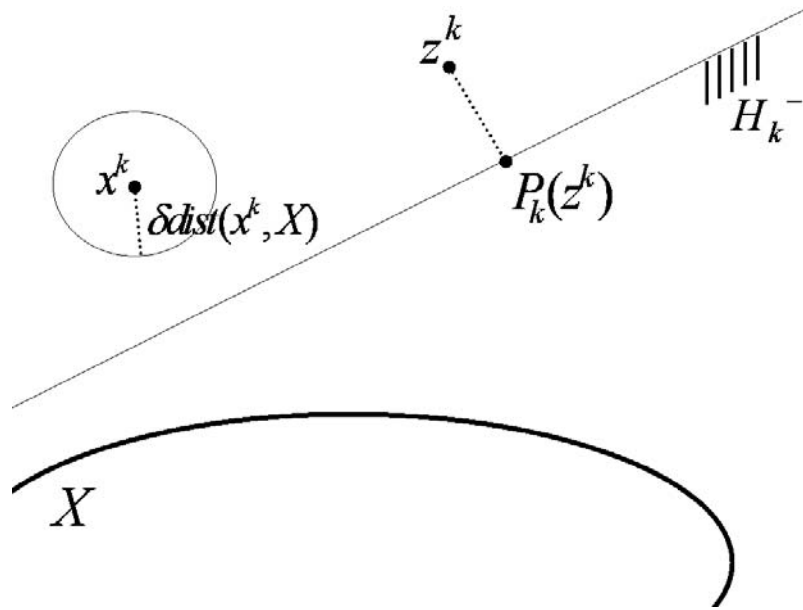The iterative step of this algorithmic scheme is illustrated in Figure 1.

Figure 1: Illustration of the algorithmic scheme in Algorithm 7.

Remark 8 *Observe that there is no need to calculate in practice the radius $\delta \operatorname{dist}(x^k, X)$ of the ball $B(x^k, X, \delta)$. If there would have been a need to calculate this then it would, obviously, amount to preforming a projection of x^k*

6

onto X, which is the very thing that we are trying to circumvent. All that is needed, when deriving from the algorithmic scheme a specific algorithm, is to show that the specific algorithm indeed "chooses" the hyperplanes in concert with the requirement of separating such $B(x^k, X, \delta)$ balls from the feasible set of (1). We demonstrate this in Section 4 below.

3 Convergence

First we show that if Algorithm 7 stops then it has reached the solution of the VIP.

Theorem 9 *If $x^{k+1} = x^k$ occurs for some $k \geq 0$ in Algorithm 7, then x^k is the solution of problem (1).*

Proof. First assume that $f(x^k) \neq 0$, then $x^{k+1} = x^k$ is possible only if the radius of $B(x^k, X, \delta)$ is zero which implies that $x^k \in X$ since $\delta > 0$. By definition, $P_k(\overline{y})$ is the solution of the problem $\min \left\{ \|\overline{y} - w\|_G^2 \mid w \in H_k^- \right\}$, where H_k^- is the half-space determined by H_k and $X \subseteq H_k^-$. Applying [4, Theorem 2.4.2] with the Bregman function $h(\overline{y}) = \frac{1}{2}\|\overline{y}\|_G^2$, whose zone is R^n, for the set H_k^-, we get

$$\langle w - P_k(\overline{y}), \nabla h(\overline{y}) - \nabla h(P_k(\overline{y})) \rangle \leq 0, \quad \text{for all } w \in H_k^-,$$

where here $\nabla h(\overline{y}) = G\overline{y}$, $\nabla h(P_k(\overline{y})) = GP_k(\overline{y})$. So, we have

$$\langle w - P_k(\overline{y}), G\overline{y} - GP_k(\overline{y}) \rangle \leq 0, \quad \text{for all } w \in H_k^-. \tag{7}$$

Taking $w = \overline{x}$, $\overline{y} = z^k$ in (7) we get

$$\langle \overline{x} - P_k(z^k), Gz^k - GP_k(z^k) \rangle \leq 0, \quad \text{for all } \overline{x} \in H_k^-, \tag{8}$$

which, by (4), implies that for all $\overline{x} \in H_k^-$

$$\langle \overline{x} - x^{k+1}, G(x^k - \rho_k G^{-1} f(x^k)/\|f(x^k)\|_2) - Gx^{k+1} \rangle$$
$$= \langle \overline{x} - x^{k+1}, G(x^k - x^{k+1}) - \rho_k f(x^k)/\|f(x^k)\|_2 \rangle$$
$$= \langle \overline{x} - x^{k+1}, G(x^k - x^{k+1}) \rangle - \langle \overline{x} - x^{k+1}, \rho_k f(x^k)/\|f(x^k)\|_2 \rangle \leq 0.$$

