Sharp Penalty Mappings for Variational Inequality Problems

Evgeni Nurminski

Far Eastern Federal University, School of Natural Sciences,
Vladivostok, Russia
nurminski.ea@dvfu.ru

Abstract. First, this paper introduces a notion of a sharp penalty mapping which can replace more common exact penalty function for convex feasibility problems. Second, it uses it for solution of variational inequalities with monotone operators or pseudo-varitional inequalities with oriented operators. Appropriately scaled the sharp penalty mapping can be used as an exact penalty in variational inequalities to turn them into fixed point problems. Then they can be approximately solved by simple iteration method.

Keywords: monotone variational inequalities, oriented mappings, sharp penalty mappings, exact penalty mappings, approximate solution

Introduction

Variational inequalities (VI) became one of the common tools for representing many problems in physics, engineering, economics, computational biology, computerized medicine, to name but a few, which extend beyond optimization, see [1] for the extensive review of the subject. Apart from the mathematical problems connected with the characterization of solutions and development of the appropriate algorithmic tools to find them, modern problems offer significant implementation challenges due to their non-linearity and large scale. It leaves just a few options for the algorithms development as it occurs in the others related fields like convex feasibility (CF) problems [2] as well. One of these options is to use fixed point iteration methods with various attraction properties toward the solutions, which have low memory requirements and simple and easily parallelized iterations. These schemes are quite popular for convex optimization and CF problems but they need certain modifications to be applied to VI problems. The idea of modification can be related to some approaches put forward for convex optimization and CF problems in [3–5] and which is becoming known as superiorization technique (see also [6] for the general description).

From the point of view of this approach the optimization problem

$$\min f(x), \quad x \in X \tag{1}$$

or VI problem to find $x^\star \in X$ such that

$$F(x^\star)(x - x^\star) \geq 0, \quad x \in X \tag{2}$$

are conceptually divided into the feasibility problem $x \in X$ and the second-stage optimization or VI problems next. Then these tasks can be considered to a certain extent separately which makes it possible to use their specifics to apply the most suitable algorithms for feasibility and optimization/VI parts.

The problem is to combine these algorithms in a way which provides the solution of the original problems (1) or (2). As it turns out these two tasks can be merged together under rather reasonable conditions which basically require a feasibility algorithm to be resilient with respect to diminishing perturbations and the second-stage algorithm to be something like globally convergent over the feasible set or its small expansions.

Needless to say that this general idea meets many technical difficulties one of them is to balance in intelligent way the feasibility and optimization/VI steps. If optimization steps are essentially "smaller" than feasibility steps then it is possible to prove general convergence results [3, 4] under rather mild conditions. However it looks like that this requirements for optimization steps to be smaller (in fact even vanishing compared to feasibility) slows down the overall optimization in (1) considerably.

This can be seen in the text-book penalty function method for (1) which consists in the solution of the auxiliary problem of the kind

$$\min_x \{ \Phi_X(x) + \epsilon f(x) \} = \Phi_X(x_\epsilon) + \epsilon f(x_\epsilon) \tag{3}$$

where $\Phi_X(x) = 0$ for $x \in X$ and $\Phi_X(x) > 0$ otherwise. The term $\epsilon f(x)$ can be considered as the perturbation of the feasibility problem $\min_x \Phi(x)$ and for classical smooth penalty functions the penalty parameter $\epsilon > 0$ must tend to zero to guarantee convergence of x_ϵ to the solution of (1). Definitely it makes the objective function $f(x)$ less influential in solution process of (1) and hinders the optimization.

To overcome this problem the exact penalty functions $\Psi_X(\cdot)$ can be used which provide the exact solution of (1)

$$\min_x \{ \Psi_X(x) + \epsilon f(x) \} = \epsilon f(x^\star) \tag{4}$$

for small enough $\epsilon > 0$ under rather mild conditions. The price for the conceptual simplification of the solution of (2) is the inevitable non-differentiability of the penalty function $\Psi_X(x)$ and the corresponding worsening of convergence rates for instance for gradient-like methods (see [8, 9] for comparison). Nevertheless the idea has a certain appeal, keeping in mind successes of nondiffereniable optimization, and the similar approaches with necessary modifications were used for VI problems starting from [10] and followed by [11–14] among others. In these works the penalty functions were introduced and their gradient fields direct the iterations to feasibility.

Here we suggest a more general definition of a sharp penalty mapping $P : E \to \mathcal{C}(E)$, not necessarily potential, which is oriented toward a feasible set (for details of notations see the section 1). It also admits a certain problem-dependent penalty constant $\lambda > 0$ such that the sum $F + \lambda P$ of variational operator F of

(2) and P scaled by λ possesses a desirable properties to make the iteration algorithm converge at least to a given neighborhood of solution of (2). In the preliminary form this idea was suggested in [15] using a different definition of a sharp penalty mapping which resulted in rather weak convergence result. Here we show that it is possible to reach stronger convergence result with all limit points of the iteration method being an ϵ-solutions of VI problem (2).

