ARTICLE TEMPLATE

Subgradient method with feasible inexact projections for constrained convex optimization problems

A. A. Aguiar[a] and O. P. Ferreira[a] and L. F. Prudente[a]

[a]IME, Universidade Federal de Goiás, Goiânia, GO, 74001-970, Brazil

ARTICLE HISTORY
Compiled March 8, 2021

ABSTRACT
In this paper, we propose a new inexact version of the projected subgradient method to solve nondifferentiable constrained convex optimization problems. The method combine ε-subgradient method with a procedure to obtain a feasible inexact projection onto the constraint set. Asymptotic convergence results and iteration-complexity bounds for the sequence generated by the method employing the well known exogenous stepsizes, Polyak's stepsizes, and dynamic stepsizes are established.

KEYWORDS
Subgradient method; feasible inexact projection; constrained convex optimization

1. Introduction

The Subgradient method is one of the most interesting iterative method for solving nondifferentiable convex optimization problems, which has its origin and development in the 60's, see [1,2]. Since then, the subgradient method has attracted the attention of the scientific community working on optimization. One of the factors that explains this interest is its simplicity and ease of implementation. In particular, allowing a low cost of storage and ready exploitation of separability and sparsity. For these reasons, several variants of this method have emerged and properties of it have been discovered throughout the years, resulting in a wide literature on the subject; see, for exemple [3–8] and the references therein.

The aim of this paper is to present an inexact version of the projected subgradient method, which consists in using an inexact projection instead of the exact one, for minimizing a convex function $f : \mathbb{R}^n \to \mathbb{R}$ onto a closed and convex subset C of $\mathbb{R}^n$. The proposed method, that we call *Subgradient-InexP method*, generates a sequence $\{x^k\}$ where each iteration consists of two stages. The first stage performs a step from the current iterate x^k in the opposite direction of a ε-subgradient of f at x^k and the second inexactly projects the resulting vector onto the feasible set C. From the theoretical point of view, considering methods that use inexact projections are particularly interesting for the following reasons. Even when the projection onto a convex set is an easy problem, iterative methods provide only approximated solutions with small errors, due to round-off errors in floating-point arithmetics. Therefore, the study of inexact methods gives theoretical support for real computational implementations of exact schemes. On the other hand, in general, one drawback of methods that use exact projections is having to solve a quadratic problem at each stage, which may substantially increasing the cost

A. A. Aguiar. Email: ademiraguia@discente.ufg.br

per iteration if the number of unknowns is large. In fact, it may not be justified to compute exact projection when the current iterate x^k is far from the solution of the problem in consideration. Moreover, a procedure for computing a feasible inexact projection may present a low computation cost per iteration in comparison with one that computes the exact projection. Thus, it seems reasonable to consider versions of projected subgradient method that compute the projection only approximately. In order to present formally and analyze the Subgradient-InexP method, we use the concept of feasible inexact projection with relative error, which was appeared in [9] (see also[10]). It is worth noting that the concept of feasible inexact projection also accepts an exact projection when it is easy to obtain. For instance, the exact projections onto a box or a second order cone is very easy to obtain; see, respectively, [11, p. 520] and [12, Proposition 3.3]. A feasible inexact projection onto a polyhedral closed convex set can be obtained using quadratic programming methods that generate feasible iterates, such as feasible active set methods and interior point methods; see, for example, [11,13,14]. It is worth mentioning that, if the exact projection is used, then Subgradient-InexP method becomes the projected subgradient method considered in [3]. Several methods similar to the projected subgradient method have been studied in different papers, see [6,15]. However, as far as we know, none of them use the concept of feasible inexact projection.

The main tool used in our analysis of Subgradient-InexP method is a version of the inequality obtained in [16, Lemma 1.1]; see also a variant of it in [17, Lemma 2.1]. By using this inequality, we establish asymptotic convergence results and iteration-complexity bounds for the sequence generated by our method employing the well known exogenous stepsizes, Polyak's stepsizes, and dynamic stepsizes. We point out that these stepsizes have been discussed extensively in the related literature, including [3,6,8,17–19], where many of our results were inspired. Let us describe the results in the present and their relationship with the literature on the subject. With respect to the exogenous stepsize we establish convergence results without any compactness assumption, existence of a solution, and the iteration-complexity bound, which are similar to the well known bound presented in [3,17]. In particular, for $C = \mathbb{R}^n$, the convergence results merge into the ones presented in [16] and the iteration-complexity bound into [20, Theorem 3.2.2]. The asymptotic convergence result and the iteration-complexity bound obtained using Polyak's stepsizes are similar to the corespondent ones in [17,21,22] and [21], respectively. Regarding to the dynamic stepsize, we establish global convergence in objective values as address, for example, in [6,17]. In [18, Proposition 2.15], the authors presented the rate of convergence for another variant of subgradient method, known as incremental subgradient algorithms. This study allowed us to estimate an iteration-complexity bound for the dynamic stepsize.

The organization of the paper is as follows. In Section 2, we present some notation and basic results used in our presentation. In Section 3 we describe the Subgradient-InexP method with different choices for the stepsize. The main results of the present paper, including the converge theorems and iteration-complexity, are presented in Section 4. Some numerical experiments are provided in Section 5. We conclude the paper with some remarks in Section 6.

2. Notation and definitions

In this section, we present some notations, definitions, and results used throughout the paper. We are interested in

$$\min\{f(x) : \; x \in C\}, \tag{1}$$

where C is a closed and convex subset of $\mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function. We denote by

$$f^* := \inf_{x \in C} f(x), \tag{2}$$

its infimal value (possibly $-\infty$) and by Ω^* its solution set (possibly $\Omega^* = \varnothing$). The next concept will be useful in the analysis of the sequence generated by the subgradient method to solve (1).

Definition 2.1. A sequence $\{y^k\} \subset \mathbb{R}^n$ is said to be quasi-Fejér convergent with respect to a set $W \subset \mathbb{R}^n$ if, for every $w \in W$, there exists a sequence $\{\delta_k\} \subset \mathbb{R}$ such that $\delta_k \geq 0$, $\sum_{k=1}^{\infty} \delta_k < +\infty$, and

$$\|y^{k+1} - w\|^2 \leq \|y^k - w\|^2 + \delta_k, \qquad \forall\, k = 0, 1, \ldots.$$

When, $\delta_k = 0$, for all $k = 0, 1, \ldots$, $\{y^k\}$ is called Fejér convergent to a set W.

The main property of the quasi-Fejér convergent sequence is stated in the next result, and its proof can be found in [23].

Theorem 2.2. *Let $\{y^k\}$ be a sequence in $\mathbb{R}^n$. If $\{y^k\}$ is quasi-Fejér convergent to a nonempty set $W \subset \mathbb{R}^n$, then $\{y^k\}$ is bounded. If furthermore, a cluster point y of $\{y^k\}$ belongs to W, then $\lim_{k\to\infty} y^k = y$.*

To describe the method for solving the problem (1) we need to define, for each $\varepsilon \geq 0$, the ε-subdifferential $\partial_\varepsilon f(x)$ of a convex function f at $x \in \mathbb{R}^n$,

$$\partial_\varepsilon f(x) := \{s \in \mathbb{R}^n : \ f(y) \geq f(x) + \langle s, y - x\rangle - \varepsilon, \ \forall y \in \mathbb{R}^n\}. \tag{3}$$

We end this section by presenting important properties of the set ε-subdifferential of a convex function, which proofs follow by combining [24, Proposition 4.3.1(a)] and [25, Proposition 4.1.1, Proposition 4.1.2].

Proposition 2.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and $\varepsilon \geq 0$. The set $\partial_\varepsilon f(x)$ is nonempty, convex, and compact. Moreover, if $B \subset \mathbb{R}^n$ is a bounded set, then there exists a real number $L > 0$ such that $\|s\| < L$, for all $s \in \cup_{x\in B} \partial_\varepsilon f(x)$. In addition, if $\{\varepsilon_k\}$ is a bounded sequence of nonnegative real numbers, the sequence $\{x^k\}$ converges to $x \in \mathbb{R}^n$, and $s^k \in \partial_{\varepsilon_k} f(x^k)$ for all k, then the sequence $\{s^k\}$ is bounded.*

3. Subgradient-InexP method

Next, we present the subgradient method with a feasible inexact projections, which will be called *Subgradient-InexP method*. We begin by presenting the concept of relative feasible inexact projection, which is a variation of those presented in [9,10].

