

A Simple Adaptive Step-size Choice for Iterative Optimization Methods

I.V. KONNOV¹

Abstract

We suggest a simple adaptive step-size procedure, which does not require any line-search, for a general class of nonlinear optimization methods and prove convergence of a general method under mild assumptions. In particular, the goal function may be non-smooth and non-convex. Unlike the descent line-search methods, it does not require monotone decrease of the goal function values along the iteration points and reduces the implementation cost of each iteration essentially. The key element of this procedure consists in inserting a majorant step-size sequence such that the next element is taken only if the current iterate does not give a sufficient descent. Its applications yield in particular a new gradient projection method for smooth constrained optimization problems and a new projection type method for minimization of the gap function of a general variational inequality. Preliminary results of computational experiments confirm efficiency of the proposed modification.

Key words: Optimization problems; non-smooth functions; iterative methods; adaptive step-size; non-monotone function values; variational inequalities; gap function; convergence properties.

1 Introduction

Iterative methods are utilized for various difficult problems whose solution in the closed form either is not known or has significant computational drawbacks. For instance, if a system of linear equations has large dimensionality and inexact data, it is better to apply one of well known iterative methods. These methods are a standard tool for nonlinear constrained optimization problems; see e.g. [1, 2]. One of the most popular approaches to generation of an iterative sequence consists of solution of direction finding and step-size choice subproblems at each iteration.

During rather long time, most efforts were concentrated on developing more powerful and rapidly convergent methods, such as Newton and interior point type ones, which admit complex transformations at each iteration and attain high accuracy of

¹Department of System Analysis and Information Technologies, Kazan Federal University, ul. Kremlevskaya, 18, Kazan 420008, Russia.

*AMO – Advanced Modeling and Optimization. ISSN: 1841-4311

approximations. That is, the direction finding subproblem was considered as the main one, whereas the step-size was chosen by one of the few well known procedures; see e.g. [2]–[4]. However, new significant areas of applications related to data processing in information and communication systems, having large dimensionality and inexact data together with scattered necessary information force one to avoid complex transformations and apply mostly simple methods such as the projection and linearization type methods, whose iteration computation expenses and accuracy requirements are rather low; see e.g. [5]–[8]. Therefore, one is also interested in suggesting new step-size choice rules that reduce the total computational expenses of the method.

In fact, the existing rules are not completely satisfactory. The exact or approximate one-dimensional minimization line-search requires significant computational expenses per iteration especially in the case where calculation of the function value is almost similar to calculation of its derivative (gradient) and needs solution of complex auxiliary problems; see e.g. [1, 3]. In order to remove the line-search, one can calculate the step-size value via utilization of a priori information such as Lipschitz type constants for the gradient, but then one must take only some their inexact estimates, which leads to slow convergence. This is also the case for the known divergent series rule; see e.g. [1, 6].

In this paper, we suggest a new simple adaptive step-size procedure for a general class of iterative optimization methods, which does not require any line-search. In particular, this procedure can be applied to the known projective optimization methods. In creation of the adaptive step-size rule we follow the approach from [9] where a step-size procedure for the conditional gradient method was proposed. However, the procedure in [9] admits only decrease of the step-size and can not be extended to the other optimization methods since it requires the boundedness of the feasible set and is adjusted to the set-valued solution mapping of the direction finding subproblem. Our new step-size procedure admits different changes of the step-size and wide variety of implementation rules. The key element of this procedure consists in inserting a majorant step-size sequence tending to zero. However, we do not change these majorant values continuously such that we take the next element only if the current iterate has not given a sufficient descent. In such a way, the procedure takes into account behavior of the iteration sequence. We show that this strategy can be implemented within a rather general framework of iterative solution methods applied both for smooth and non-smooth optimization problems and variational inequalities. It does not utilize a priori information such as Lipschitz constants of the gradient, besides, the Lipschitz continuity of the gradient of the goal function is not necessary for convergence of the method. Preliminary results of computational experiments confirm efficiency of the proposed modification.

The remainder of the paper is organized as follows. Section 2 contains necessary definitions and properties from the theory of non-smooth optimization and set-valued analysis. In Section 3, we describe a general framework of iterative methods with the new step-size procedure applied to constrained non-smooth and non-convex optimiza-

tion problems and prove its convergence. Afterwards, we describe its specializations as projection type methods. Namely, a new gradient projection method for smooth constrained optimization problems is given in Section 4, whereas a new projection type method for a variational inequality with a general non-integrable mapping, which is re-formulated as a constrained gap function minimization problem, is given in Section 5. We show that they both fall into the general framework of Section 3 and obtain their convergence directly from the basic convergence theorem of Section 3. Section 6 contains results of preliminary computational experiments.