Since we assume that $x^{k+1} = x^k$

$$\langle \overline{x} - x^{k+1}, G(x^k - x^{k+1}) \rangle = 0,$$

so we get

$$\langle \overline{x} - x^{k+1}, \rho_k f(x^k)/\|f(x^k)\|_2 \rangle \geq 0, \quad \text{for all } \overline{x} \in H_k^-,$$

and then $\rho_k/\|f(x^k)\|_2 > 0$, $X \subseteq H_k^-$, and $x^{k+1} = x^k$ lead to

$$\langle \overline{x} - x^k, f(x^k) \rangle \geq 0, \ \text{for all } \overline{x} \in X.$$

On the other hand, if $f(x^k) = 0$, then (8) holds with $z^k = x^k$, i.e.,

$$\langle \overline{x} - P_k(x^k), Gx^k - GP_k(x^k) \rangle \leq 0, \quad \text{for all } \overline{x} \in H_k^-,$$

or, by (6) and $x^{k+1} = x^k$,

$$\langle \overline{x} - x^k, 0 \rangle \leq 0, \quad \text{for all } \overline{x} \in H_k^-,$$

which is true for all $\overline{x} \in X$ since $X \subseteq H_k^-$ and the proof is complete. ∎

In the remainder of this section we suppose that Algorithm 7 generates an infinite sequence $\{x^k\}_{k=0}^{\infty}$ and establish the next lemmas that will be useful in proving the convergence of our algorithmic scheme. The next lemma was proved in [8] for $G = I$ and under the assumption that E is compact.

Lemma 10 *Let A, E and F be nonempty closed convex sets in R^n, such that $A \subset E \subset F$. For any point $x \in F$, let y be the point in E closest to x. Then we have*

$$\|y - z\|_G^2 \leq \|x - z\|_G^2 - \|y - x\|_G^2, \quad \text{for all } z \in A. \tag{9}$$

Furthermore,

$$(\text{dist}(y, A))^2 \leq (\text{dist}(x, A))^2 - \|y - x\|_G^2. \tag{10}$$

Proof. Since y is the optimal solution of the problem $\min \{\|e - x\|_G^2 \mid e \in E\}$, it must satisfy the inequality (see, e.g., [4, Theorem 2.4.2])

$$\langle e - y, Gx - Gy \rangle \leq 0, \ \text{for all } e \in E.$$

Thus, for all $e \in E$,

$$\|x - e\|_G^2 - \|y - x\|_G^2 - \|y - e\|_G^2 = 2 \langle y - e, Gx - Gy \rangle$$
$$= -2 \langle e - y, Gx - Gy \rangle \geq 0,$$

and, since $A \subset E$, we get

$$\|x - z\|_G^2 - \|y - x\|_G^2 - \|y - z\|_G^2 \geq 0, \quad \text{for all } z \in A,$$

which proves (9). In order to prove (10), let $\tilde{z} \in A$ be the closest point to $x \in F$. Then, using (9), we obtain

$$(\mathrm{dist}(y, A))^2 \leq \|y - \tilde{z}\|_G^2 \leq \|x - \tilde{z}\|_G^2 - \|y - x\|_G^2$$
$$= (\mathrm{dist}(x, A))^2 - \|y - x\|_G^2,$$

which completes the proof. ∎

Lemma 11 *Let $\overline{y} \in R^n$ be an arbitrary point. Then, in Algorithm 7, for all $k \geq 0$, we have the inequalities*

$$\|P_k(\overline{y}) - w\|_G^2 \leq \|\overline{y} - w\|_G^2 - \|P_k(\overline{y}) - \overline{y}\|_G^2, \quad \text{for all } w \in X.$$

Proof. This follows from Lemma 10 with $A = X$, $E = H_k^-$ and $F = R^n$. ∎

The next lemma is quoted from [10, Lemma 2].

Lemma 12 *Let $\{a_k\}_{k=0}^{\infty}$ and $\{b_k\}_{k=0}^{\infty}$ be sequences of nonnegative numbers, and let $\mu \in [0, 1)$ be a constant. If the inequalities*

$$a_{k+1} \leq \mu a_k + b_k, \quad k = 0, 1, \ldots,$$

hold and if $\lim_{k \to \infty} b_k = 0$, then $\lim_{k \to \infty} a_k = 0$.