Definition 3.1. Let $\varphi_\gamma : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ be a relative error tolerance function such that

$$\varphi_\gamma(u, v, w) \leq \gamma_1 \|v - u\|^2 + \gamma_2 \|w - v\|^2 + \gamma_3 \|w - u\|^2, \qquad \forall\, u, v, w \in \mathbb{R}^n, \tag{4}$$

where $\gamma = (\gamma_1, \gamma_2, \gamma_3) \in \mathbb{R}^3_+$ are given forcing parameters, and $C \subset \mathbb{R}^n$ be a closed convex set. The *feasible inexact projection mapping* relative to $u \in C$ with relative error tolerance function φ_γ, denoted by $\mathscr{P}_C(\varphi_\gamma, u, \cdot) : \mathbb{R}^n \rightrightarrows C$ is the set-valued mapping defined as follows

$$\mathscr{P}_C(\varphi_\gamma, u, v) := \left\{ w \in C : \ \langle v - w, z - w \rangle \leq \varphi_\gamma(u, v, w), \ \forall\, z \in C \right\}. \tag{5}$$

Each point $w \in \mathscr{P}_C(\varphi_\gamma, u, v)$ is called a *feasible inexact projection of v onto C relative to u and with relative error tolerance function φ_γ.*

In the following, we present some remarks about the definition of the feasible inexact projection mapping onto the convex set C.

Remark 1. Let $C \subset \mathbb{R}^n$, $u \in C$ and φ_γ be as in Definition 3.1. Therefore, for all $v \in \mathbb{R}^n$, it follows from (5) that $\mathscr{P}_C(0, u, v)$ is the exact projection of v onto C; see [5, Proposition 2.1.3, p. 201]. Moreover, $\mathscr{P}_C(0, u, v) \in \mathscr{P}_C(\varphi_\gamma, u, v)$ concluding that $\mathscr{P}_C(\varphi_\gamma, u, v) \neq \varnothing$, for all $u \in C$ and $v \in \mathbb{R}^n$. Consequently, the set-valued mapping $\mathscr{P}_C(\varphi_\gamma, u, \cdot)$ is well-defined.

Next lemma is a variation of [26, Lemma 6]. It will play an important role in the remainder of this paper.

Lemma 3.2. *Let $v \in \mathbb{R}^n$, $u \in C$, $\gamma = (\gamma_1, \gamma_2, \gamma_3) \in \mathbb{R}^3_+$ and $w \in \mathscr{P}_C(\varphi_\gamma, u, v)$. Then, there holds*

$$\|w - x\|^2 \leq \|v - x\|^2 + \frac{2\gamma_1 + 2\gamma_3}{1 - 2\gamma_3} \|v - u\|^2, \qquad \forall\, x \in C,$$

for all $\gamma_2, \gamma_3 \in [0, 1/2)$.

Proof. Let $x \in C$. First note that $\|w - x\|^2 = \|v - x\|^2 - \|w - v\|^2 + 2\langle v - w, x - w \rangle$. Since $w \in \mathscr{P}_C(\varphi_\gamma, u, v)$ and $0 \leq \gamma_2 < 1/2$, combining the last equality with (4) and (5) we obtain

$$\begin{aligned} \|w - x\|^2 &\leq \|v - x\|^2 - (1 - 2\gamma_2)\|v - w\|^2 + 2\gamma_1 \|v - u\|^2 + 2\gamma_3 \|w - u\|^2 \\ &\leq \|v - x\|^2 + 2\gamma_1 \|v - u\|^2 + 2\gamma_3 \|w - u\|^2. \end{aligned} \tag{6}$$

On the other hand, we also have

$$\begin{aligned} \|w - u\|^2 &= \|v - u\|^2 + \|w - v\|^2 + 2\langle v - w, u - v \rangle \\ &= \|v - u\|^2 + \|w - v\|^2 + 2\langle v - w, u - w \rangle - 2\|w - v\|^2 \\ &= \|v - u\|^2 - \|w - v\|^2 + 2\langle v - w, u - w \rangle. \end{aligned}$$

Thus, due to $w \in \mathscr{P}_C(\varphi_\gamma, u, v)$ and $u \in C$, using (4), (5), $0 \leq \gamma_2 < 1/2$ and $0 \leq \gamma_3 < 1/2$, we have

$$\|w - u\|^2 \leq \frac{1 + 2\gamma_1}{1 - 2\gamma_3} \|v - u\|^2 - \frac{1 - 2\gamma_2}{1 - 2\gamma_3} \|w - v\|^2 \leq \frac{1 + 2\gamma_1}{1 - 2\gamma_3} \|v - u\|^2.$$

Therefore, combining the last inequality with (6), we obtain the desired inequality. $\square$

The conceptual subgradient method with feasible inexact projections for solving the Problem (1) is formally defined as follows:

Algorithm 1: Subgradient-InexP method

Step 0. Let $\{\varepsilon_k\}$ and $\{\gamma^k\}$ be sequences of nonnegative real numbers, where $\gamma^k = (\gamma_1^k, \gamma_2^k, \gamma_3^k)$. Let $x^0 \in C$ and set $k = 0$.

Step 1. If $0 \in \partial f(x^k)$, then **stop**. Otherwise, choose a non-null element $s^k \in \partial_{\varepsilon_k} f(x^k)$, compute a stepsize $t_k > 0$, (to be specified later), and take the next iterate as any point such that

$$x^{k+1} \in \mathscr{P}_C\left(\varphi_{\gamma^k}, x^k, x^k - t_k s^k\right).$$

Step 2. Set $k \leftarrow k+1$, and go to **Step 1**.

Let us describe the main features of the subgradient-InexP method. Firstly, we check if the current iterate x^k is a solution of Problem (1). If x^k is not a solution, then we choose a non-null element $s^k \in \partial_{\varepsilon_k} f(x^k)$, compute a stepsize $t_k > 0$, and take the next iterate $x^{k+1} \in C$ as any feasible inexact projection of $x^k - t_k s^k$ onto C relative to x^k with error tolerance given by $\varphi_{\gamma^k}(x^k, x^k - t_k s^k, x^{k+1})$, i.e., $x^{k+1} \in \mathscr{P}_C\left(\varphi_{\gamma^k}, x^k, x^k - t_k s^k\right)$. We remark that if $\varphi_{\gamma^k} \equiv 0$, then $\mathscr{P}_C\left(0, x^k, x^k - t_k s^k\right)$ is the exact projection of $x^k - t_k s^k$ onto C, and Algorithm 1 amounts to the classical projected subgradient method, see for example [3] and the references therein. Among the several possible choices that appeared in literature on subject, see for example [2,5,17], we studied three well known strategies, beginning with exogenous stepsize.

Rule 1 (Exogenous stepsize)**.** Let $\mu \geq 0$. Take exogenous sequences $\{\alpha_k\}$ and $\{\varepsilon_k\}$ of non-negative real numbers satisfying the following conditions: the sequence $\{\varepsilon_k\}$ is nonincreasing and

$$\sum_{k=0}^{\infty} \alpha_k = +\infty, \qquad \sum_{k=0}^{\infty} \alpha_k^2 < +\infty, \qquad \varepsilon_k \leq \mu \alpha_k, \qquad k = 0, 1, \ldots. \tag{7}$$

Given $s^k \in \partial_{\varepsilon_k} f(x^k)$, define the stepsize t_k as the following nonnegative real number

$$t_k := \frac{\alpha_k}{\eta_k}, \qquad \eta_k := \max\left\{1, \|s^k\|\right\}, \qquad k = 0, 1, \ldots. \tag{8}$$

The stepsize in Rule 1 is one the most popular. It have been used in several paper for analyzing subgradient method; see for example, [3,16–19].

From now on we assume that there exist $0 \leq \bar{\gamma}_2 < 1/2$ *and* $0 \leq \bar{\gamma}_3 < 1/2$*, such that* $\{\gamma_2^k\} \subset [0, \bar{\gamma}_2)$*,* $\{\gamma_1^k\} \subset [0, \bar{\gamma}_1)$ *and* $\{\gamma_3^k\} \subset [0, \bar{\gamma}_3)$*. For future references define*

$$\theta := \frac{1 + 2\bar{\gamma}_1}{1 - 2\bar{\gamma}_3} > 0. \tag{9}$$

To define the next stepsize, we need to known the optimum value f^* given in (2). In [27, p.142], is present some examples of problems for which the optimum value are known. The statement of the Polyak's stepsize is as follows.

Rule 2 (Polyak's stepsize)**.** Assume that $\Omega^* \neq \varnothing$ and the optimal value $f^* > -\infty$ is known. Let $\mu \geq 0$, $\underline{\beta} > 0$, $\bar{\beta} > 0$ and take exogenous sequences $\{\beta_k\}$ and $\{\varepsilon_k\}$ of real numbers satisfying the following conditions: the sequence $\{\varepsilon_k\}$ is nonincreasing and

$$0 < \underline{\beta} \le \beta_k \le \bar{\beta} < \frac{1}{2\mu+\theta}, \qquad 0 < \varepsilon_k \le \mu\beta_k[f(x^k) - f^*], \qquad k = 0, 1, \ldots. \quad (10)$$

Given $s^k \in \partial_{\varepsilon_k} f(x^k)$, $s^k \neq 0$, define the stepsize t_k as the following nonnegative real number

$$t_k = \beta_k \frac{f(x^k) - f^*}{\|s^k\|^2}, \qquad k = 0, 1, \ldots. \quad (11)$$

The stepsize in Rule 2 was introduced in [22] and has been used in several papers, including the ones [17–19]. In general, in practical problems the optimum value (2) is not known. In this case, we may modify the stepsize (11) by replacing the optimum value (2) with a suitable estimate in each iteration. This leads to the dynamic stepsize rule as follows.