2 Basic Preliminaries

We intend to develop the new method for a wide class of optimization problems whose goal functions can be non-smooth and non-convex. For this reason, we first recall some concepts and properties from Non-smooth Analysis; see [10] for more details. If some function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is Lipschitz continuous in a neighborhood of any point x of a set X , it is called locally Lipschitz on X . Then we can define its generalized gradient set at x :

$$\partial^\uparrow f(x) = \{g \in \mathbb{R}^n : \langle g, p \rangle \leq f^\uparrow(x; p) \quad \forall p \in \mathbb{R}^n\},$$

which must be non-empty, convex and closed. Here $f^\uparrow(x; p)$ denotes the upper Clarke derivative:

$$f^\uparrow(x; p) = \limsup_{y \rightarrow x, \alpha \searrow 0} ((f(y + \alpha p) - f(y))/\alpha).$$

It follows that

$$f^\uparrow(x, p) = \sup_{g \in \partial^\uparrow f(x)} \langle g, p \rangle.$$

In general, the locally Lipschitz function f need not be differentiable. At the same time, it has the gradient $\nabla f(x)$ a.e. in X , furthermore, it holds that

$$\partial^\uparrow f(x) = \text{conv} \left\{ \lim_{y \rightarrow x} \nabla f(y) : y \in X_f, y \notin S \right\}, \quad (1)$$

where X_f denotes the set of points where f is differentiable, and S denotes an arbitrary subset of measure zero. If f is convex, then $\partial^\uparrow f(x)$ coincides with the subdifferential $\partial f(x)$ in the sense of Convex Analysis, i.e.,

$$\partial f(x) = \{g \in \mathbb{R}^n : f(y) - f(x) \geq \langle g, y - x \rangle \quad \forall y \in \mathbb{R}^n\}.$$

In this case the upper derivative coincides with the usual directional derivative:

$$f^\uparrow(x; p) = f'(x; p). \quad (2)$$

Also, if f is differentiable at x , (2) obviously holds and we have

$$f'(x; p) = \langle \nabla f(x), p \rangle \text{ and } \partial^\uparrow f(x) = \{\nabla f(x)\};$$

cf. (1).

The general optimization problem consists in finding the minimal value of some goal function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ on a feasible set $D \subseteq \mathbb{R}^n$. For brevity, we write this problem as

$$\min_{x \in D} f(x), \tag{3}$$

its solution set is denoted by D^* and the optimal value of the function by f^* , i.e. $f^* = \inf_{x \in D} f(x)$. We will use the following first set of basic assumptions for problem (3).

(A1) *The set D is nonempty, convex, and closed, the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is locally Lipschitz on D .*

(A2) *There exists a number $\gamma > f^*$ such that the set*

$$D_\gamma = \{x \in D : f(x) \leq \gamma\}$$

is bounded.

Clearly, **(A2)** is a general coercivity condition that is necessary in the case where the set D is unbounded. If **(A1)** and **(A2)** hold, problem (3) has a solution. This means that we intend to present a method for non-smooth and non-convex optimization problems.

Together with problem (3) we will consider the following set-valued variational inequality (VI for short): Find a point $x^* \in D$ such that

$$\exists g^* \in \partial^\uparrow f(x^*), \quad \langle g^*, x - x^* \rangle \geq 0 \quad \forall x \in D. \tag{4}$$

We denote by D^0 the solution set of VI (4). Solutions of VI (4) are called *stationary points* of (3) due to the known necessary optimality condition; see e.g. [10, 11].

Proposition 2.1 Let **(A1)** hold. Then each solution of problem (3) is a solution of VI (4).

The reverse implication needs additional conditions. We recall that a function $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ is called

(a) *pseudo-convex* on a set X , if for each pair of points $x, y \in X$, we have

$$\varphi'(x; y - x) \geq 0 \implies \varphi(y) \geq \varphi(x);$$

(b) *semi-convex* (or *upper pseudo-convex*) if for each pair of points $x, y \in X$, we have

$$\varphi^\uparrow(x; y - x) \geq 0 \implies \varphi(y) \geq \varphi(x);$$

see [12] and also [11]. In case (2), these concepts coincide, but in general (b) implies (a). Besides, the class of convex functions is strictly contained in that of pseudo-convex functions. If **(A1)** holds and f is semi-convex, then each solution of VI (4) clearly solves problem (3), i.e. $D^* = D^0$.

We need several continuity properties of set-valued mappings; see e.g. [13, 11]. Here and below $\Pi(A)$ denotes the family of all nonempty subsets of a set A .

Let X be a convex set in \mathbb{R}^n . A set-valued mapping $Q : X \rightarrow \Pi(\mathbb{R}^n)$ is said to be

(a) *upper semicontinuous (u.s.c.)*, if for each point $y \in X$ and for each open set U such that $Q(y) \subset U$, there is a neighborhood Y of y such that $Q(z) \subset U$ whenever $z \in X \cap Y$;

(b) *closed*, if for each pair of sequences $\{x^k\} \rightarrow x$, $\{q^k\} \rightarrow q$ such that $x^k \in X$ and $q^k \in Q(x^k)$, we have $q \in Q(x)$;

(c) a *K-mapping* (Kakutani-mapping), if it is u.s.c. and has nonempty, convex, and compact values.

It is known (see e.g. [13, Chapter 1, Lemma 4.4]), that each u.s.c. mapping with closed values is closed and that each closed mapping which maps any compact set into a compact set is u.s.c. Also, if a function $f : Y \rightarrow \mathbb{R}$ is locally Lipschitz on an open convex set Y , then $\partial^\dagger f$ is a K-mapping on Y ; see [10, Section 2.1].