Lemma 13 *If Condition 5 is satisfied, then any sequence $\{x^k\}_{k=0}^{\infty}$, generated by Algorithm 7, is bounded.*

Proof. The proof is structured along the lines of [10, Lemma 3]. First assume that $f(x^k) \neq 0$. Let $y \in X$ be a point for which Condition 5 holds and let $M > 0$ be such that $\|x - y\|_G < M$, for all $x \in D$, where D is a bounded set given in Condition 5. Lemma 11 implies that, for each $z^k \in R^n$,

$$\|P_k(z^k) - y\|_G^2 \leq \|z^k - y\|_G^2. \tag{11}$$

Therefore,

$$\|x^{k+1} - y\|_G^2 \leq \left\|x^k - \rho_k G^{-1} \frac{f(x^k)}{\|f(x^k)\|_2} - y\right\|_G^2 = \|x^k - y\|_G^2$$
$$- 2\frac{\rho_k}{\|f(x^k)\|_2}\langle f(x^k), x^k - y\rangle + \frac{\rho_k^2}{\|f(x^k)\|_2^2}\langle f(x^k), G^{-1}f(x^k)\rangle. \tag{12}$$

9

Thus, if $\|x^k - y\|_G \geq M$, then we have, by (12) and Condition 5,

$$\|x^{k+1} - y\|_G^2 \leq \|x^k - y\|_G^2 - 2\rho_k\beta + \frac{\rho_k^2}{\|f(x^k)\|_2^2}\langle f(x^k), G^{-1}f(x^k)\rangle. \qquad (13)$$

From Demmel [6, Equation (5.2), page 199] we have

$$\langle f(x^k), G^{-1}f(x^k)\rangle \leq \upsilon\langle f(x^k), f(x^k)\rangle,$$

where υ is the largest eigenvalue of G^{-1}, so that

$$\langle f(x^k), G^{-1}f(x^k)\rangle \leq \nu^{-1}\langle f(x^k), f(x^k)\rangle, \qquad (14)$$

where ν is the smallest eigenvalue of G. By (13) and (14)

$$\|x^{k+1} - y\|_G^2 \leq \|x^k - y\|_G^2 - \rho_k(2\beta - \rho_k\nu^{-1}).$$

Since $\lim_{k\to\infty}\rho_k = 0$, the last inequality implies

$$\|x^{k+1} - y\|_G < \|x^k - y\|_G, \qquad (15)$$

provided that k is sufficiently large. On the other hand, since $y \in X$,

$$\left\|x^{k+1} - y\right\|_G \leq \left\|z^k - y\right\|_G = \left\|x^k - \rho_k G^{-1}\frac{f(x^k)}{\|f(x^k)\|_2} - y\right\|_G$$
$$= \left\|(x^k - y) - \rho_k G^{-1}\frac{f(x^k)}{\|f(x^k)\|_2}\right\|_G.$$

Using the triangle inequality with G-norms and (14) we obtain

$$\left\|(x^k - y) - \rho_k G^{-1}\frac{f(x^k)}{\|f(x^k)\|_2}\right\|_G \leq \left\|x^k - y\right\|_G + \rho_k\nu^{-1/2},$$

so that
$$\left\|x^{k+1} - y\right\|_G \leq \left\|x^k - y\right\|_G + \varepsilon, \qquad (16)$$

for all sufficiently large k, where $\varepsilon > 0$ is a small constant. Inequalities (15) and (16) imply that $\{x^k\}_{k=0}^\infty$ is bounded. If $f(x^k) = 0$ we have by (6) and (11)

$$\|x^{k+1} - y\|_G^2 \leq \|x^k - y\|_G^2,$$

which implies (16) and the rest follows. ∎

Lemma 14 *Let X be a nonempty closed convex set $X \subseteq R^n$ and let $0 < \delta \leq 1$. Let $W \subset R^n$ be a nonempty convex compact set, let $W \setminus X := \{x \in W \mid x \notin X\}$. Denote by $\sigma(x)$ a selection from $\mathcal{A}_{X,\delta}(x,x)$, then there exists a constant $\mu \in [0,1)$ such that*

$$\operatorname{dist}(\sigma(x), X) \leq \mu \operatorname{dist}(x, X), \quad \text{for all } x \in W \setminus X.$$

Proof. Since $\sigma(x)$ is, by definition, the closest point to x in the convex set H^-, and $X \subseteq H^-$, we get

$$(\operatorname{dist}(\sigma(x), X))^2 \leq (\operatorname{dist}(x, X))^2 - \|\sigma(x) - x\|_G^2. \tag{17}$$

This holds because

$$(\operatorname{dist}(\sigma(x), X))^2 \leq \|\sigma(x) - P_X(x)\|_G^2, \tag{18}$$

and, by Lemma 10, we get

$$\|\sigma(x) - P_X(x)\|_G^2 \leq \|x - P_X(x)\|_G^2 - \|\sigma(x) - x\|_G^2$$
$$= (\operatorname{dist}(x, X))^2 - \|\sigma(x) - x\|_G^2. \tag{19}$$

(18) and (19) imply (17). On the other hand, since $\sigma(x)$ is a selection from $\mathcal{A}_{X,\delta}(x,x)$ and $x \notin X$, we have that $\sigma(x) = P_{H^-}(x)$ where $H \in \mathcal{H}(x, X, \delta)$, so

$$\operatorname{dist}(x, X) \cdot \delta \leq \|\sigma(x) - x\|_G, \text{ for all } x \in R^n, \tag{20}$$