Rule 3 (Dynamic stepsize)**.** Let $\mu \ge 0$, $\underline{\beta} > 0$, $\bar{\beta} > 0$, and take exogenous sequences $\{\beta_k\}$ and $\{\varepsilon_k\}$ of real numbers satisfying the following conditions: the sequence $\{\varepsilon_k\}$ is nonincreasing and

$$0 < \underline{\beta} \le \beta_k \le \bar{\beta} < \frac{2}{2\mu+\theta}, \qquad 0 < \varepsilon_k \le \mu\beta_k[f(x^k) - f^k_{lev}], \quad (12)$$

for all $k = 0, 1, \ldots$, where f^k_{lev} will be specified later (see Section 4.3). Given $s^k \in \partial_{\varepsilon_k} f(x^k)$ such that $s^k \neq 0$, define the stepsize t_k as the following nonnegative real number

$$t_k = \frac{\tilde{t}_k}{\|s^k\|}, \qquad \tilde{t}_k = \beta_k \frac{f(x^k) - f^k_{lev}}{\|s^k\|}, \qquad k = 0, 1, \ldots. \quad (13)$$

The dynamic stepsize in Rule 3 is based on the ideas of [28]; see also [6]. This rule has been used in several papers, see for example [8,17,19].

From now on we assume that the sequence $\{x^k\}$ is generated by Algorithm 1, with one of the three above strategies for choosing the stepsize, is infinite.

4. Analysis of the subgradient-InexP method

In the following, we state and prove our first result to analyze the sequence $\{x^k\}$ generated by Algorithm 1. The obtained inequality in next lemma is its counterpart for unconstrained optimization provided in [16, Lemma 1.1)]. As we shall see, this inequality will be the main tool in our asymptotic convergence analysis, as well as in the iteration-complexity analysis.

Lemma 4.1. *Let $\theta > 0$ be as defined* (9)*. For all $x \in C$, the following inequality holds*

$$\|x^{k+1} - x\|^2 \le \|x^k - x\|^2 + \theta t_k^2 \|s^k\|^2 - 2t_k \left[f(x^k) - f(x) - \varepsilon_k \right], \quad (14)$$

for all $k = 0, 1, \ldots$.

Proof. Let $x \in C$. To simply the notations we set $z^k := x^k - t_k s^k$. Due to $x^{k+1} \in \mathscr{P}_C\left(\varphi_{\gamma^k}, x^k, z^k\right)$ and $x^k \in C$, we apply Lemma 3.2 with $w = x^{k+1}$, $v = z^k$, $u = x^k$, and $\varphi_\gamma = \varphi_{\gamma^k}$

to conclude

$$\|x^{k+1}-x\|^2 \leq \|z^k-x\|^2 + \left(\frac{2\gamma_1^k+2\gamma_3^k}{1-2\gamma_3^k}\right) t_k^2\|s^k\|^2. \tag{15}$$

On the other hand, due to $z^k = x^k - t_k s^k$, after some algebraic manipulations, we obtain

$$\|z^k-x\|^2 = \|x^k-x\|^2 + t_k^2\|s^k\|^2 + 2t_k\langle s^k, x-x^k\rangle.$$

Since $s^k \in \partial_{\varepsilon_k} f(x^k)$, (3) implies that $\langle s^k, z-x^k\rangle \leq f(z)-f(x^k)+\varepsilon_k$. Thus,

$$\|z^k-x\|^2 \leq \|x^k-x\|^2 + t_k^2\|s^k\|^2 + 2t_k\left[f(x)-f(x^k)+\varepsilon_k\right].$$

Therefore, combining last inequality with (15) we conclude that

$$\|x^{k+1}-x\|^2 \leq \|x^k-x\|^2 + \left(\frac{1+2\gamma_1^k}{1-2\gamma_3^k}\right) t_k^2\|s^k\|^2 - 2t_k\left[f(x^k)-f(x)-\varepsilon_k\right].$$

Considering that $0 \leq \gamma_3^k < \bar{\gamma}_3 < 1/2$ and $0 \leq \gamma_1^k < \bar{\gamma}_1$, and using (9), we obtain (14). □

4.1. Analysis of the subgradient-InexP method with exogenous stepsize

In this section we will analyze the subgradient-InexP method with stepsizes satisfying Rule 1. For that, *throughout this section we assume also that $\{x^k\}$ is a sequence generated by Algorithm 1 with the stepsize given by Rule 1 and, define*

$$\rho := 2\mu + \theta > 0. \tag{16}$$

First of all, note that under the above assumptions, Lemma 4.1 becomes as follows.

Lemma 4.2. *Let $\rho > 0$ be as in* (16)*. For all $x \in C$, the following inequality holds*

$$\|x^{k+1}-x\|^2 \leq \|x^k-x\|^2 + \rho\alpha_k^2 - 2\frac{\alpha_k}{\eta_k}\left[f(x^k)-f(x)\right], \qquad k = 0,1,\ldots.$$

Proof. The definition of t_k in (8) implies $t_k \leq \alpha_k$, which combined with the last inequality in (7) yields $2t_k\varepsilon_k \leq 2\mu\alpha_k^2$. Moreover, (8) also implies that $t_k^2\|s^k\|^2 \leq \alpha_k^2$. Therefore, using (16), the desired inequality follows directly from (14). □

To proceed with the analysis of Algorithm 1, we also need the following auxiliary set

$$\Omega := \left\{x \in C:\ f(x) \leq \inf_k f(x^k)\right\}. \tag{17}$$

It is worth mentioning that, in principle, set Ω can be empty and, in such case, $f^* = -\infty$. In the next lemma we analyze the behavior of the sequence $\{x^k\}$ under the hypothesis that $\Omega \neq \varnothing$.

Lemma 4.3. *If $\Omega \neq \varnothing$, then $\{x^k\}$ is quasi-Féjer convergent to Ω. Consequently, $\{x^k\}$ is bounded.*

Proof. Since $\Omega \neq \varnothing$, take $x \in \Omega$. Thus, by using the definition of Ω in (17) and Lemma 4.2, we conclude that $\|x^{k+1}-x\|^2 \leq \|x^k - x\|^2 + \rho\alpha_k^2$, for all $k = 0,1,\ldots$. Hence, using the first inequality in (7), the first statement of the lemma follows from Definition 2.1. The second statement of the lemma follows from the first part of Theorem 2.2. $\square$

Now, we are ready to prove the main result of this section, which refers to the asymptotic convergence of $\{x^k\}$. We remark that in the first part of the next theorem we do not assume neither $\Omega^* \neq \varnothing$ nor that f^* is finite.

Theorem 4.4. *The following equality holds*

$$\liminf_k f(x^k) = f^*. \tag{18}$$

In addition, if $\Omega^* \neq \varnothing$ then the sequence $\{x^k\}$ converges to a point $x^* \in \Omega^*$.

Proof. Assume by contradiction that $\liminf_k f(x^k) > f^*$. In this case, we have $\Omega \neq \varnothing$. Consequently by Lemma 4.3, we conclude that $\{x^k\}$ is bounded. Letting $x \in \Omega$, there exist $\tau > 0$ and $k_0 \in \mathbb{N}$ such that $f(x) < f(x^k) - \tau$, for all $k \geq k_0$. Hence, using Lemma 4.2, we have

$$\|x^{k+1}-x\|^2 \leq \|x^k - x\|^2 + \rho\alpha_k^2 - 2\frac{\alpha_k}{\eta_k}\tau, \qquad k = k_0, k_0+1, \ldots. \tag{19}$$

On the other hand, it follows from (7) that the sequence $\{\varepsilon_k\}$ is bounded. Thus, considering that $\{x^k\}$ is bounded, Proposition 2.3 implies that $\{s^k\}$ is also bounded. Let $c > 0$ be such that $\|s^k\| \leq c$, for all $k \geq 0$. Hence, using second equality in (8), we have $\eta_k = \max\{1, \|s^k\|\} \leq \max\{1, c\} =: \Gamma$. Thus, letting $\ell \in \mathbb{N}$ and using (19), we conclude that

$$\frac{2\tau}{\Gamma}\sum_{j=k_0}^{\ell+k_0}\alpha_j \leq \|x^{k_0}-x\|^2 - \|x^{k_0+\ell+1}-x\| + \rho\sum_{j=k_0}^{\ell+k_0}\alpha_j^2 \leq \|x^{k_0}-x\|^2 + \rho\sum_{j=k_0}^{\ell+k_0}\alpha_j^2.$$