1 Notations and Preliminaries

Let E denotes a finite-dimensional space with the inner product xy for $x, y \in E$, and the standard Euclidean norm $\|x\| = \sqrt{xx}$. The one-dimensional E is denoted as $\mathbb{R}$ and $\mathbb{R}_\infty = \mathbb{R} \cup \{\infty\}$. The unit ball in E is denoted as $B = \{x : \|x\| \leq 1\}$. The space of bounded closed convex subsets of E is denoted as $\mathcal{C}(E)$. The distance function $\rho(x, X)$ between point x and set $X \subset E$ is defined as $\rho(x, X) = \inf_{y \in X} \|x - y\|$. The norm of a set X is defined as $\|X\| = \sup_{x \in X} \|x\|$.

For any $X \subset E$ its interior is denoted as $\mathrm{int}(X)$, the closure of X is denoted as $\mathrm{cl}(X)$ and the boundary of X is denoted as $\partial X = X \setminus \mathrm{int}(X)$.

The sum of two subsets A and B of E is denoted as $A + B$ and understood as $A + B = \{a + b, a \in A, b \in B\}$. If A is a singleton $\{a\}$ we write just $a + B$.

Any open subset of E containing zero vector is called a neighborhood of zero in E. We use the standard definition of upper semi-continuity and monotonicity of set-valued mappings:

Definition 1. *A set-valued mapping $F : E \to \mathcal{C}(E)$ is called upper semi-continuous if at any point $\bar{x}$ for any neighborhood of zero U there exists a neighborhood of zero V such that $F(x) \subset F(\bar{x}) + U$ for all $x \in \bar{x} + V$.*

Definition 2. *A set-valued mapping $F : E \to \mathcal{C}(E)$ is called a monotone if $(f_x - f_y)(x - y) \geq 0$ for any $x, y \in E$ and $f_x \in F(x), f_y \in F(y)$.*

We use standard notations of convex analysis: if $h : E \to \mathbb{R}_\infty$ is a convex function, then $\mathrm{dom}(h) = \{x : h(x) < \infty\}$, $\mathrm{epi}\, h = \{(\mu, x) : \mu \geq h(x), x \in \mathrm{dom}(h)\} \subset \mathbb{R} \times E$, the sub-differential of h is defined as follows:

Definition 3. *For a convex function $h : E \to \mathbb{R}_\infty$ a sub-differential of h at point $\bar{x} \in \mathrm{dom}(h)$ is the set $\partial h(\bar{x})$ of vectors g such that $h(x) - h(\bar{x}) \geq g(x - \bar{x})$ for any $x \in \mathrm{dom}(h)$.*

This defines a convex-valued upper semi-continuous maximal monotone set-valued mapping $\partial h : \mathrm{int}(\mathrm{dom}(h)) \to \mathcal{C}(E)$. At the boundaries of $\mathrm{dom}(h)$ the sub-differential of h may or may not exist. For differentiable $h(x)$ the classical gradient of h is denoted as $h'(x)$.

We define the convex envelope of $X \subset E$ as follows.

Definition 4. *An inclusion-minimal set $Y \in \mathcal{C}(E)$ such that $X \subset Y$ is called a convex envelope of X and denoted as $\mathrm{co}(X)$.*

Our main interest consists in finding a solution $x^\star$ of a following finite-dimensional VI problem with a single-valued operator $F(x)$:

$$\text{Find } x^\star \in X \subset \mathcal{C}(E) \text{ such that } F(x^\star)(x - x^\star) \geq 0 \text{ for all } x \in X. \tag{5}$$

This problem has its roots in convex optimization and for $F(x) = f'(x)$ VI (5) is the geometrical formalization of the optimality conditions for (1).

If F is monotone, then the pseudo-variational inequality (PVI) problem

$$\text{Find } x^\star \in X \text{ such that } F(x)(x - x^\star) \geq 0 \text{ for all } x \in X. \tag{6}$$

has a solution $x^\star$ which is a solution of (5) as well. However it is not necessary for F to be monotone to have a solution of (6) which coincides with a solution of (5) as Fig. 1 demonstrates.

For simplicity we assume that both problems (5) and (6) has unique and hence coinciding solutions.

To have more freedom to develop iteration methods for the problem (6) we introduce the notions of oriented and strongly oriented mappings according to the following definitions.

Definition 5. *A set-valued mapping $G : E \to \mathcal{C}(E)$ is called oriented toward $\bar{x}$ at point x if*

$$g_x(x - \bar{x}) \geq 0 \tag{7}$$

for any $g_x \in G(x)$.