(if $x \in X$ then both sides of (20) are zero.) Since δ is a positive constant such that $0 < \delta \leq 1$, we have

$$-\delta^2 (\operatorname{dist}(x, X))^2 \geq -\|\sigma(x) - x\|_G^2, \quad \text{for all } x \in W \setminus X,$$

so, by (17), for all $x \in W \setminus X$

$$(\operatorname{dist}(\sigma(x), X))^2 \leq (\operatorname{dist}(x, X))^2 - \delta^2 (\operatorname{dist}(x, X))^2$$
$$= (1 - \delta^2)(\operatorname{dist}(x, X))^2.$$

Since also $0 \leq 1 - \delta^2 < 1$, taking $\mu = \sqrt{1 - \delta^2} \in [0, 1)$, we get

$$\operatorname{dist}(\sigma(x), X) \leq \mu \operatorname{dist}(x, X), \quad \text{for all } x \in W \setminus X,$$

which completes the proof. ∎

Lemma 15 *For any sequence $\{x^k\}_{k=0}^\infty$, generated by Algorithm 7, we have* $\lim_{k\to\infty} \mathrm{dist}(x^k, X) = 0$.

Proof. Assume that $f(x^k) \neq 0$, by (10) of Lemma 10 we have

$$(\mathrm{dist}(P_k(x^k), X))^2 \leq (\mathrm{dist}(x^k, X))^2 - \|P_k(x^k) - x^k\|_G^2, \text{ for all } k \geq 0,$$

and, by Lemma 14, there exists a constant $\mu \in [0, 1)$ such that

$$\mathrm{dist}(P_k(x^k), X) \leq \mu \, \mathrm{dist}(x^k, X), \quad \text{for all } k \geq 0. \tag{21}$$

On the other hand, the nonexpansiveness of the projection operator and reuse of (14) implies

$$\|x^{k+1} - P_k(x^k)\|_G^2 = \|P_k(z^k) - P_k(x^k)\|_G^2 \leq \|z^k - x^k\|_G^2$$

$$= \left\| x^k - \rho_k G^{-1} \frac{f(x^k)}{\|f(x^k)\|_2} - x^k \right\|_G^2$$

$$= \left\| \rho_k G^{-1} \frac{f(x^k)}{\|f(x^k)\|_2} \right\|_G^2 \leq \rho_k^2 \nu^{-1}.$$

Therefore,

$$\|x^{k+1} - P_k(x^k)\|_G \leq \frac{\rho_k}{\sqrt{\nu}}. \tag{22}$$

Let $s^k = P_X\left(P_k\left(x^k\right)\right)$, namely,

$$\|P_k(x^k) - s^k\|_G = \mathrm{dist}(P_k(x^k), X).$$

Then, by the triangle inequality, we get

$$\|x^{k+1} - s^k\|_G = \|x^{k+1} - P_k(x^k) + P_k(x^k) - s^k\|_G$$

$$\leq \|x^{k+1} - P_k(x^k)\|_G + \|P_k(x^k) - s^k\|_G.$$

Now, since $s^k \in X$, we have

$$\mathrm{dist}(x^{k+1}, X) \leq \|x^{k+1} - s^k\|_G.$$

From the last three inequalities we get

$$\mathrm{dist}(x^{k+1}, X) \leq \|x^{k+1} - P_k(x^k)\|_G + \mathrm{dist}(P_k(x^k), X),$$

and, using (21) and (22), we have

$$\text{dist}(x^{k+1}, X) \le \frac{\rho_k}{\sqrt{\nu}} + \mu \, \text{dist}(x^k, X), \quad \text{for all } k \ge 0.$$

So the result of this case is obtained by Lemma 12 and (3). In case $f(x^k) = 0$, (11) becomes

$$\text{dist}(x^{k+1}, X) \le \mu \, \text{dist}(x^k, X), \quad \text{for all } k \ge 0,$$

and, by Lemma 12 with $\{a_k\}_{k=0}^{\infty} = \{\text{dist}(x^k, X)\}_{k=0}^{\infty}$ and $b_k = 0$, for all $k \ge 0$, the desired result is obtained. ∎

Lemma 16 *For any iterative sequence $\{x^k\}_{k=0}^{\infty}$, generated by Algorithm 7, $\lim_{k \to \infty} \|x^{k+1} - x^k\|_G = 0$.*

