Noname manuscript No.
(will be inserted by the editor)

Weak subgradient algorithm for solving nonsmooth nonconvex unconstrained optimization problems

Gulcin Dinc Yalcin and Refail Kasimbeyli

Received: date / Accepted: date

Abstract This paper presents a weak subgradient based method for solving nonconvex unconstrained optimization problems. The method uses a weak subgradient of the objective function at a current point, to generate a new one at every iteration. The concept of the weak subgradient is based on the idea of using supporting cones to the graph of a function under consideration, which replaces in some sense, the supporting hyperplanes used for subgradient notion of convex analysis. Because of this reason, the method developed in this paper, does not require convexity assumption neither on the objective function nor on the set of feasible solutions. The new method is similar to subgradient methods of convex analysis and can be considered as a generalization of those methods. The paper investigates different stepsize parameters and provides convergence theorems for all cases. The significant difficulty of subgradient methods is, an estimation of subgradients at every iteration. In this paper, a method for estimating the weak subgradients, is also presented. The new method is tested on well-known test problems from the literature and computational results are reported and interpreted.

Keywords subgradient · weak subgradient · nonconvex optimization · nonsmooth optimization · nonlinear optimization · solution method

Mathematics Subject Classification (2000) 90C26 · 90C30 · 65K05 · 49M37

G. Dinc Yalcin and R. Kasimbeyli
Faculty of Engineering, Department of Industrial Engineering, Eskisehir Technical University, Iki Eylul Campus, Eskisehir 26555, Turkey
Tel.: +90-222-3213550
E-mail: gdinc@eskisehir.edu.tr; rkasimbeyli@eskisehir.edu.tr

1 Introduction

In this paper, we focus on the following box constrained optimization problem:

$$\min_{x \in \mathbb{X}} f(x), \tag{1}$$

where $f : \mathbb{X} \to \mathbb{R}$ is a real-valued function and $\mathbb{X} \subset \mathbb{R}^n$ is a nonempty and compact set of feasible solutions. The aim of this paper is to present a solution method for this problem, which is based on the weak subgradient concept.

Subgradient methods originated with works of Polyak [62] and Ermoliev [24], and further developed by Shor [68] in the 1970s, are practical techniques used to minimize a nondifferentiable convex function. Since then, many properties of subgradient methods have been investigated and different extensions have been proposed (see [16–18, 41, 42, 44, 57] and references therein).

One of the most popular extensions of subgradient methods, are approximate subgradient methods (also called $\varepsilon-$subgradient methods) developed for solving convex and quasi-convex problems (see [4, 23, 33, 45, 48, 58, 68]).

Versions of subgradient methods for solving stochastic problems where the convexity condition has been weakened, were studied in [25, 55, 60, 69].

Dual subgradient algorithms, which use subgradients of dual functions, were investigated in, e.g., [59] for convex, and in [28] for nonconvex constrained optimization problems. The method given in [28] was further extended in [13, 19, 29, 39].

In all above mentioned works, the subgradient notion of convex analysis is used. The idea of this subgradient is based on the use of a supporting hyperplane to the graph of convex (or concave dual) function at the point under investigation. Clearly, it should be very difficult to solve problems which do not satisfy the convexity (or concavity) assumption, by using this tool.

In this paper, we present a generalization of the subgradeint method, by using the so-called weak subgradients (instead of subgradients), whose definition is based on the idea of using a supporting conic surface (instead of supporting hyperplanes) to the graph of a function at the point under consideration. This notion was introduced by Azimov and Gasimov in [5] (see also [6]) where a special class of superlinear functions (instead of linear ones in convex analysis) with conical hypograph, is used. Due to this construction, the notion of weak subgradient does not require any kind of convexity in investigations of minimization problems. The idea of using supporting cones, has been used to characterize optimal solutions in a wide range of nonconvex optimization problems in last years (see, e.g., [34–38]), and is firstly used in this paper, to develop a method for solving unconstrained nonconvex optimization problems.