We shall also use the mean value theorem by G. Lebourg for locally Lipschitz functions.

Proposition 2.2 [10, Theorem 2.3.7] Let x and y be given points in \mathbb{R}^n and let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a Lipschitz continuous function on an open set containing the segment $[x, y]$. Then there exists a point $z \in (x, y)$ such that

$$f(y) - f(x) \in \langle \partial^\dagger f(z), y - x \rangle.$$

3 The Basic Method and Its Convergence

We first describe conditions for the solution mapping of the direction finding subproblem.

(A3) *There exists a single-valued mapping $x \mapsto y(x)$, which maps the set D into D such that*

- (i) *it is continuous;*
- (ii) *$\bar{x} = y(\bar{x})$ if and only if \bar{x} is a solution of VI (4);*
- (iii) *for each $x \in D$ and for all $g \in \partial^\dagger f(x)$ it holds that*

$$\langle g, y(x) - x \rangle \leq -\tau \|y(x) - x\|^2 \tag{5}$$

for some $\tau > 0$.

We now describe the basic method for problem (3), which involves a simple adaptive step-size procedure without line-search.

Method (SBM).

Step 0: Choose a point $x^0 \in D_\gamma$, a number $\beta \in (0, 1)$ and a sequence $\{\tau_l\} \rightarrow 0$, $\tau_l \in (0, 1)$. Set $k = 0$, $l = 0$, $u^0 = x^0$, choose a number $\lambda_0 \in (0, \tau_0]$.

Step 1: Take the point $y^k = y(x^k)$. If $y^k = x^k$, stop. Otherwise set $d^k = y^k - x^k$ and $z^{k+1} = x^k + \lambda_k d^k$.

Step 2: If

$$f(z^{k+1}) \leq f(x^k) - \beta \lambda_k \|d^k\|^2, \quad (6)$$

take $\lambda_{k+1} \in [\lambda_k, \tau_l]$, set $x^{k+1} = z^{k+1}$ and go to Step 4.

Step 3: Set $\lambda'_{k+1} = \min\{\lambda_k, \tau_{l+1}\}$, $l = l+1$ and take $\lambda_{k+1} \in (0, \lambda'_{k+1}]$. If $f(z^{k+1}) \leq \gamma$, set $x^{k+1} = z^{k+1}$ and go to Step 4. Otherwise set $x^{k+1} = u^k$, $u^{k+1} = u^k$, $k = k+1$ and go to Step 1.

Step 4: If $f(x^{k+1}) < f(u^k)$, set $u^{k+1} = x^{k+1}$, $k = k+1$ and go to Step 1.

Therefore, (SBM) in fact represents a general framework of iterative methods for nonlinear optimization problems. Observe that the sequence $\{u^k\}$ simply contains the best current points of the sequence $\{x^k\}$, i.e.

$$f(u^k) = \min_{0 \leq i \leq k} f(x^i).$$

Due to **(A3)**, termination of (SBM) yields a point of D^0 . Hence, we will consider only the case where the sequence $\{x^k\}$ is infinite.

Theorem 3.1 Let the assumptions **(A1)**–**(A3)** be fulfilled and $\beta < \tau$. Then:

(i) The sequence $\{x^k\}$ has a limit point, which belongs to the set D^0 .

(ii) If $D^* = D^0$, then all the limit points of the sequence $\{x^k\}$ belong to the set D^* , besides, we have

$$\lim_{k \rightarrow \infty} f(x^k) = f^*. \quad (7)$$

Proof. First we observe that the sequence $\{x^k\}$ belongs to the bounded set D_γ and must have limit points. By **(A3)**, so is the sequence $\{y^k\}$. Hence $\{d^k\}$ is also bounded and has limit points. Next, we take the subsequence of indices $\{i_s\}$ such that

$$f(z^{i_s+1}) > \gamma, \quad f(x^{i_s}) \leq \gamma, \quad (8)$$

$$f(z^{i_s+1}) > f(x^{i_s}) - \beta \lambda_{i_s} \|d^{i_s}\|^2, \quad z^{i_s+1} = x^{i_s} + \lambda_{i_s} d^{i_s}. \quad (9)$$

Let us consider several possible cases.

Case 1: The subsequence $\{x^{i_s}\}$ is infinite.

Take an arbitrary limit point x' of the subsequence $\{x^{i_s}\}$. Without loss of generality we can suppose that

$$\lim_{s \rightarrow \infty} x^{i_s} = x' \quad \text{and} \quad \lim_{s \rightarrow \infty} y^{i_s} = y',$$

where $y' = y(x')$ by **(A3)**. Note that

$$\lambda_{i_s} \in (0, \tau_{l_s}], \quad \lambda_{i_s+1} \in (0, \tau_{l_s+1}],$$

for some infinite subsequence of indices $\{l_s\}$ where

$$\lim_{s \rightarrow \infty} \tau_{l_s} = 0.$$

Since the sequence $\{d^{i_s}\}$ is bounded, the limit points of the subsequences $\{x^{i_s}\}$ and $\{z^{i_s+1}\}$ coincide due to (9). From (8) we now obtain