Proof. Assume first that $f(x^k) \ne 0$, using the triangle inequality with G-norms we have

$$
\begin{aligned}
\|x^{k+1} - x^k\|_G &= \|x^{k+1} - P_k(x^k) + P_k(x^k) - x^k\|_G \\
&\le \|x^{k+1} - P_k(x^k)\|_G + \|P_k(x^k) - x^k\|_G \\
&\le \frac{\rho_k}{\sqrt{\nu}} + (\text{dist}(x^k, H_k^-)), \quad \text{for all } k \ge 0, \qquad (23)
\end{aligned}
$$

where the last inequality follows from (22) and the equality

$$\text{dist}(x^k, H_k^-) = \|P_k(x^k) - x^k\|_G.$$

Since $X \subseteq H_k^-$, we have

$$\text{dist}(x^k, H_k^-) \le \text{dist}(x^k, X),$$

thus,

$$\|x^{k+1} - x^k\|_G \le \frac{\rho_k}{\sqrt{\nu}} + \text{dist}(x^k, X).$$

By Lemma 15 and (3) we now obtain the required result. In case $f(x^k) = 0$, (23) becomes

$$
\begin{aligned}
\|x^{k+1} - x^k\|_G &\le \|x^{k+1} - P_k(x^k)\|_G + \|P_k(x^k) - x^k\|_G = \|P_k(x^k) - x^k\|_G \\
&= \text{dist}(x^k, H_k^-) \le \text{dist}(x^k, X),
\end{aligned}
$$

where the last inequality follows from $X \subseteq H_k^-$. By Lemma 15 and taking limits as $k \to \infty$, we get the required result. ∎

Theorem 17 *Consider the $VIP(X, f)$ (1) and assume that Conditions 3–5 hold. Then any sequence $\{x^k\}_{k=0}^{\infty}$, generated by Algorithm 7, converges to the unique solution x^* of problem (1).*

Proof. By Lemma 15, $x^k \in X_\varepsilon$, for all sufficiently large k, where X_ε is the set given in Conditions 3 and 4. (We assume without loss of generality that the value of ε is common in both conditions.) From Condition 4 we have

$$\langle f(x^k) - f(x^*), x^k - x^* \rangle \geq \alpha \|x^k - x^*\|_2^2,$$

and

$$
\begin{aligned}
\langle f(x^k) - f(x^*), x^k - x^* \rangle &= \langle f(x^k) - f(x^*), x^k - x^{k+1} + x^{k+1} - x^* \rangle \\
&= \langle f(x^k), x^k - x^{k+1} \rangle + \langle f(x^k), x^{k+1} - x^* \rangle \\
&\quad - \langle f(x^*), x^k - x^* \rangle \\
&\geq \alpha \|x^k - x^*\|_2^2.
\end{aligned}
$$

So, we obtain

$$\langle f(x^k), x^{k+1} - x^* \rangle \geq \alpha \|x^k - x^*\|_2^2 + \langle f(x^*), x^k - x^* \rangle + \langle f(x^k), x^{k+1} - x^k \rangle. \tag{24}$$

Let λ be an arbitrary positive number. Since x^* satisfies (1), it follows from Lemmas 13 and 15 that the following inequalities hold, for all sufficiently large k,

$$\langle f(x^*), x^k - x^* \rangle \geq -\lambda. \tag{25}$$

Using the Cauchy-Schwarz inequality

$$\langle f(x^k), x^{k+1} - x^k \rangle \geq -\|f(x^k)\|_2 \|x^{k+1} - x^k\|_2.$$

By Lemma 13, $\{x^k\}_{k=0}^{\infty}$ is bounded, and with Condition 3 we get that $\{f(x^k)\}_{k=0}^{\infty}$ is also bounded. Lemma 16 guarantees

$$\langle f(x^k), x^{k+1} - x^k \rangle \geq -\lambda, \tag{26}$$

for all sufficiently large k. Applying (25) and (26) to (24), we obtain

$$\langle f(x^k), x^{k+1} - x^* \rangle \geq \alpha \|x^k - x^*\|_2^2 - 2\lambda. \tag{27}$$

Let us divide the indices of $\{x^k\}_{k=0}^{\infty}$ as follows

$$\Gamma := \{k \geq 0 \mid f(x^k) = 0\} \text{ and } \bar{\Gamma} := \{k \geq 0 \mid f(x^k) \neq 0\}.$$

14

Equation (27) implies, due to the arbitrariness of λ, that for $f(x^k) = 0$

$$\lim_{k \to \infty,\ k \in \Gamma} x^k = x^*.$$

Now consider the indices in $\bar{\Gamma}$ and suppose that there exists a $\zeta > 0$ such that

$$\|x^k - x^*\|_2 \geq \zeta, \quad \text{for all } k \in \bar{\Gamma}. \tag{28}$$

By Lemma 11

$$
\begin{aligned}
\|x^{k+1} - x^*\|_G^2 &= \|P_k(z^k) - x^*\|_G^2 \leq \|z^k - x^*\|_G^2 - \|P_k(z^k) - z^k\|_G^2 \\
&= \left\| \left(x^k - \rho_k G^{-1} \frac{f(x^k)}{\|f(x^k)\|} \right) - x^* \right\|_G^2 \\
&\quad - \left\| x^{k+1} - \left(x^k - \rho_k G^{-1} \frac{f(x^k)}{\|f(x^k)\|} \right) \right\|_G^2 \\
&= \|x^k - x^*\|_G^2 - \|x^{k+1} - x^k\|_G^2 - 2\frac{\rho_k}{\|f(x^k)\|_2} \langle f(x^k), x^{k+1} - x^* \rangle \\
&\leq \|x^k - x^*\|_G^2 - 2\frac{\rho_k}{\|f(x^k)\|_2} \langle f(x^k), x^{k+1} - x^* \rangle.
\end{aligned}
$$