Definition 6. *A set-valued mapping $G : E \to \mathcal{C}(E)$ is called strongly oriented toward $\bar{x}$ on a set X if for any $\epsilon > 0$ there is $\gamma_\epsilon > 0$ such that*

$$g_x(x - \bar{x}) \geq \gamma_\epsilon \tag{8}$$

for any $g_x \in G(x)$ and all $x \in X \setminus \{\bar{x} + \epsilon B\}$.

If G is oriented (strongly oriented) toward $\bar{x}$ at all points $x \in X$ then we will call it oriented (strongly oriented) toward $\bar{x}$ on X.

Of course if $\bar{x} = x^\star$, a solution of PVI problem (6), then G is oriented toward $x^\star$ on X by definition and the other way around.

The notion of oriented mappings is somewhat related to attractive mappings introduced in [2], which can be defined for our purposes as follows.

Definition 7. *A mapping $F : E \to E$ is called attractive with respect to $\bar{x}$ at point x if*

$$\|F(x) - \bar{x}\| \leq \|x - \bar{x}\| \tag{9}$$

It is easy to show that if F is an attractive mapping, then $G(x) = F(x) - x$ is an oriented mapping, however $G(x) = -10x$ is the oriented mapping toward $\{0\}$ on $[-1, 1]$ but neither $G(x)$ nor $G(x) + x$ are attractive.

Despite the fact that the problem (5) depends upon the behavior of F on X only, we need to make an additional assumption about global properties of F to avoid certain problems with possible divergence of iteration method due to "run-away" effect. Such assumption is the long-range orientation of F which is frequently used to ensure the desirable global behavior of iteration methods.

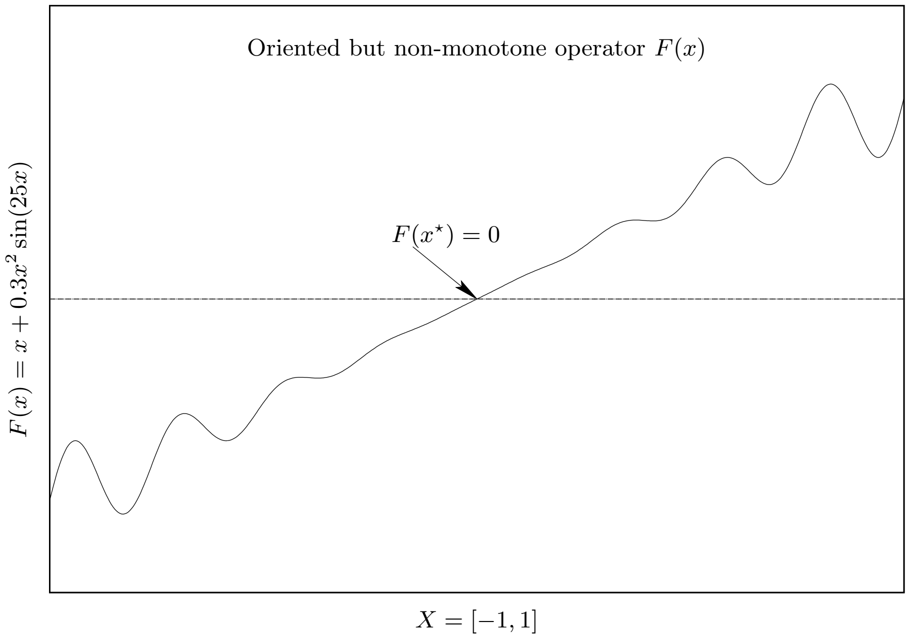

Fig. 1. Non-monotone operator $F(x)$ oriented toward $x^\star = 0$.

Definition 8. *A mapping $F : E \to E$ is called long-range oriented toward a set X if there exists $\rho_F \geq 0$ such that for any $\bar{x} \in X$*

$$F(x)(x - \bar{x}) > 0 \textit{ for all } x \textit{ such that } \|x\| \geq \rho_F \tag{10}$$

.

We will call ρ_F the radius of long-range orientation of F toward X.

2 Sharp Penalty Mappings

In this section we present the key construction which makes possible to reduce an approximate solution of VI problem into calculation of the limit points of iterative process, governed by strongly oriented operators.

For this purpose we modify slightly the classical definition of a polar cone of a set X.

Definition 9. *The set $K_X(x) = \{p : p(x - y) \geq 0$ for all $y \in X\}$ we will call the polar cone of X at a point x.*

Of course $K_X(x) = \{0\}$ if $x \in \operatorname{int} X$.

For our purposes we need also a stronger definition which defines a certain sub-cone of $K_X(x)$ with stronger pointing toward X.

Definition 10. *Let $\epsilon \geq 0$ and $x \notin X + \epsilon B$. The set*

$$K^\epsilon_X(x) = \{p : p(x - y) \geq 0 \textit{ for all } y \in X + \epsilon B\} \tag{11}$$

will be called ϵ-strong polar cone of X at x.