Since the last inequality holds for all $\ell \in \mathbb{N}$ then, by using the first two conditions on $\{\alpha_k\}$ in (7), we have a contraction. Therefore, (18) holds. For proving the last statement, let us assume that $\Omega^* \neq \varnothing$. In this case, we also have $\Omega \neq \varnothing$ and, from Lemma 4.3, the sequence $\{x^k\}$ is bounded and quasi-Féjer convergent to Ω. The equality (18) implies that $\{f(x^k)\}$ has a decreasing monotonous subsequence $\{f(x^{k_j})\}$ such that $\lim_{j\to\infty} f(x^{k_j}) = f^*$. Without lose of generality, we can assume that $\{f(x^k)\}$ is decreasing, is monotonous, and converges to f^*. Being bounded, the sequence $\{x^k\}$ has a convergent subsequence $\{x^{k_\ell}\}$. Let us say that $\lim_{\ell\to\infty} x^{k_\ell} = x^*$, which by the continuity of f implies $f(x^*) = \lim_{\ell\to\infty} f(x^{k_\ell}) = f^*$, and then $x^* \in \Omega^*$. Hence, $\{x^k\}$ has an cluster point $x^* \in \Omega$, and due to $\{x^k\}$ be quasi-Féjer convergent to Ω, Theorem 2.2 implies that $\{x^k\}$ converges to x^*. $\square$

Next theorem presents an iteration-complexity bound; similar bound can be found in [20, Theorem 3.2.2].

Theorem 4.5. *Assume that the sequence $\{x^k\}$ converges to a point $x^* \in \Omega^*$. Then, for every $N \in \mathbb{N}$, the following inequality holds*

$$\min\{f(x^k) - f^* : \ k = 0,1,\ldots,N\} \leq \Gamma\frac{\|x^0 - x^*\|^2 + \rho\sum_{k=0}^N \alpha_k^2}{2\sum_{k=0}^N \alpha_k}.$$

Proof. Since $\{\varepsilon_k\}$ and $\{x^k\}$ are bounded sequences, then using Proposition 2.3, it follows that $\{s^k\}$ is also bounded, i.e. there exists $c > 0$ such that $\|s^k\| \leq c$, for all $k \geq 0$. Therefore, using the definition of η_k in (8), we have $\eta_k = \max\{1, \|s^k\|\} \leq \max\{1, c\} =: \Gamma$. Now, applying Lemma 4.2 with $x = x^*$ and due to $f^* = f(x^*)$, we obtain

$$\frac{2\alpha_k}{\Gamma}[f(x^k) - f^*] \leq \|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 + \rho\alpha_k^2, \qquad k = 0, 1, \ldots.$$

Hence, performing the sum of the above inequality for $k = 0, 1, \ldots, N$, we have

$$\frac{2}{\Gamma}\sum_{k=0}^{N}\alpha_k[f(x^k) - f^*] \leq \|x^0 - x^*\|^2 - \|x^{N+1} - x^*\|^2 + \rho\sum_{k=0}^{N}\alpha_k^2.$$

Therefore,

$$\frac{2}{\Gamma}\min\left\{f(x^k) - f^* : \ k = 0, 1, \ldots, N\right\}\sum_{k=0}^{N}\alpha_k \leq \|x^0 - x^*\|^2 + \rho\sum_{k=0}^{N}\alpha_k^2,$$

which is equivalent to the desired inequality. □

4.2. Analysis of the subgradient-InexP method with Polyak's stepsize rule

In this section we will analyze the subgradient-InexP method with Polyak's step sizes. *Throughout this section, we assume also that $\Omega^* \neq \varnothing$ and $\{x^k\}$ is a sequence generated by Algorithm 1 with the stepsize given by Rule 2.*

Lemma 4.6. *Let $x \in \Omega^*$. Then, the following inequality holds*

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 - \underline{\beta}\frac{\left[f(x^k) - f^*\right]^2}{\|s^k\|^2}, \qquad k = 0, 1, \ldots. \tag{20}$$

Proof. Considering that $x \in \Omega^*$ we have $f^* = f(x)$. The combination of (10) with (11) implies $2t_k\varepsilon_k \leq 2\mu\beta_k^2[f(x^k) - f^*]^2/\|s^k\|^2$. Moreover, (11) also implies that $t_k^2\|s^k\|^2 = \beta_k^2[f(x^k) - f^*]^2/\|s^k\|^2$. Thus, we conclude form (14) that

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 - (2 - \theta\beta_k - 2\mu\beta_k)\,\beta_k\frac{\left[f(x^k) - f^*\right]^2}{\|s^k\|^2}, \qquad k = 0, 1, \ldots. \tag{21}$$

On the other hand, (10) gives us $\beta_k < 1/(2\mu + \theta)$, which is equivalent to $2 - \theta\beta_k - 2\mu\beta_k > 1$. Therefore, since (10) also gives $\underline{\beta} \leq \beta_k$, we conclude that (21) implies (20). □

In the following theorem we present our main result about the asymptotic convergence of $\{x^k\}$. It has as correspondent result in [22, Theorem 1]; see also [17].

Theorem 4.7. *The sequence $\{x^k\}$ converges to a point $x^* \in \Omega^*$.*

Proof. Let $x \in \Omega^*$. Then, Lemma 4.6 implies $\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2$, for all $k = 0, 1, \ldots$. Thus, $\{x^k\}$ is Fejér convergent to Ω^*. Since $\Omega^* \neq \varnothing$, Theorem 2.2 implies that $\{x^k\}$ is bounded. By using Proposition 2.3, we conclude that there exists $c > 0$ such that $\left\|s^k\right\| \leq c$, for $k = 0, 1, \ldots$. Then, from (20), after some algebra, we have

$$\left[f(x^k) - f^*\right]^2 \leq \frac{c^2}{\underline{\beta}} \left(\|x^k - x\|^2 - \|x^{k+1} - x\|^2\right), \qquad k = 0, 1, \ldots.$$

Thus, performing the sum of the this inequality for $j = 0, 1, \ldots, \ell$, we obtain

$$\sum_{j=0}^{\ell} \left[f(x^j) - f^*\right]^2 \leq \frac{c^2}{\underline{\beta}} \left(\|x^0 - x\|^2 - \|x^{\ell+1} - x\|^2\right) \leq \frac{c^2}{\underline{\beta}} \|x^0 - x\|^2.$$

Considering that the last inequality holds for all $\ell \in \mathbb{N}$, we conclude that $\lim_{k\to+\infty} f(x^k) = f^*$. Let x^* be a cluster point of $\{x^k\}$ and $\{x^{k_j}\}$ a subsequence of $\{x^k\}$ such that $\lim_{j\to+\infty} x^{k_j} = x^*$. Since f is continuous, we have $f(x^*) = \lim_{j\to+\infty} f(x^{k_j}) = f^*$. Therefore, $x^* \in \Omega^*$. Since $\{x^k\}$ is quasi-Fejér convergent to a set Ω^*, it follows from Theorem 2.2 that $\{x^k\}$ converges to x^*. □

The next result presents an iteration-complexity bound, which is a version of [21, Theorem 1].

Theorem 4.8. *For every $N \in \mathbb{N}$, the following inequality holds*

$$\min\{f(x^k) - f^* : \ k = 0, 1, \ldots, N\} \leq \frac{c}{\sqrt{\underline{\beta}(N+1)}} \|x^0 - x^*\|,$$

where $c \geq \max\{\|s^k\| : \ k = 0, 1, \ldots\}$.

Proof. Applying Lemma 4.6 with $x = x^*$, where $f^* = f(x^*)$, we obtain

$$\underline{\beta} \frac{\left[f(x^k) - f^*\right]^2}{\|s^k\|^2} \leq \|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2, \qquad k = 0, 1, \ldots.$$

Performing the sum of the above inequality for $k = 0, 1, \ldots, N$, we conclude that

$$\sum_{k=0}^{N} \frac{\left[f(x^k) - f^*\right]^2}{\|s^k\|^2} \leq \frac{1}{\underline{\beta}} \|x^0 - x^*\|^2.$$

Since $\{x^k\}$ is bounded, by using Proposition 2.3, we conclude that there exists $c > 0$ such that $\left\|s^k\right\| \leq c$, for $k = 0, 1, \ldots$. Thus, we have

$$\sum_{k=0}^{N} \left[f(x^k) - f^*\right]^2 \leq \frac{c^2}{\underline{\beta}} \|x^0 - x^*\|^2.$$

Therefore,

$$(N+1) \min\left\{\left[f(x^k) - f^*\right]^2 : \ k = 0, 1, \ldots, N\right\} \leq \frac{c^2}{\underline{\beta}} \|x^0 - x^*\|^2,$$

which is equivalent to the desired inequality. □

4.3. Analysis of the subgradient-InexP method with dynamic stepsize

Next we consider the Subgradient-InexP method employing the dynamic stepsize Rule 3, which guarantees that $\{f_{lev}^k\}$ converges to the optimum value f^*. In the following we present formally the algorithm which compute f_{lev}^k. This scheme was introduced in [28]; see also [6].