For solving nonsmooth nonconvex unconstrained optimization problems, different methods were proposed in the literature. Some examples are bundle-type methods (see, e.g., [27, 32, 43, 49, 56]), gradient sampling algorithm (see, e.g., [20, 46]), variable metric method (see, e.g., [52, 71]), trust region method (see, e.g., [1, 22, 66]), cutting planes (see, e.g., [26]). The general disadvantege of these methods is that, the computation of descent directions requires usually, solving (quadratic) subproblems. Therefore, they are not as easy as the implementation of subgradient methods which do not involve any subproblems either to find search directions or to compute stepsizes.

The weak subgradient based method presented in this paper, is both easy to implement and efficient for solving (smooth and) nonsmooth, (convex and) nonconvex unconstrained optimization problems. The weak subgradient is defined to be a pair $(v, c) \in \mathbb{R}^n \times \mathbb{R}_+$ such that, the graph of a superlinear function $\eta_{(v,c)}(x) = \langle v, x \rangle - c\|x\|$ supports the epigraph of the function under consideration, at a given point. That is, a weak subgradient is a pair, which consists of a vector component $v \in \mathbb{R}^n$ and a scalar component $c \in \mathbb{R}_+$. The method developed in this paper, generates a sequence $\{x_k\}_{k=1}^{\infty} \subset \mathbb{X}$ with

$$x_{k+1} = \mathcal{P}_{\mathbb{X}}[x_k - \alpha_k v_k],$$

where $\mathcal{P}_{\mathbb{X}}$ denotes the projection on the set $\mathbb{X}$, α_k is a positive stepsize, and v_k is a vector component of the weak subgradient, for which there exists a nonnegative number c_k such that (v_k, c_k) belongs to the *weak* subdifferential $\partial^w f(x_k)$ of f at x_k :

$$\partial^w f(x_k) = \{(v, c) : f(x) \geq f(x_k) + \langle v, x - x_k \rangle - c\|x - x_k\|, \forall x \in \mathbb{R}^n\}.$$

At every iteration, by using the current feasible solution x_k, a new weak subgradient (v_k, c_k) of f at x_k, is (approximately) computed, and by using a stepsize parameter α_k, a new feasible solution x_{k+1} is calculated. In this paper, we use three types of stepsize rules: a constant stepsize rule (where α_k is fixed at some positive value α), a diminishing stepsize rule (where $\alpha_k \to 0$), and a dynamic stepsize rule (where the value of α_k is calculated using an exact or approximate knowledge on the optimal objective function value). These stepsize rules, actually are standard rules used in subgradient methods (like the formula for calculating the new iteration) (see e.g. [15,33,44,45,57,58,64,65]).

In this paper, we not only analyze convergence properties of the proposed method for all of these stepsize rules, but also implement them on benchmark test problems from the literature and compare their perfomances. The crucial problem in the implementation of subgradient methods, is a computing of subgradients at every iteration. In this paper, an simple method for approximate computing of weak subgradients, is also proposed. This method makes use of the idea of discrete gradient method, introduced in [11] (see also [12]), which uses discrete gradients to approximate the Clarke subgradients. The presented method for computing of weak subgradients, uses the relation between the usual directional derivative and the weak subdifferential (see, [38, Theorem 1] or [37, Theorem 4.5]. The method is interpreted on demonstrative examples.

The performance of the weak subgradient method proposed in this paper, is demonstrated through a set of 71 experimental evaluations with benchmark test problems from the literature. It is remarkable that for some test problems, the proposed method improved the best known solutions, and demonstrated comparable performance to existing algorithms.

The rest of the paper is organized as follows. Section 2 recalls the definition and some significant theorems on weak subdifferentials and presents new relations for the weak subdifferentials and directional derivatives. All these features are used to approximate weak subgradients in next section. The method for approximation of weak subgradients, is given in Section 3. In this section, two illustrative examples are also provided. Section 4 presents the weak subgradient method and convergence properties for different stepsize rules. Section 5 gives computational results. In particular, we compare the performance of the weak subgradient method under the different types of stepsize rules. Finally, Section 6 draws some conclusions from the paper.