$$f(x') = \gamma > f^*. \quad (10)$$

Applying Proposition 2.2 in (9), we have

$$\langle g^{i_s}, d^{i_s} \rangle \geq -\beta \|d^{i_s}\|^2$$

for some $g^{i_s} \in \partial^\dagger f(x^{i_s} + \theta_{i_s} \lambda_{i_s} d^{i_s})$, and $\theta_{i_s} \in (0, 1)$. Taking the limit $s \rightarrow \infty$ gives

$$\langle g', y' - x' \rangle \geq -\beta \|y' - x'\|^2$$

for some $g' \in \partial^\dagger f(x')$. Using (5), we obtain

$$\beta \|y' - x'\|^2 \geq \tau \|y' - x'\|^2,$$

i.e. $y(x') = x'$, hence

$$x' \in D^0. \quad (11)$$

Therefore, assertion (i) is true in this case.

Case 2: The subsequence $\{x^{i_s}\}$ is finite.

Without loss of generality we can then suppose that $z^k = x^k$ for each number k . The further proof depends on the properties of the sequence $\{\lambda_k\}$.

Case 2a: The number of changes of the index l is finite.

Then we have $\lambda_k \geq \bar{\lambda} > 0$ for k large enough, hence (6) gives

$$f(x^{k+1}) \leq f(x^k) - \beta \lambda_k \|d^k\|^2 \leq f(x^k) - \beta \bar{\lambda} \|d^k\|^2$$

for k large enough. Since $f(x^k) \geq f^* > -\infty$, we must have

$$\lim_{k \rightarrow \infty} f(x^k) = \mu \quad (12)$$

and

$$\lim_{k \rightarrow \infty} \|y^k - x^k\| = 0. \quad (13)$$

Let x'' be an arbitrary limit point of the sequence $\{x^k\}$. From (13) and **(A3)** we now have

$$y(x'') = x'',$$

which gives $x'' \in D^0$. Hence in this case all the limit points of the sequence $\{x^k\}$ belong to the set D^0 . Therefore, assertion (i) is true in this case.

Case 2b: The number of changes of the index l is infinite.

Then there exists an infinite subsequence of indices $\{k_l\}$ such that

$$f(x^{k_l} + \lambda_{k_l} d^{k_l}) - f(x^{k_l}) = f(x^{k_l+1}) - f(x^{k_l}) > -\beta \lambda_{k_l} \|d^{k_l}\|^2. \quad (14)$$

besides,

$$\lambda_{k_l} \in (0, \tau_l], \quad \lambda_{k_{l+1}} \in (0, \tau_{l+1}],$$

and

$$\lim_{l \rightarrow \infty} \tau_l = 0.$$

Let \bar{x} be an arbitrary limit point of this subsequence $\{x^{k_l}\}$. Without loss of generality we can suppose that

$$\lim_{l \rightarrow \infty} x^{k_l} = \bar{x} \quad \text{and} \quad \lim_{l \rightarrow \infty} y^{k_l} = \bar{y}.$$

where $\bar{y} = y(\bar{x})$. Applying Proposition 2.2 in (14), we have

$$\langle g^{k_l}, d^{k_l} \rangle \geq -\beta \|d^{k_l}\|^2$$

for some $g^{k_l} \in \partial^\dagger f(x^{k_l} + \theta_{k_l} \lambda_{k_l} d^{k_l})$, and $\theta_{k_l} \in (0, 1)$. Since $\lambda_{k_l} \rightarrow 0$ as $l \rightarrow \infty$, taking the limit $l \rightarrow +\infty$ gives

$$\langle \bar{g}, \bar{y} - \bar{x} \rangle \geq -\beta \|\bar{y} - \bar{x}\|^2$$

for some $\bar{g} \in \partial^\dagger f(\bar{x})$. Using (5), we obtain

$$\beta \|\bar{y} - \bar{x}\|^2 \geq \tau \|\bar{y} - \bar{x}\|^2,$$

i.e. $\bar{x} = y(\bar{x})$, hence $\bar{x} \in D^0$. Therefore, all the limit points of the subsequence $\{x^{k_l}\}$ belong to the set D^0 . Since $x^{k_{l+1}} = x^{k_l} + \lambda_{k_l} d^{k_l}$, $\lambda_{k_l} \rightarrow 0$, and the sequence $\{d^{k_l}\}$ is bounded, the limit points of the subsequences $\{x^{k_l}\}$ and $\{x^{k_{l+1}}\}$ coincide and all they belong to the set D^0 . We conclude that assertion (i) is also true in this case.

We now suppose in addition that $D^0 = D^*$. Then relations (10) and (11) become inconsistent, hence Case 1 is impossible. This means that the subsequence $\{x^{i_s}\}$ is always finite. In Case 2a we now have $\mu = f^*$ in (12), which gives (7). We conclude that assertion (ii) holds true in this case.