As it is easy to see that the alternative definition of $K^\epsilon_X(x)$ is $K^\epsilon_X(x) = \{p : p(x - y) \geq \epsilon\|p\|$ for all $y \in X.\}$

To define a sharp penalty mapping for the whole space E we introduce a composite mapping

$$\tilde{K}^\epsilon_X(x) = \begin{cases} \{0\} & \text{if } x \in X \\ K_X(x) & \text{if } x \in \operatorname{cl}\{\{X + \epsilon B\} \setminus X\} \\ K^\epsilon_X(x) & \text{if } x \in \rho_F B \setminus \{X + \epsilon B\} \end{cases} \tag{12}$$

Notice that $\tilde{K}^\epsilon_X(x)$ is upper semi-continuous by construction.

Now we define a sharp penalty mapping for X as

$$P^\epsilon_X(x) = \{p \in \tilde{K}^\epsilon_X(x), \|p\| = 1\}. \tag{13}$$

Clear that $P^\epsilon_X(x)$ is not defined for $x \in \operatorname{int}\{X\}$ but we can defined it to be equal to zero on $\operatorname{int}\{X\}$ and take a convex envelope of $P^\epsilon_X(x)$ and $\{0\}$ at the boundary of X to preserve upper semi-continuity.

For some positive λ define $F_\lambda(x) = F(x) + \lambda P^\epsilon_X(x)$. Of course by construction $F_\lambda(x)$ is upper semi-continuous for $x \notin X$.

For the further development we establish the following result on construction of an approximate globally oriented mapping related to the VI problem (5).

Lemma 1. *Let $X \subset E$ is closed and bounded, F is monotone and long-range oriented toward X with the radius of orientability ρ_F and strongly oriented toward solution $x^\star$ of (5) on X with the constants $\gamma_\epsilon > 0$ for $\epsilon > 0$, satisfying (8) and $P_X^\epsilon(\cdot)$ is a sharp penalty (13). Then for any sufficiently small $\epsilon > 0$ there exists $\lambda_\epsilon > 0$ and $\delta_\epsilon > 0$ such that for all $\lambda \geq \lambda_\epsilon$ a penalized mapping $F_\lambda(x) = F(x) + \lambda P_X^\epsilon(x)$ satisfies the inequality*

$$f_x(x - x^\star) \geq \delta_\epsilon \tag{14}$$

for all $x \in \rho_F B \setminus \{x^\star + \epsilon B\}$ and any $f_x \in F_\lambda(x)$.

Proof For monotone F we can equivalently consider a pseudo-variational inequality (6) with the same solution $x^\star$. Define the following subsets of E:

$$\begin{aligned} X_\epsilon^{(1)} &= X \setminus \{x^\star + \epsilon B\}, \\ X_\epsilon^{(2)} &= \{\{X + \epsilon B\} \setminus X\} \setminus \{x^\star + \epsilon B\}, \\ X_\epsilon^{(3)} &= \rho_F B \setminus \{\{X + \epsilon B\} \setminus \{x^\star + \epsilon B\}\}. \end{aligned} \tag{15}$$

Correspondingly we consider 3 cases.

Case A. $x \in X_\epsilon^{(1)}$. In this case $f_\lambda(x) = F(x)$ and therefore

$$f_\lambda(x)(x - x^\star) = F(x)(x - x^\star) \geq \gamma_\epsilon > 0. \tag{16}$$

Case B. $x \in X_\epsilon^{(2)}$. In this case $f_\lambda(x) = F(x) + \lambda p_X(x)$ where $p_X(x) \in K_X(x)$, $\|p_X(x)\| = 1$ and therefore

$$f_\lambda(x)(x - x^\star) = F(x)(x - x^\star) + \lambda p_X(x)(x - x^\star) \geq \gamma_\epsilon/2 > 0. \tag{17}$$

as $\lambda p_X(x)(x - x^\star) > 0$ by construction.

Case C. $x \in X_\epsilon^{(3)}$. In this case $f_\lambda(x) = F(x) + \lambda p_X(x)$ where $p_X(x) \in K_X^\epsilon(x)$, $\|p_X(x)\| = 1$. By continuity of F the norm of F is bounded on $\rho_F B$ by some M and as $P_X^\epsilon(\cdot)$ is ϵ-strong penalty mapping

$$\begin{aligned} f_\lambda(x)(x - x^\star) = F(x)(x - x^\star) + \lambda p_X(x)(x - x^\star) \geq \\ -M\|x - x^\star\| + \lambda\epsilon \leq -2\rho_F M + \lambda\epsilon \geq \rho_F M > 0 \end{aligned} \tag{18}$$

for $\lambda \geq \rho_F M/\epsilon$.