2 Weak Subdifferential

We begin this section by recalling the definition and some important properties of the weak subdifferential, as well as the well-known "*sup*" relation between the directional derivative and the weak subdifferential, from [5,6,37,38]. Then we will present new relations and properties which are used in future sections to evaluate weak subgradients approximately.

Definition 1 Let $f : \mathbb{X} \to R$ and $\overline{x} \in \mathbb{X}$. A pair $(v, c) \in \mathbb{R}^n \times \mathbb{R}_+$ is called a weak subgradient of f at $\overline{x}$ on $\mathbb{X}$ if

$$f(x) \geq f(\overline{x}) + \langle v, x - \overline{x} \rangle - c\|x - \overline{x}\|, \quad \forall x \in \mathbb{X}. \tag{2}$$

The set

$$\partial^w f(\overline{x}) = \{(v, c) \in \mathbb{R}^n \times \mathbb{R}_+ : f(x) \geq f(\overline{x}) + \langle v, x - \overline{x} \rangle - c\|x - \overline{x}\|, \quad \forall x \in \mathbb{X}\}$$

of all weak subgradients of f at $\overline{x}$ is called the weak subdifferential of f at $\overline{x}$ on $\mathbb{X}$.

Remark 1 Consider the function $\tilde{f} := f + \delta_{\mathbb{X}}$, where $\delta_{\mathbb{X}}$ is indicator function of $\mathbb{X}$ defined as $\delta_{\mathbb{X}}(x) = 0$ if $x \in \mathbb{X}$, and ∞ otherwise. Then $\partial^w f(\overline{x}) = \partial^w \tilde{f}(\overline{x})$ at any $\overline{x} \in \mathbb{X}$ and hence, in the definition of weak subgradients and the weak subdifferential, the relation (2) can be required for all $x \in \mathbb{R}^n$ instead of $x \in \mathbb{X}$, and consequently, in this case we can simply say that, pair $(v, c) \in \mathbb{R}^n \times \mathbb{R}_+$ is a weak subgradient of f at $\overline{x}$, not emphasizing the phrase "on $\mathbb{X}$", in Definition 1. Note also that $inf_{x \in \mathbb{X}} f(x) = inf_{x \in \mathbb{R}^n} \tilde{f}(x)$.

Remark 2 It is clear that, when f is subdifferentiable at $\overline{x}$ (in the sense of convex analysis) then f is also weakly subdifferentiable at $\overline{x}$, that is if $v \in \partial f(\overline{x})$, then by defnition, $(v, c) \in \partial^w f(\overline{x})$ for every $c \geq 0$. The well-known optimality condition of convex analysis says that, if f is a convex function subdifferentiable at $\overline{x}$, then f has a global minimum at $\overline{x}$ if and only if $0 \in \partial f(\overline{x})$. Similarly, it follows from the definition of the weak subdifferential that, if f is weakly subdifferentiable at $\overline{x}$, then f has a global minimum at $\overline{x}$ if and only if $(0_{\mathbb{R}^n}, 0_{\mathbb{R}}) \in \partial^w f(\overline{x})$.

Remark 3 It follows from Definition 1 that, the pair $(v, c) \in \mathbb{R}^n \times \mathbb{R}_+$ is a weak subgradient of f at $\overline{x} \in \mathbb{X}$, if there exists a continuous (superlinear) concave function

$$g(x) = f(\overline{x}) + \langle v, x - \overline{x} \rangle - c\|x - \overline{x}\|, \tag{3}$$

such that $g(x) \leq f(x)$, for all $x \in \mathbb{X}$ and $g(\overline{x}) = f(\overline{x})$. The set $hypo(g) = \{(x, \alpha) \in \mathbb{X} \times \mathbb{R} : g(x) \geq \alpha\}$ is a closed convex cone in $\mathbb{X} \times \mathbb{R}$ with vertex at $(\overline{x}, f(\overline{x}))$. Indeed,

$$hypo(g) - (\overline{x}, f(\overline{x})) = \{(x - \overline{x}, \alpha - f(\overline{x})) \in \mathbb{X} \times \mathbb{R} : \langle v, x - \overline{x} \rangle - c\|x - \overline{x}\| \geq \alpha - f(\overline{x})\}$$

$$= \{(u, \beta) \in \mathbb{X} \times \mathbb{R} : \langle v, u \rangle - c\|u\| \geq \beta\}$$