Algorithm 2: Subgradient-InexP employing the dynamic stepsize rule (SInexPD)

Step 0. Select $x^0 \in C, \delta_0 > 0$, and $R > 0$. Set $k = 0, \sigma_0 = 0, f_{rec}^{-1} = \infty, \ell = 0, k(\ell) = 0$.
Step 1. If $f(x^k) < f_{rec}^{k-1}$, set $f_{rec}^k = f(x^k)$ and $x_{rec}^k = x^k$, else set $f_{rec}^k = f_{rec}^{k-1}$ and $x_{rec}^k = x_{rec}^{k-1}$.
Step 2. If $0 \in \partial f(x^k)$, then **stop**.
Step 3. If $f(x^k) \le f_{rec}^{k(\ell)} - \frac{1}{2}\delta_\ell$, set $k(\ell+1) = k, \sigma_k = 0, \delta_{\ell+1} = \delta_\ell$, replace ℓ by $\ell+1$, and go to Step 5.
Step 4. If $\sigma_k > R$, set $k(\ell+1) = k, \sigma_k = 0, \delta_{\ell+1} = \frac{1}{2}\delta_\ell$, $x^k = x_{rec}^k$, and $\ell \leftarrow \ell+1$.
Step 5. Set $f_{lev}^k := f_{rec}^{k(\ell)} - \delta_\ell$. Select $\beta_k \in [\underline{\beta}, \bar{\beta}]$ and calculate x^{k+1} via Algorithm 1 with the stepsize given by Rule 3.
Step 6. Set $\sigma_{k+1} := \sigma_k + \tilde{t}_k$, where $\tilde{t}_k$ is defined in (13), $k \leftarrow k+1$, and go to Step 1.

Following [6,17–19], we describe the main features of the Subgradient-InexP method.

Remark 2. Note that in Step 1, f_{rec}^k keeps the record of the smallest functional value attained by the iterates generated so far, i.e., $f_{rec}^k := \min\{f(x^j) : \ j = 0, \dots, k\}$. Splitting the iterations into groups

$$K_\ell := \{k(\ell), k(\ell)+1, \dots, k(\ell+1)-1\}, \quad \ell = 0, 1, \dots,$$

Algorithm 2 uses the same target level $f_{lev}^k = f_{rec}^{k(\ell)} - \delta_\ell$, for $k \in K_\ell$. Also, note that the target level is update only if sufficient descent or oscillation is detected (Step 3 or Step 4, respectively). Whenever σ_k exceeds the upper bound R, the parameter δ_ℓ is decreased, which increases the target level f_{lev}^k.

From now on, we assume that Algorithm 2 generates an infinite sequence. In the next theorem, we present the result about the asymptotic convergence of the sequence $\{x^k\}$. It is the versions of [6, Theorem 1] and [17, Proposition 2.7] by using inexact projections.

Theorem 4.9. *There holds* $\inf_{k \ge 0} f(x^k) = f^*$.

Proof. Since $x^k \in C$ and $x^{k+1} \in \mathscr{P}_C\left(\varphi_{\gamma^k}, x^k, x^k - t_k s^k\right)$, by the first equality in (13) and applying Lemma 3.2 with $w = x^{k+1}$, $v = x^k - t_k s^k$, $x = x^k$, $u = x^k$ and $\varphi_\gamma = \varphi_{\gamma^k}$, we conclude that

$$\|x^{k+1} - x^k\| \le \sqrt{\frac{1+2\bar{\gamma}_1}{1-2\bar{\gamma}_3}}\, t_k \|s^k\| \le \sqrt{\frac{1+2\bar{\gamma}_1}{1-2\bar{\gamma}_3}}\, \tilde{t}_k, \qquad k = 0, 1, \dots.$$

We claim that the index ℓ goes to $+\infty$ and either $\inf_{k \ge 0} f(x^k) = -\infty$ or $\lim_{l \to \infty} \delta_\ell = 0$. Indeed, assume that ℓ takes only a finite number of values, i.e., $\ell < \infty$. Thus, from the last inequality we have

$$\|x^k - x^{k(\ell)}\| \le \sum_{j=k(\ell)}^{k} \|x^{j+1} - x^j\| \le \sqrt{\frac{1+2\bar{\gamma}_1}{1-2\bar{\gamma}_3}} \sum_{j=k(\ell)}^{k} \tilde{t}_j.$$

Hence, considering that $\sigma_k + \tilde{t}_k = \sigma_{k+1} \le R$, for all $k \ge k(\ell)$, we conclude that

$$\|x^k - x^{k(\ell)}\| \le \sqrt{\frac{1+2\bar{\gamma}_1}{1-2\bar{\gamma}_3}} \sum_{j=k(\ell)}^{k} \tilde{t}_j = \sqrt{\frac{1+2\bar{\gamma}_1}{1-2\bar{\gamma}_3}}\, \sigma_{k+1} \le \sqrt{\frac{1+2\bar{\gamma}_1}{1-2\bar{\gamma}_3}}\, R. \tag{22}$$

Hence, $\{x^k\}$ is bounded. Besides, from the last condition in (12), the sequence $\{\varepsilon_k\}$ is bounded and, by using Proposition 2.3, $\{s^k\}$ is also bounded. Moreover, by (22) we also conclude $\sum_{j=k(\ell)}^{+\infty} \tilde{t}_j < +\infty$, which implies $\lim_{k\to\infty} \tilde{t}_k = 0$. Thus, due to $\beta_k \in [\underline{\beta}, \bar{\beta}]$, it follows from second equality in (13) that

$$\lim_{k\to\infty} [f(x^k) - f_{lev}^k] = 0. \tag{23}$$

On the other hand, Steps 3 and 5 of Algorithm 2 yield

$$f(x^k) > f_{rec}^{k(\ell)} - \frac{1}{2}\delta_\ell = f_{lev}^k + \delta_\ell - \frac{1}{2}\delta_\ell = f_{lev}^k + \frac{1}{2}\delta_\ell \qquad k = k(\ell), k(\ell)+1, \ldots,$$

contradicting (23). Therefore, ℓ goes to $+\infty$. Now, suppose that $\lim_{\ell\to\infty} \delta_\ell = \delta > 0$. Then, from Steps 3 and 4 of Algorithm 2, it follows that for all ℓ large enough, we have $\delta_\ell = \delta$ and $f_{rec}^{k(\ell+1)} \le f_{rec}^{k(\ell)} - \frac{1}{2}\delta$, implying that $\inf\limits_{k\ge 0} f(x^k) = -\infty$, which concludes the claim. If $\lim_{\ell\to\infty} \delta_\ell > 0$ then, according to above claim, we have $\inf_{k\ge 0} f(x^k) = -\infty$, obtain the desired result. Now, we assume by contradiction that $\lim_{\ell\to\infty} \delta_\ell = 0$ and $\inf_{k\ge 0} f(x^k) > f^*$. Thus, it follows from Remark 2 that $\inf_{k\ge 0} f_{rec}^k = \inf_{k\ge 0} f(x^k)$. Hence, we conclude that $\inf_{k\ge 0} f_{rec}^k > f^*$. In this case, by using the definition of $\{f_{lev}^k\}$ in Step 5 and taking into account that $\lim_{\ell\to\infty} \delta_\ell = 0$, we conclude that

$$\inf_{k\ge 0} f_{lev}^k = \inf_{\ell\ge 0} (f_{rec}^{k(\ell)} - \delta_\ell) = \inf_{\ell\ge 0} f_{rec}^{k(\ell)} > f^*.$$

Therefore, there exist $\bar{\delta} > 0$, $\bar{x} \in C$ and $\bar{k} \in \mathbb{N}$ such that

$$f_{lev}^k - f(\bar{x}) \ge \bar{\delta}, \qquad k = \bar{k}, \bar{k}+1, \ldots. \tag{24}$$

Hence, by using the definition of $\tilde{t}_k$ in (13), it follows from (24) that

$$\tilde{t}_k = \beta_k \frac{f(x^k) - f_{lev}^k}{\|s^k\|} < \bar{\beta} \frac{f(x^k) - f(\bar{x}) - \bar{\delta}}{\|s^k\|}, \qquad k = \bar{k}, \bar{k}+1, \ldots. \tag{25}$$