In Case 2b, the limit points of the subsequences $\{x^{k_l}\}$ and $\{x^{k_{l+1}}\}$ coincide and all they now belong to the set D^* . For any index k we define the index $m(k)$ as follows:

$$m(k) = \max\{j : j \leq k, f(x^j) - f(x^{j-1}) > -\beta \lambda_{j-1} \|d^{j-1}\|^2\},$$

i. e. $m(k)$ is the closest to k but not greater index from the subsequence $\{x^{k_l+1}\}$. This means that $m(k) = k$ if $f(x^k) - f(x^{k-1}) > -\beta \lambda_{k-1} \|d^{k-1}\|^2$. By definition, we have

$$f(x^k) \leq f(x^{m(k)}). \quad (15)$$

Let now x^* be an arbitrary limit point of the sequence $\{x^k\}$, i.e. $\lim_{s \rightarrow \infty} x^{t_s} = x^*$. Create the corresponding infinite subsequence $\{x^{m(t_s)}\}$. From (15) we have $f^* \leq f(x^{t_s}) \leq f(x^{m(t_s)})$, but all the limit points of the sequence $\{x^{m(t_s)}\}$ belong to the set D^* since it is contained in the sequence $\{x^{k_l+1}\}$. Choose any limit point \tilde{x} of $\{x^{m(t_s)}\}$. Then, taking a subsequence if necessary we obtain

$$f^* \leq f(x^*) \leq f(\tilde{x}) = f^*.$$

therefore $x^* \in D^*$. This means that all the limit points of the sequence $\{x^k\}$ belong to the set D^* and that (7) holds true. We conclude that assertion (ii) is also true. \square

The method can be simplified in the case where the set D is bounded. Then we can set $\gamma = +\infty$ and remove all the calculations of the sequence $\{u^k\}$. It is easy to verify that all the assertions of Theorem 3.1 remain true.

4 Application to Smooth Optimization Problems

From the results of Section 3 it follows that we can create a number of new solution methods for optimization problems. It suffices to take an optimization problem that satisfies conditions **(A1)** and **(A2)** and a method whose solution mapping of the direction finding subproblem satisfies condition **(A3)**. Then we place this method in the framework of (SBM) and obtain its convergence properties directly from Theorem 3.1.

We illustrate diversity of possible specializations of (SBM) by only two basic examples. In this section, we take the well known class of smooth constrained optimization problems.

(A1') *The set D is nonempty, convex, and closed, the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuously differentiable on D .*

Clearly, **(A1')** implies **(A1)**. Let $\pi_X(x)$ denotes the projection of x onto a set X . Fix a number $\alpha > 0$ and define the mapping $y_\alpha(x) = \pi_D[x - \alpha^{-1}\nabla f(x)]$ on the set D . Then setting $y(x) = y_\alpha(x)$ in (SBM), we obtain a new version of the gradient projection method for problem (3). We call it (GPMS) for brevity.

We now utilize the well known properties of mapping $x \mapsto y_\alpha(x)$; see e.g. [14] and [11, Lemma 9.5].

Lemma 4.1 Let the assumptions in **(A1')** be fulfilled. Then:

- (a) $\bar{x} = y_\alpha(\bar{x})$ if and only if $\bar{x} \in D^0$;
- (b) The mapping $x \mapsto y_\alpha(x)$ is continuous on D ;
- (c) For any point $x \in D$ it holds that

$$\langle \nabla f(x), y_\alpha(x) - x \rangle \leq -\alpha \|y_\alpha(x) - x\|^2.$$

Therefore, the assumptions in **(A3)** are fulfilled with $\tau = \alpha$ and we can obtain the convergence result for (GPMS) directly from Theorem 3.1.

Theorem 4.1 Let assumptions **(A1')** and **(A2)** be fulfilled. If we apply (GPMS) with $\beta < \alpha$, then the following assertions are true:

- (i) The termination of (GPMS) yields a point of D^0 .
- (ii) The sequence $\{x^k\}$ has a limit point, which belongs to the set D^0 .
- (iii) If $D^* = D^0$, then all the limit points of the sequence $\{x^k\}$ belong to the set D^* and (7) holds.

Observe that the choice $\alpha \geq 1$ allows us to take an arbitrary value $\beta \in (0, 1)$ for convergence. The above method can be extended to the case where

$$f(x) = \mu(x) + \eta(x),$$

where μ is continuously differentiable and η is convex, but non-differentiable. We have to replace the projection mapping with the proximal mapping with respect to η and apply the so-called splitting or proximal gradient iteration. A number of these splitting based descent methods were proposed for such composite non-smooth optimization problems; see e.g. [14, 4, 6, 11, 7]. Similarly, we can create new versions of splitting methods if we place the corresponding splitting direction finding mapping in the (SBM) framework.

5 Application to Non-smooth Variational Inequality Problems

In this section, we take a variational inequality whose underlying mapping is (strongly) monotone, but non-integrable and non-smooth in general. Given a convex set D in \mathbb{R}^n and a single-valued mapping $G : D \rightarrow \mathbb{R}^n$, one can define the custom *variational inequality* (VI for short): Find $x^* \in D$ such that

$$\langle G(x^*), x - x^* \rangle \geq 0 \quad \forall x \in D. \quad (16)$$

This problem has a great number of applications in different fields, the theory and methods for VIs are investigated very extensively and great advances were made in this field; e.g. see [4, 15, 16] and the references therein. We denote by D^e the solution set of VI (16) and consider this problem under the following basic assumptions.