By combining all three bounds we obtain

$$f_\lambda(x)(x - x^\star) \geq \min\{\gamma_\epsilon/2, \rho_F M\} = \delta_\epsilon > 0 \tag{19}$$

for $\lambda \geq \Lambda_\epsilon = \rho_F M/\epsilon$ which completes the proof. ■

The elements of a polar cone for a given set X can be obtained by different means. The most common are either by projection onto set X:

$$x - \Pi_X(x) \in K_X(x) \tag{20}$$

where $\Pi_X(x) \in X$ is the orthogonal projection of x on X, or by subdifferential calculus when X is described by a convex inequality $X = \{x : h(x) \leq 0\}$. If there is a point $\bar{x}$ such that $h(\bar{x}) < 0$ (Slater condition) then $h(y) < 0$ for all $y \in \text{int}\{X\}$. Therefore $0 < h(x) - h(y) \leq g_h(x)(x - y)$ for any $y \in \text{int}\{X\}$. By continuity $0 < h(x) - h(y) \leq g_h(x)(x - y)$ for all $y \in X$ which means that $g_h \in K_X(x)$.

One more way to obtain $g_h \in K_X(x)$ relies on the ability to find some $x^c \in \text{int}\{X\}$ and use it to compute Minkowski function

$$\mu_X(x, x^c) = \inf_{\theta \geq 0}\{\theta : x^c + (x - x^c)\theta^{-1} \in X\} > 1 \text{ for } x \notin X. \tag{21}$$

Then by construction $\bar{x} = x^c + (x - x^c)\mu_X(x, x^c)^{-1} \in \partial X$, i.e. $h(\bar{x}) = 0$ and for any $g_h \in \partial h(\bar{x})$ the inequality $g_h\bar{x} \geq g_h y$ holds for any $y \in X$.

By taking $y = x^c$ obtain $g_h\bar{x} \geq g_h x^c$ and therefore

$$\begin{aligned} g_h\bar{x} = g_h x^c + g_h(x - x^c)\mu_X(x, x^c)^{-1} = \mu_X(x, x^c)^{-1}g_h x + (1 - \mu_X(x, x^c)^{-1})g_h x^c \leq \\ \mu_X(x, x^c)^{-1}g_h x + (1 - \mu_X(x, x^c)^{-1})g_h\bar{x}. \end{aligned} \tag{22}$$

Hence $g_h x \geq g_h\bar{x} \geq g_h y$ for any $y \in X$, which means that $g_h \in K_X(x)$.

As for ϵ-expansion of X it can be approximated from above (included into) by the relaxed inequality $X + \epsilon B \subset \{x : h(x) \leq L\epsilon\}$ where L is a Lipschitz constant in an appropriate neighborhood of X.

3 Iteration Algorithm

After construction of the mapping F_λ, oriented toward solution $x^\star$ of (6) at the whole space E except ϵ-neighborhood of $x^\star$ we can use it in an iterative manner like

$$x^{k+1} = x^k - \theta_k f^k,\ f^k \in F_\lambda(x^k),\ k = 0, 1, \ldots, \tag{23}$$

where $\{\theta_k\}$ is a certain prescribed sequence of step-size multipliers, to get the sequence of $\{x^k\}, k = 0, 1, \ldots$ which hopefully converges under some conditions to to at least the set $X_\epsilon = x^\star + \epsilon B$ of approximate solutions of (5).

For technical reasons, however, it would be convenient to guarantee from the very beginning the boundedness of $\{x^k\}, k = 0, 1, \ldots$. Possibly the simplest way to do so is to insert into the simple scheme (23) a safety device, which enforces restart if a current iteration x^k goes too far. This prevents the algorithm from divergence due to the "run away" effect and it can be easily shown that it keeps a sequence of iterations $\{x^k\}$ bounded.

Thus the final form of the algorithm is shown as the figure Algorithm 1, assuming that the set X, the operator F and the sharp penalty mapping P_X^ϵ satisfy conditions of the lemma 1. We prove convergence of the algorithm 1 under common assumptions on step sizes: $\theta_k \to +0$ when $k \to \infty$ and $\sum_{k=1}^K \theta_k \to \infty$ when $K \to \infty$. This is not the most efficient way to control the algorithm, but at the moment we are interested mostly in the very fact of convergence.

Data: The variational inequality operator F, sharp penalty mapping P_X, positive constant ϵ, penalty constant λ which satisfy conditions of the Lemma 1, long-range orientation radius ρ_F, a sequence of step-size multipliers $\{0 < \theta_k,\ k = 0, 1, 2, \dots\}$. and an initial point $x^0 \in \rho_F B$.

Result: The sequence of approximate solutions $\{x^k\}$ where every converging sub-sequence has a limit point which belongs to a set X_ϵ of ϵ-solution of variational inequality (5).

Initialization;

Define penalized mapping

$$F_\lambda(x) = F(x) + \lambda P_X(x), \tag{24}$$

and set the iteration counter k to 0;

while *The limit is not reached* **do**

Generate a next approximate solution x_{k+1}:

$$x^{k+1} = \begin{cases} x^k - \theta_k f^k,\ \ f^k \in F_\lambda(x^k), & \text{if } \|x^k\| \le 2\rho_F \\ x^0 & \text{otherwise.} \end{cases} \tag{25}$$

Increment iteration counter $k \longrightarrow k + 1$;

end

Complete: accept $\{x^k\}, k = 0, 1, \dots$ as a solution of (5). [a]

Algorithm 1: The generic structure of a conceptual version of the iteration algorithm with exact penalty.