Now, applying Lemma 4.1 with $x = \bar{x}$ and then using (11) and (13), we obtain

$$\|x^{k+1} - \bar{x}\|^2 \le \|x^k - \bar{x}\|^2 + \tilde{t}_k \left(\theta \tilde{t}_k - \frac{2}{\|s^k\|} \left[f(x^k) - f(\bar{x}) - \varepsilon_k \right] \right),$$

for all $k = \bar{k}, \bar{k}+1, \ldots$. Thus, the combination of the last inequality with (24), (25), and the last inequality in (12) yields

$$\|x^{k+1} - \bar{x}\|^2 \le \|x^k - \bar{x}\|^2 + \frac{\tilde{t}_k}{\|s^k\|} \left(\left[(2\mu + \theta)\bar{\beta} - 2\right] \left[f(x^k) - f(\bar{x})\right] - (2\mu + \theta)\,\bar{\beta}\,\bar{\delta} \right),$$

for all $k = \bar{k}, \bar{k}+1, \ldots$. On the other hand, (12) implies that $\bar{\beta} < 2/(2\mu + \theta)$, which is equivalent to $(2\mu + \theta)\bar{\beta} - 2 < 0$. Thus, by using that $f(x^k) \ge f^k_{lev}$ for all $k = 0, 1, \ldots$, the last inequality together (24) imply

$$\|x^{k+1} - \bar{x}\|^2 \le \|x^k - \bar{x}\|^2 - \frac{\tilde{t}_k}{\|s^k\|} (2\mu + \theta)\,\bar{\beta}\,\bar{\delta}, \qquad k = \bar{k}, \bar{k}+1, \ldots. \tag{26}$$

Hence, $\|x^{k+1} - \bar{x}\| \le \|x^{\bar{k}} - \bar{x}\|$, for all $k \ge \bar{k}$, which implies that $\{x^k\}$ is bounded. Besides, by using (23), it follows from the last condition in (12) that the sequence $\{\varepsilon_k\}$ is also bounded. Thus, using Proposition 2.3, we conclude that there exists $c > 0$ such that $\|s^k\| \le c$, for all $k \ge 0$, which together (26), yield

$$\frac{\bar{\beta}\,\bar{\delta}}{c} (2\mu + \theta) \sum_{k=\bar{k}}^{\infty} \tilde{t}_k \le \|x^{\bar{k}} - \bar{x}\|^2 < +\infty.$$

Since $\sigma_k = \sum_{j=k(\ell)}^{k(\ell+1)-1} \tilde{t}_j$, the last inequality implies that there exists $\ell_0 \in \mathbb{N}$ such that

$$\sigma_{k(\ell+1)} \le \sum_{k=k(\ell)}^{\infty} \tilde{t}_k < R, \qquad \ell = \ell_0, \ell_0 + 1 \ldots.$$

Hence, Step 4 in Algorithm 2 cannot occur infinitely to decrease δ_ℓ, contradicting the fact that $\lim_{\ell \to \infty} \delta_\ell = 0$. Therefore, the result follows and the proof is concluded. □

The next result presents an iteration-complexity bound for the subgradient-InexP method with the stepsize given by Rule 3, which is a version of [18, Proposition 2.15] for our algorithm.

Theorem 4.10. *Assume that the sequence $\{x^k\}$ converges to a point $x^* \in \Omega^*$. Let $\delta_0 > 0$ be given in Algorithm 2 and $c \ge \max\{\|s^k\| : \ k = 0, 1, \ldots\}$. Then,*

$$\min\{f(x^k) - f^* : \ k = 0, 1, \ldots, N\} \le \delta_0, \tag{27}$$

where N is the largest positive integer such that

$$\sum_{k=0}^{N-1} \left(\beta_k \left[2 - (2\mu + \theta)\beta_k\right] \delta_k^2\right) \le \left(c\|x^0 - x^*\|\right)^2. \tag{28}$$

Proof. Assume by contradiction that (27) does not holds. Thus, for all k with $0 \le k \le N$ we have $f(x^k) > f^* + \delta_0$. Hence, considering that $\delta_\ell \le \delta_0$ for all ℓ, we have

$$f^k_{lev} = f^{k(\ell)}_{rec} - \delta_\ell > f^* + \delta_0 - \delta_\ell \ge f^*, \qquad k = 0, \ldots, N. \tag{29}$$

On the other hand, the combination of the last inequality in (12) with (13) gives $2t_k\varepsilon_k \leq 2\mu\beta_k^2\left[f(x^k)-f_{lev}^k\right]^2/\|s^k\|^2$. Moreover, by using (13) we obtain that $t_k^2\|s^k\|^2 = \beta_k^2\left[f(x^k)-f_{lev}^k\right]^2/\|s^k\|^2$. Thus, using (29), Lemma 4.1 with $x = x^* \in \Omega^*$, and taking into account that $\beta_k \in [\underline{\beta},\bar{\beta}]$, we conclude that

$$\|x^{k+1}-x^*\|^2 \leq \|x^k-x^*\|^2 - \beta_k\left[2-(2\mu+\theta)\beta_k\right]\frac{\left[f(x^k)-f_{lev}^k\right]^2}{\|s^k\|^2}. \tag{30}$$

Since $\{x^k\}$ converges to $x^* \in \Omega^*$, Proposition 2.3 implies that there exists $c > 0$ such that $\left\|s^k\right\| \leq c$, for $k = 0,1,\ldots$. Furthermore, using the fact $f(x^k)-f_{lev}^k \geq \delta_k$, $0 \leq k \leq N$, (30) yields

$$\|x^{k+1}-x^*\|^2 \leq \|x^k-x^*\|^2 - \beta_k\left[2-(2\mu+\theta)\beta_k\right]\frac{\delta_k^2}{c^2}.$$

Performing the sum of the above inequality for $k = 0,1,\ldots,N$, we conclude that

$$\sum_{k=0}^{N}\left(\beta_k\left[2-(2\mu+\theta)\beta_k\right]\frac{\delta_k^2}{c^2}\right) \leq \|x^0-x^*\|^2,$$

which contradicts (28). $\square$

5. Numerical results

Our intention in this section is to report some numerical results in order to illustrate the practical behavior of SInexPD Algorithm when C is a compact convex set. We implemented SInexPD Algorithm in Fortran 90 considering set C in the general form $C = \{x \in \mathbb{R}^n : \ h(x) = 0, g(x) \leq 0\}$, where $h : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^p$ are smooth functions. At each iteration k, the Frank-Wolfe algorithm is used to compute a feasible inexact projection as explained below. The algorithm codes are freely available at `https://lfprudente.ime.ufg.br/up/948/o/SInexPD.zip`.

5.1. Frank-Wolfe algorithm to find an approximated projection

In this section we use the *Frank-Wolfe algorithm* also known *conditional gradient method* to find an inexact projection onto a compact convex set $C \subset \mathbb{R}^n$; papers dealing with this method include [29–34]. The exact projection of $v \in \mathbb{R}^n$ onto C is the solution of the following convex quadratic optimization problem

$$\min_{w\in C}\psi(w) := \frac{1}{2}\|w-v\|^2. \tag{31}$$

Assume that $v \notin C$. Let us describe the subroutine, which we nominate *FW-Procedure*, for finding an approximated solution of (31) relative to a point $u \in C$, i.e., a point belonging to the set $\mathscr{P}_C(\varphi_\gamma,u,v)$, where the error tolerance mapping φ_γ and the set valued mapping $\mathscr{P}_C(\varphi_\gamma,u,.)$ are given in Definition 3.1.

Since ψ is strictly convex, we conclude from (32) that $\psi(z) > \psi(w^k)+g_k^*$, for all $z \in C$ such that $z \neq w^k$. Setting $\psi^* := \min_{w\in C}\psi(w)$ we have $\psi(w^k) \geq \psi^* \geq \psi(w^k)+g_k^*$, which

FW-Procedure to compute $w^+ \in \mathscr{P}_C(\varphi_{\gamma,\theta,\lambda}, u, v)$

Step 0. Set $w^1 = u$ and $k = 1$.