Thus, it follows from (2) and (3) that

$$graph(g)\{(x, \alpha) \in \mathbb{X} \times \mathbb{R} : g(x) = \alpha\}$$

is a conic surface which supports

$$epi(f) = \{(x, \alpha) \in \mathbb{X} \times \mathbb{R} : f(x) \leq \alpha\}$$

at the point $(\overline{x}, f(\overline{x}))$ in the sense that

$$epi(f) \subset epi(g), \quad cl(epi(f)) \cap graph(g) \neq \emptyset.$$

The above analysis shows that the class of weakly subdifferentiable functions is essentially larger than the class of subdifferentiable functions. Azimov and Gasimov [5,6] and Kasimbeyli and Mammadov [37] showed that the certain subclasses of lower (locally) Lipschitz functions and positively homogeneous functions are weakly subdifferentiable. To explain these classes of functions, we first recall the definiton of lower Lipschitz functions.

Definition 2 A funciton $f : \mathbb{X} \to (-\infty, +\infty]$ is called locally Lipschitz at $\overline{x} \in \mathbb{X}$, if there exists a non-negtive number L (Lpschitz constant), and a neighbourhood $N(\overline{x})$ of $\overline{x}$ such taht

$$f(x) - f(\overline{x}) \geq -L\|x - \overline{x}\|, \quad \forall x \in N(\overline{x}).$$

If above inequality holds true for all $x \in \mathbb{X}$ then f is called lower Lipschitz at $\overline{x}$ with Lipschitz constant L.

The following theorem explines some classes of the weakly subdifferentiable functions.

Theorem 1 *Let $f : \mathbb{X} \to (-\infty, +\infty]$ be a given function.*

(i) [5, Theorem 3.1] Let $f(\overline{x})$ be a lower semi-continuous at $\overline{x} \in \mathbb{X}$. Then f is weakly subdifferentiable at $\overline{x}$ if f is lower locally Lipschitz at $\overline{x}$ and there exists a point $\tilde{x}$ such that f is subdifferentiable there.

(ii) [5, Theorem 3.2] f is weakly subdifferentiable at $\overline{x}$ iff f is lower Lipschitz at $\overline{x}$.

(iii) *[5, Corollary 3.1] f is weakly subdifferentiable at $\overline{x}$ if f is lower locally Lipschitz at $\overline{x}$ and bounded below.*
(iv) *[6, Theorem 1] f is weakly subdifferentiable at $\overline{x}$ if f is lower locally Lipschitz at $\overline{x}$, and there exist numbers $p \geq 0$ and q such that $f(x) \geq -p\|x\|+q$, for all $x \in \mathbb{X}$.*
(v) *[37, Lemma 2.7] If f is a positively homogeneous function bounded from below on some neighborhood of $0_{\mathbb{R}^n}$, then f is weakly subdifferentiable at $0_{\mathbb{R}^n}$.*
(vi) *[37, Lemma 2.8] If f is a positively homogeneous continuous function, then f is weakly subdifferentiable at $0_{\mathbb{R}^n}$.*

The following theorem describes an important property of the weak subdifferential.

Theorem 2 *[37, Theorem 2.4] If the weak subdifferential $\partial^w f(\overline{x})$ of a function $f : X \to \mathbb{R}$ is not empty, then it is convex and closed.*

2.1 Directional Derivative and Weak Subdifferential

We begin this section by recalling the definition of the starshaped set.

Definition 3 [38] The set $S \subseteq \mathbb{R}^n$ is called starshaped at $\overline{x} \in S$, if for each $x \in S$ the following holds:

$$\lambda x + (1-\lambda)\overline{x} \in S \quad \text{for all } \lambda \in [0,1]. \tag{4}$$

The following lemma and the subsequent theorem are proved in [38].