[a] The exact meaning of this will be clarified in the convergence theorem 1.

Theorem 1. *Let $\epsilon > 0, \Lambda_\epsilon, F, P_X$ satisfy the assumptions of the Lemma 1, $\lambda > \Lambda_\epsilon$, and $\theta_k \to +0$ when $k \to \infty$ and $\sum_{k=1}^K \theta_k \to \infty$ when $K \to \infty$. Then all limit points of the sequence $\{x^k\}$ generated by the algorithm 1 belong to the set of ϵ-solutions $X_\epsilon = x^\star + \epsilon B$ of the problem (5).*

Proof We show first the boundedness of the sequence $\{x^k\}$. To do so it is sufficient to demonstrate that the sequence $\{\|x^k\|, k = 1, 2, \dots\}$ crosses the interval $[\rho_F, 2\rho_F]$ a finite number of times only (any way, from below or from above). Show first that the sequence $\{x^k\}$ leaves the set $\frac{3}{2}\rho_F B$ a finite number of times only. Define (a finite or not) set T of indices $T = \{t_k, k = 1, 2, \dots\}$ such that

$$\|x^\tau\| < \frac{3}{2}\rho_F \text{ and } \|x^{\tau+1}\| \geq \frac{3}{2}\rho_F. \tag{26}$$

If $\tau \in T$ then

$$\begin{aligned} \|x^{\tau+1}\|^2 = \|x^\tau - \theta_k F_\Lambda(x^\tau)\|^2 = \|x^\tau\|^2 - 2\theta_\tau f^\tau x^\tau + \theta_\tau^2 \|f^\tau\|^2 \leq \\ \|x^\tau\|^2 - 2\theta_\tau f^\tau x^\tau + \theta_\tau^2 C^2 \leq \|x^\tau\|^2 - 2\theta_\tau \gamma\delta + \theta_\tau^2 C^2, \end{aligned} \tag{27}$$

where C is an upper bound for $\|F_\lambda(x)\|$ with $x \in 2\rho_F B$ and $\gamma > 0$ is a lower bound for $f^\tau x$ for $x \in 2\rho_F B$ and $f^\tau \in F_\lambda(x)$. For τ large enough $\theta_\tau < \gamma C^{-2}$ and hence

$$\|x^{\tau+1}\|^2 \leq \|x^\tau\|^2 - \theta_\tau \delta < \|x^\tau\|^2 \tag{28}$$

which contradicts the definition of the set T. Therefore T is a finite set and the sequence $\{x^k\}$ leaves the set $\frac{3}{2}\rho_F B$ a finite number of times only which proves the boundedness of $\{x^k\}$.

Define now $W(x) = \|x - x^\star\|^2$ and notice that due to the boundedness of $\{x^k\}$ and semi-continuity of $F_\lambda(x)$ and etc, $W(x^{k+1}) - W(x^k) \to 0$ when $k \to \infty$. It implies that the limit set

$$W_\star = \{w_\star : \text{ the sub-sequence } \{x^{k_s}\} \text{ exists such that } \lim_{s\to\infty} W(x^{k_s}) = w_\star\} \tag{29}$$

is a certain interval $[w_\star^l, w_\star^u] \subset \mathbb{R}_+$ and the statement of the theorem means that $w_\star^u \leq \epsilon^2$.

To prove this we assume contrary, that is $w_\star^u > \epsilon^2$ and hence there exists a sub-sequence $\{x^{k_s}, s = 1, 2, \dots\}$ such that $\lim_{s\to\infty} W(x^{k_s}) = w' > \epsilon^2$. Without loss of generality we may assume that $\lim_{s\to\infty} x^{k_s} = x'$ and of course $x' \notin X_\epsilon$. Therefore $f'(x' - x^\star) > 0$ for any $f' \in F_\lambda(x')$ and by upper semi-continuity of F_λ there exists an $\upsilon > 0$ such that $F_\lambda(x)(x - x^\star) \geq \delta$ for all $x \in x' + 4\upsilon B$ and some $\delta > 0$. Again without loss of generally we may assume that $\upsilon < (\sqrt{w'} - \epsilon)/4$ so $(x' + 4\upsilon B) \cap (x^\star + \epsilon B) = \emptyset$.

For for s large enough $x^{k_s} \in x' + \upsilon B$ and let us assume that for all $t > k_s$ the sequence $\{x^t, t > k_s\} \subset x^{k_s} + \upsilon B \subset x' + 2\upsilon B$.