Step 1. Call the linear optimization oracle (or simply LO oracle) to compute

$$z^k := \arg\min_{z\in C}\langle w^k - v, z - w^k\rangle, \qquad g_k^* := \langle w^k - v, z^k - w^k\rangle. \tag{32}$$

Step 2. If $g_k^* \geq -\varphi_\gamma(u,v,w^k)$, set $w^+ := w^k$ and **stop**; otherwise, compute

$$\tau_k := \min\left\{1, \frac{-g_k^*}{\|z^k - w^k\|^2}\right\}, \qquad w^{k+1} := w^k + \tau_k(z^k - w^k). \tag{33}$$

Step 3. Set $k \leftarrow k+1$, and go to **Step 1**.

implies $g_k^* < 0$. Thus, the stepsize τ_k given by (33) is computed using exact minimization, i.e., $0 < \tau_k := \arg\min_{\tau\in[0,1]} \psi(w^k + \tau(z^k - w^k))$. Since C is convex and z^k, $w^k \in C$, we have from (33) that $w^{k+1} \in C$, which implies that all points generated by *FW-Procedure* are in C. Moreover, (32) implies that $g_k^* = \langle w^k - v, z^k - w^k\rangle \leq \langle w^k - v, z - w^k\rangle$, for all $z \in C$. Hence, if the stopping criteria $g_k^* = \langle w^k - v, z^k - w^k\rangle \geq -\varphi_\gamma(u,v,w^k)$ in Step 2 of *FW-Procedure* is satisfied, then $\langle v - w^k, z - w^k\rangle \leq \varphi_\gamma(u,v,w^k)$, for all $z \in C$. Therefore, from Definition 3.1, we conclude that $w^+ = w^k \in \mathscr{P}_C(\varphi_\gamma, u, v)$, i.e., the output of *FW-Procedure*, is a feasible inexact projection of $v \in \mathbb{R}^n$ relative to $u \in C$. Finally, [29, Proposition A.2] implies that $\lim_{k\to+\infty} g_k^* = 0$. Thus, the stopping criteria $g_k^* \geq -\varphi_\gamma(u,v,w^k)$ in Step 2 of *FW-Procedure* is satisfied in a finite number of iterations if and only if $\varphi_\gamma(u,v,w^k) \neq 0$, for all $k = 0,1,\ldots$.

The following theorem is an import result about the convergence rate of the conditional gradient method applied to problem (31), which its proof can be found in [35]. For stating the theorem, we first note that

$$\psi(w) - \psi(w^*) \leq \frac{1}{2}\|w - w^*\|, \qquad \forall z \in C; \tag{34}$$

see also [20, Theorem 2.1.8].

Theorem 5.1. *Let $d_C := \max_{z,w\in C}\|z - w\|$ be the diameter of C. For $k \geq 1$, the iterate w^k of* FW-Procedure *satisfies $\psi(w^k) - \psi(w^*) \leq 8d_C^2/k$. Consequently, using* (34)*, we have $\|w^k - w^*\| \leq 4d_C/\sqrt{k}$, for all $k \geq 1$.*

5.2. Examples

Consider the problem

$$\min_{x\in C} f(x) := \|x\|_1, \tag{35}$$

where $C := \left\{x \in \mathbb{R}^n \colon x \geq 0 \text{ and } (x - \bar{x})^T Q(x - \bar{x}) \leq 1\right\}$ for a given vector $\bar{x} \in \mathbb{R}^n$ and a symmetric positive definite matrix $Q \in \mathbb{R}^{n\times n}$. Since the ℓ_1 norm tends to promote sparse solutions, we formulated instances of Problem (35) where there are vectors in C with only one non-null component. Thus we can verify the ability of SInexPD Algorithm to recover sparsity. Let us describe the main characteristics of the considered instances. Consider the spectral decomposition of Q given by

$$Q = \sum_{i=1}^{n} \lambda_i v^i (v^i)^T,$$

where $\lambda_1 \geq \ldots \geq \lambda_{n-1} > \lambda_n > 0$ are the given eigenvalues of Q and $\{v_1, v_2, \ldots, v_n\}$ is an orthonormal system of corresponding eigenvectors. We assume that there exists $u \in \mathbb{R}^n_{++}$ such that

$$v_n = u/\|u\|, \quad \lambda_n < 1/\|u\|^2, \qquad \text{and} \qquad \bar{x} = u + \xi e_n, \tag{36}$$

where $\xi \geq 1/\sqrt{\lambda_n}$ and $e_n \in \mathbb{R}^n$ is such that $e_n = [0, \ldots, 0, 1]^T$. We claim that $\tilde{x} := \xi e_n \in C$ and $0 \notin C$. Indeed, using (36) we have $\tilde{x} - \bar{x} = -\|u\| v_n$, which implies

$$(\tilde{x} - \bar{x})^T Q (\tilde{x} - \bar{x}) = \|u\|^2 v_n^T Q v_n = \|u\|^2 \lambda_n < 1,$$

concluding that $\tilde{x} \in C$. Now note that $0 \in C$ if and only if $\bar{x}^T Q \bar{x} \leq 1$. Since

$$\xi \geq \frac{1}{\sqrt{\lambda_n}} > -\langle u, e_n \rangle + \frac{1}{\sqrt{\lambda_n}} > \left(-\langle u, e_n \rangle + \sqrt{\langle u, e_n \rangle^2 - (\|u\|^2 - 1/\lambda_n)} \right) > 0$$

and $\|u + \xi e_n\|^2 = \xi^2 + 2\langle u, e_n \rangle \xi + \|u\|^2$, we have

$$\bar{x}^T Q \bar{x} \geq \lambda_n \|\bar{x}\|^2 = \lambda_n \|u + \xi e_n\|^2 > 1,$$

implying that $0 \notin C$.

For Problem (35), given $x \in \mathbb{R}^n$ we can get $s \in \partial f(x)$ by taking

$$[s]_i := \begin{cases} -1, & \text{if } [x]_i < 0 \\ 0, & \text{if } [x]_i = 0 \\ 1, & \text{if } [x]_i > 0, \end{cases}$$

where $[\cdot]_i$ stands for the i-th component of the corresponding vector. For computing the optimal solution z^k at Step 1 of the FW-Procedure, we use the software Algencan [36], an augmented Lagrangian code for general nonlinear optimization programming. We set $R = \|x^1 - x^0\|$ and $\delta_0 = \|s^0\|/2$ as suggested in [17] and [6], respectively. Our implementation uses the stopping criterion

$$\delta_\ell \leq 10^{-3}(1 + |f^k_{rec}|),$$

also suggested in [6]. Thus, in Algorithm 1, we have $\varepsilon_k = 0$ for all k. In our tests, we set $x^0 = \bar{x}$ and, for all k, $\gamma_1^k = 0.025$, $\gamma_2^k = 0.25$, $\gamma_3^k = 0.025$, and defined $\beta_k := 2(1 - 2\gamma_3^k)/(1 + 2\gamma_1^k) - 10^{-6}$ satisfying (12). Figure 1 shows the behavior of SInexPD Algorithm on a two-dimensional instance of Problem (35). The hatched region represents set C and only the iterates for which the target level was updated are plotted. As can be seen, the algorithm successfully found the *solution* for $\ell = 6$ iterations. We point out that the algorithm performed a total of $\ell = 14$ ($k = 189$) iterations until it met the stopping criterion. The highlight of the figure is that, before finding the solution, the iterates belong to the interior of set C. This is mostly due to the fact that SInexPD Algorithm performs inexact projections.

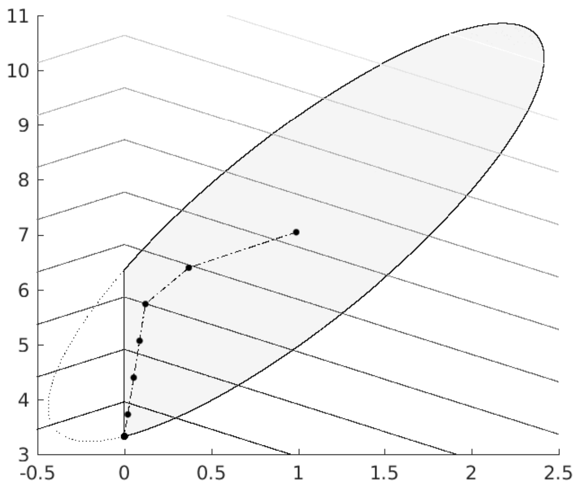

Figure 1. Behavior of SInexPD Algorithm on a two-dimensional instance of Problem (35).

Finally, we considered six instances of Problem (35) varying the dimension n. Without attempting to go into details, we mention that the problems were randomly generated such that $\lambda_n \in (10^{-2}, 10^{-6})$, $\lambda_i \in (10, 10^3)$ for $i = 1, \ldots, n-1$, vector $u \in \mathbb{R}^n_{++}$ in (36) is such that $\|u\| \in (0.8/\sqrt{\lambda_n}, 1/\sqrt{\lambda_n})$, and $\xi = 1\sqrt{\lambda_n}$. These imply that, with respect to the ellipsoid that makes up set C, the axis corresponding to the eigenvector v_n is *much larger* than the others ones. Moreover, the vectors of C that have only one non-null component are *far* from the center $\bar{x}$. These characteristics make problems more challenging for the algorithm. Table 1 shows the performance of SInexPD Algorithm. In the table, column "n" informs the considered dimension, "k" and "ℓ" are the number of iterations according to SInexPD Algorithm, "$\|x^k_{rec}\|_0$" is the number of non-null elements at the final iterate, "f^k_{rec}" and "δ_ℓ" are their corresponding values at the final iterate, and "LO" denotes the total number of linear minimization oracles required by the FW-Procedure to compute inexact projections. Note that the number of subgradient evaluations is equal to the number of iterations k plus one.

n	k	ℓ	$\|x^k_{rec}\|_0$	f^k_{rec}	δ_ℓ	LO
10	67	50	1	4.05D+01	2.47D-02	111
100	258	49	1	3.39D+01	1.95D-02	299
200	86	24	1	1.20D+01	6.91D-03	101
500	60	24	1	1.25D+01	1.09D-02	75
800	64	37	1	2.20D+01	1.38D-02	145
1000	62	38	1	2.54D+01	1.54D-02	114

Table 1. Performance of SInexPD Algorithm on six instances of Problem (35) varying the dimension.