(A1'') D is a nonempty, closed, and convex set in \mathbb{R}^n , the mapping $G : Y \rightarrow \mathbb{R}^n$ satisfies the Lipschitz condition in a neighborhood of each point of an open convex set Y such that $D \subset Y$.

Fix a number $\alpha > 0$ and define the usual gap function

$$\varphi_\alpha(x) = \max_{y \in D} \{ \langle G(x), x - y \rangle - 0.5\alpha \|x - y\|^2 \}; \quad (17)$$

e.g. see [4]. Then there exists a unique element $y_\alpha(x) \in D$ such that

$$\varphi_\alpha(x) = \langle G(x), x - y_\alpha(x) \rangle - 0.5\alpha \|x - y_\alpha(x)\|^2,$$

moreover, $y_\alpha(x) = \pi_D[x - \alpha^{-1}G(x)]$. Thus, we again have a single-valued mapping $x \mapsto y_\alpha(x)$ on D . Then, setting $y(x) = y_\alpha(x)$ and $f = \varphi_\alpha$ in (SBM), we obtain a new version of the projective gap function method for problem (16). We call it (GFPMS) for brevity. We observe that the descent method with Armijo line-search was proposed for this non-smooth VI in [17].

First of all we replace VI (16) with the optimization problem

$$\min_{x \in D} \rightarrow \varphi_\alpha(x). \quad (18)$$

From the definition of the function φ_α in (17) we can easily deduce that it is always nonnegative and that the optimal value in (18) is zero. We shall utilize the other known properties of the mapping $x \mapsto y_\alpha(x)$; see [17, 18].

Lemma 5.1 Let the assumptions in **(A1'')** be fulfilled. Then:

- (a) VI (16) is equivalent to problem (18);
- (b) $\bar{x} = y_\alpha(\bar{x})$ if and only if $\bar{x} \in D^e$;
- (c) The mapping $x \mapsto y_\alpha(x)$ is continuous on D .

Therefore, the optimization problem (18) is an equivalent re-formulation of VI (16). Although the gap function φ_α is non-smooth and non-convex, it is locally Lipschitz on D and we can calculate its generalized gradient set; see [17, Lemma 4].

Lemma 5.2 Let the assumptions in **(A1'')** be fulfilled. Then, at any point $x \in D$, there exists the generalized gradient set

$$\partial^\dagger \varphi_\alpha(x) = G(x) - [\partial^\dagger G(x)^\top - \alpha I] (y_\alpha(x) - x),$$

where $\partial^\dagger G(x)$ denotes the generalized Jacobian of G at x .

We recall that a mapping $G : X \rightarrow \mathbb{R}^n$ is said to be

- (a) *monotone* if, for each pair of points $x, y \in X$, it holds that

$$\langle G(x) - G(y), x - y \rangle \geq 0;$$

- (b) *strongly monotone* with constant $\tau > 0$ if, for each pair of points $x, y \in X$, it holds that

$$\langle G(x) - G(y), x - y \rangle \geq \tau \|x - y\|^2.$$

Let us take the strong monotonicity assumption on the mapping G .

(A4) *The mapping $G : D \rightarrow \mathbb{R}^n$ is strongly monotone with constant $\tau > 0$.*

The strong monotonicity enables us to obtain the desired coercivity and stationarity properties; see [17, Lemmas 5 and 6].

Lemma 5.3 Let the assumptions in **(A1'')** and **(A4)** be fulfilled. Then:

- (a) VI (16) has a unique solution;
- (b) There exists a number $\sigma > 0$ such that

$$\varphi_\alpha(x) \geq \sigma \|x - x^*\|^2 \quad \forall x \in D,$$

where x^* is a unique solution to VI (16);

- (c) For each point $x \in D$ and for all elements $V \in \partial^\dagger G(x)$ it holds that

$$\langle G(x) - (V^\top - \alpha I)(y_\alpha(x) - x), y_\alpha(x) - x \rangle \leq -\tau \|y_\alpha(x) - x\|^2.$$

Therefore, the assumptions in **(A2)** and **(A3)** are fulfilled with $f = \varphi_\alpha$, besides, $D^* = D^0 = D^e$, and we can obtain the convergence result for (GFPMS) directly from Theorem 3.1.

Theorem 5.1 Let assumptions **(A1'')** and **(A4)** be fulfilled. If we apply (GFPMS) with $\beta < \tau$, then the following assertions are true:

- (i) The termination of (GFPMS) yields a unique solution to VI (16).
- (ii) The sequence $\{x^k\}$ converges to a unique solution to VI (16).

There exist various gap function based methods with line-search procedures for different classes of VIs; see e.g. [4, 16, 18, 11]. Again, we can create new versions of these methods after the proper mapping substitution in (SBM).

6 Computational Experiments

In order to check the performance of the proposed methods we carried out computational experiments. The main goal was to compare them with the methods having the same direction mapping but utilizing the Armijo line-search. For more clarity, we describe now the corresponding modification of (SBM).

Method (ABM).

Step 0: Choose a point $x^0 \in D_\gamma$, numbers $\beta \in (0, 1)$ and $\theta \in (0, 1)$. Set $k = 0$.

Step 1: Take a point $y^k = y(x^k)$. If $y^k = x^k$, stop. Otherwise set $d^k = y^k - x^k$.