Then

$$\begin{aligned} W(x^{t+1}) = \|x^t - \theta_t F_\lambda(x^t) - x^\star\|^2 = W(x^t) - 2\theta_t F_\lambda(x^t)(x^t - x^\star) + \theta_t^2 \|F_\lambda(x^t)\|^2 \leq \\ W(x^t) - 2\theta_t F_\lambda(x^t)(x^t - x^\star) + \theta_t^2 C^2 \leq W(x^t) - 2\theta_t \delta + \theta_t^2 C^2 < W(x^t) - \theta_t \delta, \end{aligned} \tag{30}$$

for all $t > k_s$ and s large enough that $\sup_{t>k_s} \theta_t < \delta/C^2$. Summing up last inequalities from $t = k_s$ to $t = T - 1$ obtain

$$W(x^T) \leq W(x^{k_s}) - \delta \sum_{t=k_s}^{T-1} \theta_t \to -\infty \tag{31}$$

when $T \to \infty$ which is of course impossible.

Hence for each k_s there exists $r_s > k_s$ such that $\|x^{k_s} - x^{r_s}\| > \upsilon > 0$ Assume that r_s is in fact a minimal such index, i.e. $\|x^{k_s} - x^t\| \leq \upsilon$ for all t such that $k_s < t < r_s$ or $x^t \in x^{t_k} + \upsilon B \subset x' + 2\upsilon B$ for all such t. Without any loss of generality we may assume that $x^{r_s} \to x''$ where by construction $\|x' - x''\| \geq \upsilon > 0$ and therefore $x' \neq x''$.

As all conditions which led to (31) hold for $T = r_s$ then by letting $T = r_s$ we obtain

$$W(x^{r_s}) \leq W(x^{k_s}) - \delta \sum_{t=k_s}^{r_s-1} \theta_t. \tag{32}$$

On the other hand

$$\upsilon < \|x^{k_s} - x^{r_s}\| \leq \sum_{t=k_s}^{r_s} \|x^{t+1} - x^t\| \leq \sum_{t=k_s}^{r_s-1} \theta_t \|F_\lambda(x^t)\| \leq K \sum_{t=k_s}^{r_s} \theta_t \tag{33}$$

where K is the upper estimate of the norm of $F_\lambda(x)$ on $2\rho_F B$.

Therefore $\sum_{t=k_s}^{r_s-1} \theta_t > \upsilon/K > 0$ and finally

$$W(x^{r_s}) \leq W(x^{k_s}) - \delta\upsilon/K. \tag{34}$$

Passing to the limit when $s \to \infty$ obtain $W(x'') \leq W(x') - \delta\upsilon/K < W(x')$ Also $W(x'') > \epsilon^2$ as $x'' \in x' + 4\upsilon B$ which does not intersect with $x^\star + \epsilon B$. To save on notations denote $W(x') = w'$ and $W(x'') = w''$.

In other words, assuming that $w' > \epsilon^2$ we constructed another limit point w'' of the sequence $\{W(x^k)\}$ such that $\epsilon^2 < w'' < w'$. It follows from this that the sequence $\{W(x^k)\}$ infinitely many times crosses any sub-interval $[\tilde{w}'', \tilde{w}'] \subset (w'', w')$ both in "up" and "down" directions and hence there exist sub-sequences $\{p_s, s = 1, 2, \dots\}$ and $\{q_s, s = 1, 2, \dots\}$ such that $p_s < q_s$ and

$$W(x^{p_s}) \leq \tilde{w}'', W(x^{q_s}) \geq \tilde{w}', W(x^t) \in (w'', w') \text{ for } p_s < t < q_s \tag{35}$$

Then

$$0 < W(x^{q_s}) - W(x^{p_s}) = \sum_{t=p_s}^{q_s-1} (W(x^{t+1}) - W(x^t)) \tag{36}$$

and hence for any s there is an index $t_s : p_s < t_s < q_s$ such that

$$0 < W(x^{t_s+1}) - W(x^{t_s}). \tag{37}$$

However as $W(x^{t_s}) > w''$, $\quad x^{t_s} \notin x^\star + \epsilon B$ and therefore

$$\begin{array}{c} W(x^{t_s+1}) - W(x^{t_s}) = \|x^{t_s+1} - x^\star\|^2 - \|x^{t_s} - x^\star\|^2 = \\ \|x^{t_s} - x^\star + \theta_{t_s} f^{t_s}\|^2 - \|x^{t_s} - x^\star\|^2 = \\ 2\theta_{t_s} f^{t_s}(x^{t_s} - x^\star) + \theta_{t_s}^2 \|f^{t_s}\|^2 = \theta_{t_s}(2f^{t_s}(x^{t_s} - x^\star) + \theta_{t_s}\|f^{t_s}\|^2), \end{array} \tag{38}$$

where $f^{t_s} \in F_\lambda(x^{t_s})$. Notice that $f^{t_s}(x^{t_s} - x^\star) < -\delta > 0$ and $\|f^{t_s}\|^2 \leq C$. Using these estimates we obtain

$$W(x^{t_s+1}) - W(x^{t_s}) \leq \theta_{t_s}(-2\delta - \theta_{t_s} C) \leq -\theta_{t_s}\delta < 0 \tag{39}$$

for all s large enough. This contradicts (37) and therefore proves the theorem. ■