So, we get

$$\|x^{k+1} - x^*\|_G^2 \leq \|x^k - x^*\|_G^2 - 2\frac{\rho_k}{\|f(x^k)\|_2} \langle f(x^k), x^{k+1} - x^* \rangle.$$

Since λ in (27) is arbitrary, we choose $\lambda = \alpha \zeta^2 / 4$,

$$\langle f(x^k), x^{k+1} - x^* \rangle \geq \alpha \|x^k - x^*\|_2^2 - \frac{\alpha \zeta^2}{2},$$

provided that k are sufficiently large. So,

$$
\begin{aligned}
-2\frac{\rho_k}{\|f(x^k)\|_2} \langle f(x^k), x^{k+1} - x^* \rangle &\leq -2\frac{\rho_k}{\|f(x^k)\|_2} \left(\alpha \|x^k - x^*\|_2^2 - \frac{\alpha \zeta^2}{2} \right) \\
&\leq -\frac{\alpha \zeta^2 \rho_k}{\|f(x^k)\|_2}.
\end{aligned}
$$

By Lemma 13, $\{x^k\}_{k=0}^{\infty}$ is bounded, thus Condition 3 implies that $\{f(x^k)\}_{k=0}^{\infty}$ is also bounded. So there exists a $\tau > 0$ such that

$$-\frac{\alpha \zeta^2 \rho_k}{\|f(x^k)\|_2} \leq -\frac{\alpha \zeta^2 \rho_k}{\tau}, \quad \text{for all } k \in \bar{\Gamma},$$

so that,

$$\|x^{k+1} - x^*\|_G^2 \leq \|x^k - x^*\|_G^2 - \frac{\alpha\zeta^2\rho_k}{\tau},$$

provided that k are sufficiently large. Then there exists an integer $\overline{k} \in \overline{\Gamma}$ such that

$$\|x^{k+1} - x^*\|_G^2 \leq \|x^k - x^*\|_G^2 - \frac{\alpha\zeta^2\rho_k}{\tau}, \quad \text{for all } k \in \overline{\Gamma} \text{ such that } k \geq \overline{k}. \quad (29)$$

By adding the inequalities (29) from $k = \overline{k}$ to $\overline{k} + \ell$, over indices $k \in \overline{\Gamma}$, we have

$$\|x^{\overline{k}+\ell+1} - x^*\|_G^2 \leq \|x^{\overline{k}} - x^*\|_G^2 - \frac{\alpha\zeta^2}{\tau} \sum_{k \in \overline{\Gamma},\ k=\overline{k}}^{\overline{k}+\ell} \rho_k,$$

for any $\ell > 0$. However, this is impossible due to (3), so there exists no $\zeta > 0$ such that (28) is satisfied. Thus, $\{x^k\}_{k\in\overline{\Gamma}}$ contains a subsequence $\{x^k\}_{k\in\hat{\Gamma}}$, $\hat{\Gamma} \subseteq \overline{\Gamma}$ converging to x^*, and so there is a subsequence $\{x^k\}_{k\in\Gamma\cup\hat{\Gamma}}$ of the whole sequence $\{x^k\}_{k=0}^{\infty}$ converging to x^*. In order to prove that the entire sequence $\{x^k\}_{k=0}^{\infty}$ is convergent to x^*, suppose to the contrary that there exists a subsequence of $\{x^k\}_{k=0}^{\infty}$ converging to $\hat{x} \in X$, $\hat{x} \neq x^*$. Since, by Lemma 16, $\lim_{k\to\infty} \|x^{k+1} - x^k\|_G = 0$, there must exist a $\zeta > 0$ and an arbitrarily large integer $j \in \overline{\Gamma}$ such that

$$\|x^j - x^*\|_G \geq \zeta \quad \text{and} \quad \|x^{j+1} - x^*\|_G \geq \|x^j - x^*\|_G. \quad (30)$$

However, if j is sufficiently large, we may apply an argument similar to that used to derive (29), and obtain the inequality

$$\|x^{j+1} - x^*\|_G < \|x^j - x^*\|_G,$$

which contradicts (30). Therefore, the sequence $\{x^k\}_{k=0}^{\infty}$ must converge to the solution x^*. ∎

4 Special Cases of the Algorithmic Scheme

Our general algorithmic scheme (Algorithm 7) includes as a special case the classical projection method of Algorithm 1. This can be seen by using $\delta = 1$ in Algorithm 7, which reduces the family of potential half-spaces to a singleton

which includes only the half-space that supports X at the projection of z^k (of (4)) onto X.