Assumption 1 *Let $S \subseteq \mathbb{R}^n$ be starshaped at $\overline{x} \in S$, and let $f : S \to \mathbb{R}$ be a given function. Suppose that f has a directional derivative at $\overline{x}$ in every direction $x - \overline{x}$ with arbitrary $x \in S$ and*

$$f(x) - f(\overline{x}) \geq f'(\overline{x}; x - \overline{x}) \quad \textit{for all } x \in S - \{\overline{x}\}. \tag{5}$$

Lemma 1 *[38, Lemma 2] Let Assumption 1 hold. Then f has a global minimum at $\overline{x} \in S$ (over S), if and only if*

$$f'(\overline{x}; x - \overline{x}) \geq 0, \quad \textit{for all } x \in S - \{\overline{x}\}. \tag{6}$$

The following theorem gives the "*sup*" relation, where U denotes the unit sphere of $\mathbb{R}^n$, and $cone(S)$ denotes the cone generated by a set S.

Theorem 3 *[38, Theorem 1] Let Assumption 1 hold. In addition, let the directional derivative $f'(\overline{x};\cdot)$ of f at $\overline{x}$ be lower semi-continuous on $K = \operatorname{cone}(S - \{\overline{x}\}$ and let*

$$\inf\{f'(\overline{x};h):\ h \in K \cap U\} > -\infty. \tag{7}$$

Then f is weakly subdifferentiable at $\overline{x}$ on S; that is, $\partial_S^w f(\overline{x}) \neq \emptyset$ and

$$f'(\overline{x};h) = \sup\{\langle x^*, h\rangle - c\,\|h\| :\ (x^*, c) \in \partial_S^w f(\overline{x})\},\ \forall h \in K. \tag{8}$$

The following theorems explain some important properties of the weak subdifferential and the directional derivative, which play important roles in the remaining part of the paper.

The following theorem is probably the first generalization of the well-known "max-relation" of convex analysis to the nonconvex case for the directional derivative.

Theorem 4 *Let Assumption 1 hold. Then, clearly $0_{\mathbb{R}^n} \in int(dom(f'(\overline{x};\cdot)))$. Assume in addition that the directional derivative $f'(\overline{x};\cdot)$ of f at $\overline{x}$, is a convex function and subdifferentiable (in the sense of convex analysis) at $0_{\mathbb{R}^n} \in int(dom(f'(\overline{x};\cdot)))$. Then f is weakly subdifferentiable at $\overline{x}$, and there exists a positive number M such that*

$$f'(\overline{x};h) = \max\{\langle v, h\rangle - c\,\|h\| :\ (v,c) \in \partial_S^w f(\overline{x}), \|v\| + c \leq M\},\ \forall h \in \mathbb{R}^n. \tag{9}$$

Proof We denote

$$\phi(h) = f'(\overline{x};h),\ h \in \mathbb{R}^n. \tag{10}$$

Clearly $\phi(h)$ is positively homogenous and $\phi(\mathbf{0}) = 0$.

First we show that

$$\partial_S^w f(\overline{x}) = \partial_{S-\overline{x}}^w \phi(\mathbf{0}). \tag{11}$$

Let $(v, c) \in \partial_{S-\overline{x}}^w \phi(\mathbf{0})$. Then from (5) and (10) it follows that $(v, c) \in \partial_S^w f(\overline{x})$. If $(v, c) \in \partial_S^w f(\overline{x})$, then for any fixed $x \in S$ we have:

$$\begin{aligned}
\phi(x - \overline{x}) &= f'(\overline{x}; x - \overline{x}) \\
&= \lim_{\lambda\to 0^+} \frac{1}{\lambda}(f(\overline{x} + \lambda(x - \overline{x})) - f(\overline{x})) \\
&\geq \lim_{\lambda\to 0^+} \frac{1}{\lambda}(\langle v, \lambda(x - \overline{x})\rangle - c\|\lambda(x - \overline{x})\|) \\
&= \langle v, x - \overline{x}\rangle - c\|x - \overline{x}\|,
\end{aligned}$$

that leads to $(v, c) \in \partial_{S-\overline{x}}^w \phi(\mathbf{0})$.