Step 2: Determine m as the smallest nonnegative integer such that

$$f(x^k + \theta^m d^k) \leq f(x^k) - \beta \theta^m \|d^k\|^2,$$

set $\lambda_k = \theta^m$, $x^{k+1} = x^k + \lambda_k d^k$, $k = k + 1$ and go to Step 1.

Various implementations of this method were investigated in many works; e.g. see [14, 4, 18, 11] and the references therein. Its convergence properties are similar to those of the other known descent methods with line-search and it does not require a priori information. The presence of this line-search at each iteration is the main difference from (SBM). Also, we took for comparison the known non-monotone method with the divergent series step-size rule, which does not require line-search or a priori information; see e.g. [1, Chapters V and VII].

We compared all the methods for different dimensionality. They were implemented in Delphi with double precision arithmetic. Namely, we indicate the number of iterations (it) and the total number of calculations of the goal function value (kf) for attaining the same accuracy $\varepsilon = 0.01$ with respect to the error function

$$\Delta(x) = \|x - \pi_D[x - y(x)]\|.$$

We took $\theta = 0.5$ for (ABM). For (SBM), we simply set $\lambda_{k+1} = \lambda_k$ if (6) holds, and $\lambda_{k+1} = \sigma \lambda_k$ with $\sigma = 0.9$ otherwise.

Table 1: Convex optimization: Test 1 (*it* is the number of iterations, *kf* is the number of function calculations)

		(GPMA)		(GPMS)		(GPMD)	
<i>m</i>	<i>n</i>	<i>it</i>	<i>kf</i>	<i>it</i>	<i>kf</i>	<i>it</i>	<i>kf</i>
2	5	4	14	12	21	17	18
4	5	15	57	30	35	40	41
5	10	18	76	28	47	-	-
25	50	344	2683	637	679	-	-
50	100	1229	12025	2633	2689	-	-

First we applied the methods to smooth convex optimization problems of form (3). More precisely, we chose

$$f(x) = 0.5\|Px - q\|^2, \tag{19}$$

the elements of the $m \times n$ matrix P were defined by

$$p_{ij} = \begin{cases} \sin(i) \cos(j) & \text{if } i \neq j, \\ \sin(i) \cos(j) + 2 & \text{if } i = j; \end{cases}$$

and

$$q_i = \sum_{j=1}^n p_{ij}, \quad i = 1, \dots, m.$$

We utilized the mapping $y_\alpha(x) = \pi_D[x - \alpha^{-1}\nabla f(x)]$ with $\alpha = 1$ and (GPMS). Analogously, setting $y(x) = y_1(x)$ in (ABM), we obtain the well known gradient projection method with Armijo line-search. We call it (GPMA) for brevity. We set $\beta = 0.5$ for both the methods. We also implemented the gradient projection method with the divergent series step-size rule:

$$x^{k+1} = \pi_D[x^k - \lambda_k \nabla f(x^k)], \quad \lambda_k = 1/(k+1), \quad k = 0, 1, \dots;$$

see e.g. [1, Section 7.2.2]. We call it (GPMD) for brevity.

In the first series, we took the feasible set $D = \mathbb{R}_+^n$ where

$$\mathbb{R}_+^n = \{x \in \mathbb{R}^n : x_j \geq 0, \quad j = 1, \dots, n\}$$

and the starting point $x_j^0 = n/2 + \sin(j)$ for $j = 1, \dots, n$. The results are given in Table 1. (GPMD) showed very slow convergence when $n > 5$. For the case where $m = 5$ and $n = 10$, it attained only the accuracy 0.108 in 5000 iterations. For this reason, we made the further comparison only for (GPMA) and (GPMS).

In the second series, we took the same cost function from (19), the feasible set

$$D = \{x \in \mathbb{R}^n : -5 \leq x_j \leq 5, \quad j = 1, \dots, n\},$$

Table 2: Convex optimization: Test 2 (*it* is the number of iterations, *kf* is the number of function calculations)

		(GPMA)		(GPMS)	
<i>m</i>	<i>n</i>	<i>it</i>	<i>kf</i>	<i>it</i>	<i>kf</i>
2	5	4	24	14	21
4	5	17	65	35	38
5	10	19	80	60	66
25	50	225	1778	440	463
50	100	748	7445	1624	1660

and the starting point $x_j^0 = -5$ for $j = 1, \dots, n$. The results are given in Table 2.