Another important special case is obtained from Algorithm 7 if we choose the convex function $g(x) := \text{dist}(x, X)$. This is no restriction since any convex set X can be presented in this way and Fukushima's method of Algorithm 2 is obtained as a special case of our algorithmic scheme.

4.1 An interior anchor point algorithm

To illustrate that additional algorithms can be derived from Algorithm 7 we present below an algorithm that uses other hyperplanes to project on. This particular realization requires that (the interior) $\text{int}\, X$ is nonempty. The idea of using an interior point as an anchor to generate a separating hyperplane appeared previously in [1] for the convex feasibility problem, and in [9] for an outer approximation method.

Algorithm 18
Initialization*: Let $y \in \text{int}\, X$ be fixed and given. Select an arbitrary starting point $x^0 \in R^n$ and set $k = 0$.*
Iterative step*: Given the current iterate x^k,*
(1) if $x^k \in X$ set $x^k = x^{k+1}$ and stop.
(2) Otherwise, calculate the "shifted point"

$$z^k := \begin{cases} x^k - \rho_k G^{-1} f(x^k)/\|f(x^k)\|_2, & \text{if } f(x^k) \neq 0, \\ x^k, & \text{if } f(x^k) = 0. \end{cases} \qquad (31)$$

and construct the line L_k through the points x^k and y.
(3) Denote by w^k the point closet to x^k in the set $L_k \cap X$.
(4) Construct a hyperplane H_k separating x^k from X and supporting X at w^k.
(5) Compute $x^{k+1} = P_{H_k^-}(z^k)$, where H_k^- is the half-space whose bounding hyperplane is H_k and $X \subseteq H_k^-$, set $k = k + 1$ and return to (1).

The iterative step of this algorithm is illustrated in Figure 2. We show that Algorithm 18 generates sequences that converge to a solution of problem (1) by showing that it is a special case of Algorithm 7.

Theorem 19 *Consider the $VIP(X, f)$ (1) and assume that Conditions 3–5 hold and that $\text{int}\, X \neq \emptyset$. Then any sequence $\{x^k\}_{k=0}^{\infty}$, generated by Algorithm 18, converges to the solution x^* of problem (1).*
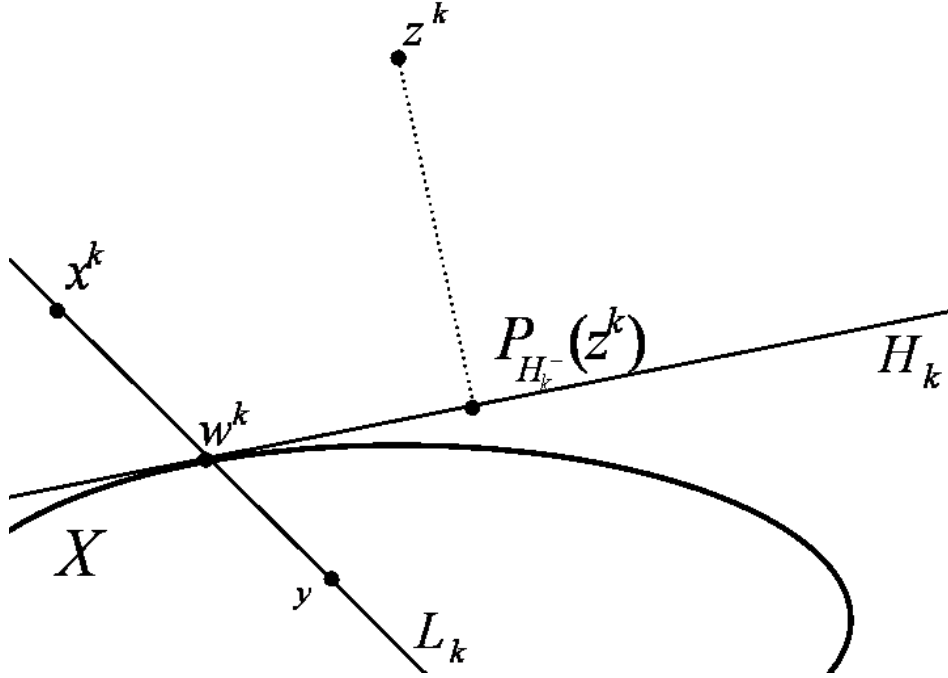
17