By assumption, the directional derivative $f'(\overline{x};\cdot)$ of f at $\overline{x}$ is a convex function and subdifferentiable at $0_{\mathbb{R}^n}$ in the sense of convex analysis. In this case, $\partial\phi(\mathbf{0})$ is a compact set (see [67, Theorem 23.4, page 217]). Then it is clear that, there exists a positive number M_1 such that $\|v\| \leq M_1$ for every $v \in \partial\phi(\mathbf{0})$. On the other hand, for every $v \in \partial\phi(\mathbf{0})$, we have $(v, c) \in \partial_{S-\overline{x}}^w \phi(\mathbf{0})$ for arbitrary nonnegative number c, by Remark 2.

Now let L be a fixed positive number. Then, clearly

$$\phi(h) = max\{\langle v, h\rangle : v \in \partial\phi(\mathbf{0})\} \quad (12)$$
$$= max\{\langle v, h\rangle - c\,\|h\| : (v, c) \in \partial^w_{S-\overline{x}}\phi(\mathbf{0}), v \in \partial\phi(\mathbf{0}), 0 \le c \le L\} \quad (13)$$

for every $h \in \mathbb{R}^n$.

On the other hand, by Theorem 3 and by relations (11) - (13), we have:

$$\begin{aligned}\phi(h) &= \sup\{\langle v, h\rangle - c\,\|h\| :\ (v, c) \in \partial^w_S f(\overline{x})\}\\ &= sup\{\langle v, h\rangle - c\,\|h\| : (v, c) \in \partial^w_{S-\overline{x}}\phi(\mathbf{0})\}\\ &\ge max\{\langle v, h\rangle - c\,\|h\| : (v, c) \in \partial^w_{S-\overline{x}}\phi(\mathbf{0}), v \in \partial\phi(\mathbf{0}), 0 \le c \le L\}\\ &= max\{\langle v, h\rangle - c\,\|h\| : (v, c) \in \partial^w_S f(\overline{x}), \|v\| \le M_1, 0 \le c \le L\} = \phi(h)\end{aligned}$$

for every $h \in \mathbb{R}^n$, which leads to (9), where it can be set $M = M_1 + L$, and hence completes the proof. □

Remark 4 It is clear that, convexity of the directional derivative of some function, does not necessitate convexity of the function itself. Therefore, Theorem 4 is very important from point of view the representation of the directional derivative of some nonconvex function in the form of "max-relation", by using a compact subset of the weak subdifferential, which can be considered as a generalization of the well-known "max-relation" (12) of convex analysis.

Theorem 5 *Let f be a continuous and weakly subdifferentiable function at $\overline{x} \in \mathbb{R}^n$ and let $L > 0$ be a real number with*

$$\partial^w_L f(\overline{x}) = \{(v, c) \in \partial^w_S f(\overline{x}) : c \le L\} \ne \emptyset. \quad (14)$$

Then there exists a number $D > 0$ such that $\|v\| \le D$ for all $(v, c) \in \partial^w_L f(\overline{x})$, that is $\partial^w_L f(\overline{x})$ is a compact set.

Proof Let $(v, c) \in \partial^w f(\overline{x})$. Then by assumption $0 \le c \le L$. Let $\varepsilon > 0$ be an arbitrary positive number. Then due to the continuity of f at $\overline{x}$, there exists $\delta > 0$ such that

$$|f(x) - f(\overline{x})| < \varepsilon \quad if \quad \|x - \overline{x}\| \le \delta.$$

Consider

$$x(\delta, v) = \overline{x} + \frac{\delta v}{\|v\|}.$$

From the definiton of the weak subgradient, we have:

$$\begin{aligned}\varepsilon > f(\overline{x} + \frac{\delta v}{\|v\|}) - f(\overline{x}) &\ge \langle v, \overline{x} + \frac{\delta v}{\|v\|} - \overline{x}\rangle - c\|\overline{x} + \frac{\delta v}{\|v\|} - \overline{x}\|\\ &= \delta\|v\| - c\delta \Rightarrow \frac{\varepsilon}{\delta} + c > \|v\|.\end{aligned}$$