Conclusions

In this paper we define and use a sharp penalty mapping to construct the iteration algorithm converging to an approximate solutions of monotone variational inequalities. Sharp penalty mappings are analogues of gradient fields of exact penalty functions but do not need to be potential mappings. Three examples of sharp penalty mappings are given with one of them seems to be a new one. The algorithm consists in recursive application of a penalized variational inequality operator, but scaled by step-size multipliers which satisfy classical diverging series condition. As for practical value of these result it is generally believed that the conditions for the step-size multipliers used in this theorem result in rather slow convergence of the order $O(k^{-1})$. However the convergence rate can be improved by different means following the example of non-differentiable optimization. The promising direction is for instance the least-norm adaptive regulation, suggested probably first by A.Fiacco and McCormick [16] as early as 1968 and studied in more details in [17] for convex optimization problems. With some modification in can be easily used for VI problems as well. Experiments show that under favorable conditions it produces step multipliers decreasing as geometrical progression which gives a linear convergence for the algorithm. This may explain the success of [7] where geometrical progression for step multipliers was independently suggested and tested in practice.

Acknowledgements.

This work is supported by the Ministry of Science and Education of Russian Federation, project 1.7658.2017/6.7

References

1. Facchinei, F.; Pang, J.-S.: Finite-dimensional Variational Inequalities and Complementarity Problems. Berlin: Springer-Verlag, (2003)
2. Bauschke, H.; Borwein, J: On Projection Algorithms for Solving Convex Feasibility Problems. SIAM Rev., 38, 367–426 (1996)

3. Nurminski, E.A.: Féjer Processes with Diminishing Disturbances. Doklady Mathematics, 78(2), 755–758 (2008)
4. Nurminski, E.A.: Use of Additional Diminishing Disturbances in Féjer Models of Iterative Algorithms. Computational Mathematics and Mathematical Physics, 48(12), 2154–2161 (2008)
5. Censor, Y.; Davidi, R.; Herman, G.T.: Perturbation Resilience and Superiorization of Iterative Algorithms. Inverse Problems, 26, (2010) DOI:10.1088/0266-5611/26/6/065008
6. Censor, Y.; Davidi,R.; Herman, G.T.; Schulte R. W.; Tetruashvili, L.: Projected Subgradient Minimization Versus Superiorization. Journal of Optimization Theory and Applications, 160, 730–747, (2014)
7. Butnariu, D.; Davidi, R.; Herman G.T.; Kazantsev, I.G.: Stable Convergence Behavior under Summable Perturbations of a Class of Projection Methods for Convex Feasibility and Optimization Problems. IEEE Journal of Selected Topics in Signal Processing, 1, 540–547, (2007) DOI:10.1109/JSTSP.2007.910263
8. Nesterov, Yu.E.: A Method for Solving the Convex Programming Problem with Convergence Rate $O(1/k^2)$. Soviet Math. Doklady, 27, 372-–376 (1983)
9. Nesterov, Yu.E.: Introductory Lectures on Convex Optimization. Boston, MA: Kluwer, (2004)
10. Browder, F.E.: On the Unification of the Calculus of Variations and the Theory of Monotone Nonlinear Operators in Banach Spaces Proc.Nat.Acad.Sci.USA 56, 419–425. (1966)
11. Konnov, I.V.; Pinyagina, O.V.: D-gap Functions and Descent Methods for a Class of Monotone Equilibrium Problems. Lobachevskii J. Math., 13, 57–65 (2003)
12. de André, T., Silva, P.S.: Exact Penalties for Variational Inequalities With Applications to Nonlinear Complementarity Problems. Computational Optimization and Application, 47(3), 401–429 (2010) DOI: 10.1007/s10589-008-9232-3,
13. Kokurin, M., Yu.: An Exact Penalty Method for Monotone Variational Inequalities and Order Optimal Algorithms for Finding Saddle Points. Izv.vyssh.Uchebn.Zaved.Mat., 8, 23–33 (2011)
14. Patriksson, M.: Nonlinear Programming and Variational Inequality Problems: A Unified Approach. Springer Science & Business Media, (2013)
15. Nurminski E.A.: arXiv:1611.09697 [math.OC]
16. Fiacco, A.V.; McCormick, G.P.: Nonlinear Programming: Sequential Unconstrained Minimization Techniques. New York: John Wile & Sons, (1968)
17. Nurminski E.A.: Envelope Step-size Control for Iterative Algorithms Based on Féjer Processes with Attractants. Optimization Methods and Software, 25(1), 97–108 (2010)