Figure 2: Illustration of Algorithm 18: Interior anchor point

Proof. First assume that $G = I$, the unit matrix. Algorithm 18 is obviously a special case of Algorithm 7 where we choose at each step a separating hyperplane which also supports X at the point w^k. The stopping criterion is valid by Theorem 9. In order to invoke Theorem 17 we have to show that for such an algorithm $\delta > 0$ always holds. By Lemma 13, $\{x^k\}_{k=0}^{\infty}$ is bounded and, since $x^k \notin X$, we have

$$\|P_{H_k^-}(x^k) - x^k\|_2 = \frac{\|x^k - w^k\|_2\|y - P_{H_k}(y)\|_2}{\|y - w^k\|_2}, \tag{32}$$

and we also have

$$\|x^k - w^k\|_2 \geq \mathrm{dist}(x^k, X).$$

Defining $d := \mathrm{dist}(y, \mathrm{bd}\, X)$, since $y \in \mathrm{int}\, X$,

$$\|y - P_{H_k}(y)\|_2 \geq d > 0.$$

18

From the boundedness of $\{x^k\}_{k=0}^\infty$ we know that there exists a positive N such that $\|y - w^k\|_2 \leq N$, for all $k \geq 0$. Combining these inequalities with (32) implies that

$$\|P_{H_k^-}(x^k) - x^k\|_2 \geq \frac{d}{N} \operatorname{dist}(x^k, X),$$

which shows that the algorithm is of the same type of Algorithm 7 with $\delta := d/N > 0$.

To show convergence for a general symmetric positive definite matrix G we recall that all norms are equivalent in R^n so that there exists constants M_1 and M_2 such that, for all $x \in R^n$,

$$M_1\|x\|_2 \leq \|x\|_G \leq M_2\|x\|_2.$$

Actually,
$$\|x\|_G \leq \rho(G)\|x\|_2,$$

where $\rho(G)$ is the largest eigenvalue of G (see, e.g., [6, Equation (5.2), page 199]). So, we get

$$\|x^k - w^k\|_2 \geq \frac{1}{\rho(G)}\|x^k - w^k\|_G \geq \frac{1}{\rho(G)} \operatorname{dist}(x^k, X),$$

thus,

$$\|y - P_{H_k}(y)\|_2 \geq \frac{1}{\rho(G)}\|y - P_{H_k}(y)\|_G \geq \frac{\widehat{d}}{\rho(G)},$$

where $\widehat{d} := \operatorname{dist}(y, \operatorname{bd} X)$. Also,

$$\|y - w^k\|_G \leq M_2\|y - w^k\|_2 \leq M_2 N.$$

So, by (32) and the last three inequalities, we get

$$\|P_{H_k^-}(x^k) - x^k\|_G \geq M_1 \cdot \|P_{H_k^-}(x^k) - x^k\|_2 = M_1 \cdot \frac{\|x^k - w^k\|_2\|y - P_{H_k}(y)\|_2}{\|y - w^k\|_2}$$

$$\geq M_1 \cdot \frac{\frac{1}{\rho(G)} \operatorname{dist}(x^k, X) \cdot \frac{\widehat{d}}{\rho(G)}}{M_2 N} = \widehat{M} \cdot \frac{\widehat{d}}{N} \operatorname{dist}(x^k, X),$$

(33)

19

where $\widehat{M} := M_1/(M_2(\rho(G))^2)$ and this completes the proof. ∎

References

[1] R. Aharoni, A. Berman and Y. Censor, An interior points algorithm for the convex feasibility problem, *Advances in Applied Mathematics* **4** (1983), 479–489.

[2] A. Auslender, *Optimisation: Méthodes Numériques*, Masson, Paris, 1976.

[3] A. Auslender and M. Teboulle, Interior projection-like methods for monotone variational inequalities, *Mathematical Programming*, Series A, **104** (2005), 39–68.

[4] Y. Censor and S.A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications*, Oxford University Press, New York, NY, USA, 1997.

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

[6] J.W. Demmel, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, PA, USA, 1997.

[7] F. Facchinei and J.S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume I and Volume II,* Springer-Verlag, New York, NY, USA, 2003.

[8] M. Fukushima, An outer approximation algorithm for solving general convex programs, *Operations Research* **31** (1983), 101–113.

[9] M. Fukushima, On the convergence of a class of outer approximation algorithms for convex programs, *Journal of Computational and Applied Mathematics* **10** (1984), 147–156.

[10] M. Fukushima, A relaxed projection method for variational inequalities, *Mathematical Programming* **35** (1986), 58–70.

[11] D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and Their Applications*, Academic Press, New York, NY,USA, 1980.

[12] M. Aslam Noor, Some developments in general variational inequalities, *Applied Mathematics and Computation* **152** (2004), 197–277.

[13] N. Xiu and J. Zhang, Some recent advances in projection-type methods for variational inequalities, *Journal of Computational and Applied Mathematics* **152** (2003), 559–585.

[14] I. Yamada and N. Ogura, Hybrid steepest descent method for variational inequality problem over the fixed point set of certain quasi-nonexpansive mappings, *Numerical Functional Analysis and Optimization* **25** (2004), 619–655.

[15] Q. Yang, On variable-step relaxed projection algorithm for variational inequalities, *Journal of Mathematical Analysis and Applications* **302** (2005), 166–179.