Since $c \le L$, this leads to

$$\frac{\varepsilon}{\delta} + L \ge \|v\|,$$

which together with the closedness property of the weak subdifferential, proves the theorem. □

Remark 5 If Assumption 1 hold and the directional derivative $f'(\overline{x};\cdot)$ of f at $\overline{x}$, is lower semi-continuous on $K = \operatorname{cone}(S - \{\overline{x}\}$, then by Theorem 3, f is weakly subdifferentiable at $\overline{x}$ and the "sup" relation (8) is satisfied. By Theorem 5, the set $\partial_L^w f(\overline{x})$ defined by (14), is a compact set for every $L > 0$ for which $\partial_L^w f(\overline{x}) \neq \emptyset$. Then for every $h \in K \cap U$, we have:

$$\sup\{\langle v, h\rangle - c\,\|h\| : \ (v,c) \in \partial_L^w f(\overline{x})\} = \max\{\langle v, h\rangle - c : \ (v,c) \in \partial_L^w f(\overline{x})\}. \tag{15}$$

Since the function $\ell(v,c) = \langle v, h\rangle - c$ is linear, and the set $\partial_L^w f(\overline{x})$ is convex (by Theorem 2) and compact, the "maximum" in (15) is attained at some extreme point (v_L, c_L) of $\partial_L^w f(\overline{x})$. By construction we have:

$$\lim_{L\to+\infty} \partial_L^w f(\overline{x}) = \partial_S^w f(\overline{x}).$$

Then clearly,

$$f'(\overline{x}; h) \geq \max\{\langle v, h\rangle - c\,\|h\| : \ (v,c) \in \partial_L^w f(\overline{x})\} \tag{16}$$

and the right-hand side of (16) can be used as an approximation for the directional derivative $f'(\overline{x}; h)$, for sufficiently large values of the parameter L for every $h \in K \cap U$.

□

Next, we present two illustrative examples for interpreting the weak subdifferential, and the sup and max relations.

Illustrative Example 1 *Let $f(x) = -|x|, x \in \mathbb{R}$. The directional derivative at $x = 0$ is:*

$$f'(0; h) = \lim_{\lambda\to+0} \frac{-|0 + \lambda h| - |0|}{\lambda} = \lim_{\lambda\to+0} \frac{-\lambda|h|}{\lambda} = -|h|.$$

Now calculate the weak subdifferential of f at $x = 0$. By the definition,

$$\partial^w f(0) = \{(v,c) \in \mathbb{R} \times \mathbb{R}^+ : vx - c|x| \leq -|x|, \forall x \in \mathbb{R}\}.$$

First, consider the following cases separately:

- *If $x \geq 0$, the inequality $vx - c|x| \leq -|x|$ becomes $vx - cx \leq -x$ and thus $v \leq c - 1$ is obained.*
- *If $x \leq 0$, then $vx - c|x| \leq -|x|$ becomes $vx + cx \leq x$, and thus $v \geq -c + 1$ is obtained.*

Hence, the weak subdifferential $\partial^w f(0)$ can be written as follows:

$$\partial^w f(0) = \{(v,c) \in \mathbb{R} \times \mathbb{R}^+ : -c + 1 \leq v \leq c - 1\}.$$

By Theorem 3, the sup relation can be written as follows:

$$-|h| = \sup\{vh - c|h| : -c + 1 \leq v \leq c - 1\}.$$

For $h = 1$ we obtain:

$$-1 = sup\{v - c : -c + 1 \leq v \leq c - 1\},$$

which shows that the sup relation is satisfied as a "maximum" along the line $v = c - 1$.
For $h = -1$ we have:

$$-1 = sup\{-v - c : -c + 1 \leq v \leq c - 1\},$$

which shows that the sup relation is satisfied as a "maximum" along the line $v = -c + 1$.

This example demonstrates that the class of functions, for which the sup relation is satisfied as a max relation, is sufficiently large.