Next we applied the methods to variational inequality problems. We chose the nonlinear strongly monotone mapping in VI (16) as follows:

$$G(x) = Ax + b + \mu C(x), \quad A = A' + A'',$$

the elements of the $n \times n$ matrix A' were defined by

$$a'_{ij} = \begin{cases} \sin(i) \cos(j)/(i+j) & \text{if } i < j, \\ \sin(j) \cos(i)/(i+j) & \text{if } i > j, \\ \sum_{s \neq i} |a'_{is}| + 2 & \text{if } i = j; \end{cases}$$

the elements of the $n \times n$ matrix A'' were defined by

$$a''_{ij} = \begin{cases} \sin(ij) \ln(1 + i/j) & \text{if } i < j, \\ -\sin(ij) \ln(1 + j/i) & \text{if } i > j, \\ 0 & \text{if } i = j; \end{cases}$$

and

$$b_i = -10 \sum_{j=1}^n a_{ij}, \quad i = 1, \dots, n.$$

This means that the matrix A is positive definite and asymmetric. The parameter μ was set to be 10, the mapping $C(x)$ was chosen to be diagonal with the elements

$$C_i(x) = \arctan(x_i - 2), \quad i = 1, \dots, n.$$

We utilized the mapping $y_\alpha(x) = \pi_D[x - \alpha^{-1}G(x)]$ with $\alpha = 1$ and (GFPMS). Analogously, setting $y(x) = y_1(x)$ in (ABM), we obtain the well known descent projection method with Armijo line-search. We call it (GFPMA) for brevity. We set $\beta = 0.4$ for both the methods.

Table 3: Variational inequality: Test 3 (*it* is the number of iterations, *kf* is the number of function calculations)

<i>n</i>	(GFPMA)		(GFPMS)	
	<i>it</i>	<i>kf</i>	<i>it</i>	<i>kf</i>
5	4	14	20	26
10	8	23	21	27
20	14	48	40	45
50	47	161	48	53
100	85	320	92	97
200	148	660	145	150
500	375	2143	345	351
1000	761	5076	708	716

In the third series, we took the feasible set

$$D = \{x \in \mathbb{R}^n : 1 \leq x_j \leq 6, j = 1, \dots, n\},$$

and starting point $x_j^0 = 6$ for $j = 1, \dots, n$. The results are given in Table 3.

In almost all the cases, the implementations of (SBM), which do not use line-search, showed rather rapid convergence, they outperformed the implementations of (ABM) in the total number of goal function calculations.

7 Conclusions

We suggested a new simple adaptive step-size procedure in a general class of solution methods for optimization problems, whose goal function may be non-smooth and non-convex. This procedure does not require any line-search or a priori information, but takes into account behavior of the iteration sequence. Therefore, it reduces the implementation cost of each iteration essentially in comparison with the descent line-search methods. We established convergence of the method under mild assumptions involving the usual coercivity condition. We showed that this new procedure yields in fact a general framework for optimization methods. In particular, a new gradient projection method for smooth constrained optimization problems and a new projection type method for minimization of the gap function of a general variational inequality can be obtained within this framework. The preliminary results of computational tests showed efficiency of the new procedure.

Acknowledgement

The results of this work were obtained within the state assignment of the Ministry of Science and Education of Russia, project No. 1.460.2016/1.4. In this work, the author was also supported by Russian Foundation for Basic Research, project No. 16-01-00109a.

References

- [1] Polyak, B.T.: Introduction to Optimization. Nauka, Moscow (1983) [Engl. transl. in Optimization Software, New York (1987)]
- [2] Bertsekas, D.P.: Nonlinear Programming. Athena Scientific, Belmont (1999)
- [3] Boyd, S.P., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2009)
- [4] Patriksson, M.: Nonlinear Programming and Variational Inequality Problems: A Unified Approach. Kluwer, Dordrecht (1999)
- [5] Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Mining Know. Disc. 2, 121–167 (1998)
- [6] Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imag. Sci. 2, 183–202 (2009)
- [7] Cevher, V., Becker, S., Schmidt, M.: Convex optimization for big data. Signal Process. Magaz. 31, 32–43 (2014)
- [8] Scutari, G., Facchinei, F., Song, P., Palomar, D.P., Pang, J.-S.: Decomposition by partial linearization: parallel optimization of multi-agent systems. IEEE Trans. Signal Process. 62, 641–656 (2014)
- [9] Konnov, I.V.: Conditional gradient method without line-search. Russ. Mathem. (Izv. VUZ). 62, 82–85 (2018)
- [10] Clarke, F.H.: Optimization and Nonsmooth Analysis. John Wiley and Sons, New York (1983)
- [11] Konnov, I.V.: Nonlinear Optimization and Variational Inequalities. Kazan Univ. Press, Kazan (2013) [In Russian]
- [12] Mifflin, R.: Semismooth and semiconvex functions in constrained optimization. SIAM J. Contr. Optim. 15, 959–972 (1977)

- [13] Nikaido, H.: *Convex Structures and Economic Theory*. Academic Press, New York (1968)
- [14] Patriksson, M.: Cost approximation: a unified framework of descent algorithms for nonlinear programs. *SIAM J. Optim.* 8, 561–582 (1998)
- [15] Konnov, I.V.: *Combined Relaxation Methods for Variational Inequalities*. Springer-Verlag, Berlin (2001)
- [16] Facchinei, F., Pang, J.-S.: *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer-Verlag, Berlin (2003)
- [17] Konnov, I.V.: Descent method for nonsmooth variational inequalities. *Comp. Maths and Math. Phys.* 46, 1286–1292 (2006)
- [18] Konnov, I.V.: Descent methods for mixed variational inequalities with non-smooth mappings. In: Reich, S. and Zaslavski, A.J. (eds.) *Optimization Theory and Related Topics*. Contemporary Mathematics, 568. Amer. Math. Soc., Providence, 121–138 (2012)