As showed in Table 1, the algorithm found vectors with only one non-null component in all instances, showing its ability to recover sparsity in this class of problems. The required number of linear minimization oracles with respect to the number of iterations k is, on average, equal to 2.2, 1.3, 1.2, 1.3, 2.3, and 1.9 for each instance, respectively. This means that, in general, the FW-Procedure uses few iterations (typically, one or two) to compute a feasible inexact projection. Remembering that the table data corresponds to the values when the stop criterion was met, we reported that the *final* iterates were found with $\ell = 44, 40, 16, 13, 26$ and 27 iterations, respectively. We point out that, due to the inexact projections and mimicking the behavior of SInexPD Algorithm in the two-dimensional case, in each instance the iterates remained in the interior of C before the corresponding solution was found.

6. Conclusions

It is well known that the application of the subgradient method is only suitable for certain specific classes of non-differentiable convex optimization problems. However, this method is basic in the sense that it is the first step towards designing more efficient methods for solving that problems. Indeed, it is intrinsically related to cutting-plane and bundle methods; see [25]. These considerations lead us to conclude that the knowledge of new properties of the subgradient method has great theoretical value. In particular, our inexact version of the projected subgradient method will be useful in this theoretical context. Finally, one issue we believe deserves attention is the construction of inexact projected versions of cutting-plane and bundle methods. It is worth pointing out that if the method studied in [37] uses a suitable operator involving the normal cone of the set C, then the solutions of the proximal subproblems can be interpreted as approximate projections on the set C. This interesting feature deserves to be explored in the context of the subgradient method. Although the subproblem addressed in the present paper is quite different from that of [37], it is worthwhile to see if it is possible to establishes any kind of relationship between them. Finally, since the subgradient method uses an inner solver to compute inexact feasible projections, an interesting issue is to obtain a complexity bound in terms of the total number of inner iterations. In order to obtain such complexity bounds it is necessary to compute first a complexity bound involving the function φ_γ and the inexact projection condition in (5).

Funding

This work was supported by CNPq (Grants 305158/2014-7, 302473/2017-3, 424860/2018-0), FAPEG (Grants PRONEM- 201710267000532, PPP03/15-201810267001725) and CAPES.

References

[1] Ermol'ev YM. Methods of solution of nonlinear extremal problems. Cybernetics.1966;2(4):1–14.
[2] Shor NZ. Minimization methods for nondifferentiable functions. (Springer Series in Computational Mathematics; Vol. 3). Springer-Verlag, Berlin; 1985.
[3] Alber YI, Iusem AN, Solodov MV. On the projected subgradient method for nonsmooth convex optimization in a Hilbert space. Math Programming. 1998;81(1, Ser. A):23–35.
[4] Bello Cruz JY. A subgradient method for vector optimization problems. SIAM J Optim. 2013; 23(4):2169–2182.
[5] Bertsekas DP. Nonlinear programming. 2nd ed. Athena Scientific, Belmont, MA; 1999. Athena Scientific Optimization and Computation Series.
[6] Goffin JL, Kiwiel KC. Convergence of a simple subgradient level method. Math Program. 1999; 85(1, Ser. A):207–211.
[7] Kiwiel KC. Methods of descent for nondifferentiable optimization. (Lecture Notes in Mathematics; Vol. 1133). Springer-Verlag, Berlin; 1985.
[8] Nedić A, Bertsekas DP. The effect of deterministic noise in subgradient methods. Math Program. 2010;125(1, Ser. A):75–99.
[9] Villa S, Salzo S, Baldassarre L, et al. Accelerated and inexact forward-backward algorithms. SIAM J Optim. 2013;23(3):1607–1633.

[10] de Oliveira FR, Ferreira OP, Silva GN. Newton's method with feasible inexact projections for solving constrained generalized equations. Comput Optim Appl. 2019;72(1):159–177.
[11] Nocedal J, Wright SJ. Numerical optimization. 2nd ed. Springer, New York; 2006. Springer Series in Operations Research and Financial Engineering.
[12] Fukushima M, Luo ZQ, Tseng P. Smoothing functions for second-order-cone complementarity problems. SIAM J Optim. 2001/02;12(2):436–460.
[13] Gould NIM, Toint PL. Numerical methods for large-scale non-convex quadratic programming. In: Trends in industrial and applied mathematics (Amritsar, 2001). (Appl. Optim.; Vol. 72). Kluwer Acad. Publ., Dordrecht; 2002. p. 149–179.
[14] Vanderbei RJ. Linear programming: foundations and extensions. (International Series in Operations Research & Management Science; Vol. 4. Kluwer Academic Publishers, Boston, MA; 1996.
[15] Maingé PE. Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization. Set-Valued Anal. 2008;16(7-8):899–912.
[16] Correa R, Lemaréchal C. Convergence of some algorithms for convex minimization. Math Programming. 1993;62(2, Ser. B):261–275.
[17] Nedić A, Bertsekas DP. Incremental subgradient methods for nondifferentiable optimization. SIAM J Optim. 2001;12(1):109–138.
[18] Nedić A, Bertsekas D. Convergence rate of incremental subgradient algorithms. In: Stochastic optimization: algorithms and applications (Gainesville, FL, 2000). (Appl. Optim.; Vol. 54). Kluwer Acad. Publ., Dordrecht; 2001. p. 223–264.
[19] Wang XM. Subgradient algorithms on Riemannian manifolds of lower bounded curvatures. Optimization. 2018;67(1):179–194.
[20] Nesterov Y. Introductory lectures on convex optimization. (Applied Optimization; Vol. 87). Kluwer Academic Publishers, Boston, MA; 2004.
[21] Nesterov Y. Subgradient methods for huge-scale optimization problems. Math Program. 2014; 146(1-2, Ser. A):275–297.
[22] Polyak BT. Minimization of nonsmooth functionals. Ž Vyčisl Mat i Mat Fiz. 1969;9:509–521.
[23] Burachik R, Drummond LMG, Iusem AN, et al. Full convergence of the steepest descent method with inexact line searches. Optimization. 1995;32(2):137–146.
[24] Bertsekas DP, Nedic A, Ozdaglar AE. Convex analysis and optimization. Athena Scientific, Belmont, MA; 2003.
[25] Hiriart-Urruty JB, Lemaréchal C. Convex analysis and minimization algorithms. II. (Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]; Vol. 306).
[26] Díaz Millán R, Ferreira OP, Prudente LF. Alternating conditional gradient method for convex feasibility problems. arXiv e-prints. 2019 Dec;:arXiv:1912.04247.
[27] Polyak BT. Introduction to optimization. Optimization Software, Inc., Publications Division, New York; 1987. Translations Series in Mathematics and Engineering; translated from the Russian, With a foreword by Dimitri P. Bertsekas.
[28] Brannlund U. On relaxation methods for nonsmooth convex optimization. ProQuest LLC, Ann Arbor, MI; 1993. Thesis (Takn.dr)–Kungliga Tekniska Hogskolan (Sweden).
[29] Beck A, Teboulle M. A conditional gradient method with linear rate of convergence for solving convex linear systems. Math Methods Oper Res. 2004;59(2):235–247.
[30] Frank M, Wolfe P. An algorithm for quadratic programming. Naval Res Logist Quart.1956;3:95–110.
[31] Jaggi M. Revisiting frank-wolfe: Projection-free sparse convex optimization. Proceedings of the 30th International Conference on Machine Learning. 2013;28(1):427–435.
[32] Konnov IV. Simplified versions of the conditional gradient method. Optimization. 2018; 67(12):2275–2290.
[33] Lan G, Zhou Y. Conditional gradient sliding for convex optimization. SIAM J Optim. 2016; 26(2):1379–1409.
[34] Ravi SN, Collins MD, Singh V. A deterministic nonsmooth frank wolfe algorithm with coreset guarantees. INFORMS Journal on Optimization. 2019;1(2):120–142.
[35] Garber D, Hazan E. Faster rates for the frank-wolfe method over strongly-convex sets. Proceed-

ings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37.
[36] Birgin EG, Martínez JM. Practical augmented lagrangian methods for constrained optimization. SIAM; 2014.
[37] Solodov MV, Svaiter BF. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. Set-Valued Anal. 1999;7(4):323–345.