The following example emphasizes the role of the condition (5) used in Assumption 1.

Counterexample 1 *Let $f(x) = |x^2 - 1|$. The directional derivative of f at $x = -1$ is:*

$$f'(-1; h) = \lim_{\lambda \to +0} \frac{|(-1 + \lambda h)^2 - 1| - |(-1)^2 - 1|}{\lambda} = 2|h|.$$

The weak subdifferential at $x = -1$ is:

$$\partial^w f(-1) = \{(v, c) \in \mathbb{R} \times \mathbb{R}^+ : v(x + 1) - c|x + 1| \leq |x^2 - 1|, \forall x \in \mathbb{R}\}.$$

- *If $x \leq -1$, the inequality $v(x + 1) - c|x + 1| \leq |x^2 - 1|$ becomes $v(x + 1) + c(x + 1) \leq (x^2 - 1)$ and thus implies $2 \geq -v - c$.*
- *If $x \geq 1$, the inequality $v(x + 1) - c|x + 1| \leq |x^2 - 1|$ becomes $v(x + 1) - c(x + 1) \leq (x^2 - 1)$ and thus implies $v - c \leq 0$.*
- *If $-1 \leq x \leq 1$, the inequality $v(x + 1) - c|x + 1| \leq |x^2 - 1|$ becomes $v(x + 1) - c(x + 1) \leq -(x^2 - 1)$ and thus $v - c \leq 0$ is obtained.*

Hence,

$$\partial^w f(0) = \{(v, c) \in \mathbb{R} \times \mathbb{R}^+ : 2 \geq -v - c \quad and \quad v - c \leq 0\}.$$

By Theorem 3, the sup relation can be written as follows:

$$2|h| = \sup\{vh - c|h| : 2 \geq -v - c \quad and \quad v - c \leq 0\}.$$

Let $h = -1$. Then we have:

$$2 = sup\{-v - c : 2 \geq -v - c \quad and \quad v - c \leq 0\},$$

which shows that the sup relation is satisfied at $h = -1$ as "maximum", along the line $v = -c - 2$.
Now, let $h = 1$. Then,

$$sup\{v - c : 2 \geq -v - c \quad and \quad v - c \leq 0\} = 0 < 2,$$

which shows that at $h = 1$ the sup relation is not satisfied. The reason is that the condition (5) is not satisfied for this function, at $x = -1$.

The properties presented in this section, will be used to develop a method for approximate computing weak subgradients, in the next section.

3 Approximation of Weak Subgradients

This section presents a method for numerical computing of weak subgradients at a given point. This method is based on the results and properties given in the previous section, and uses the idea of the discrete gradient method given by Bagirov [8].

We will assume that all conditions given in Assumption 1, are satisfied. The method presented in this section, is based on the exact or approximate representation of the directional derivative as "sup" or "max" relations given in Theorems 3 and 4 respectively, and also on Theorem 5 on the compactness of the subset $\partial_L^w f(\overline{x})$ of the weak subdifferential. The analysis given in Remark 5 shows that the directional derivative $f'(\overline{x}; h)$ can be approximated by using the "max" relation on the right-hand side of (16), and this approximation becomes more accurate if the value of L is as large as possible. Throughout in this section we will use the compact set $\partial_L^w f(\overline{x})$ as an approximation for the weak subdifferential $\partial^w f(\overline{x})$ for some positive number L, and assume that the directional derivative can (approximately) be represented using the above mentioned "max" relation.

Let $\overline{x} \in \mathbb{X}$, assume that $\partial_L^w f(\overline{x}) \neq \emptyset$, where L is sufficiently large positive number. Let $C = \{c \in \mathbb{R}_+ : (v, c) \in \partial_L^w f(\overline{x}) \text{ for some } v \in \mathbb{R}^n\}$ and let $\overline{c} \in C$. Then, the set $V_{\overline{c}} = \{v \in \mathbb{R}^n : (v, \overline{c}) \in \partial_L^w f(\overline{x})\}$ is not empty. Let $G = \{e = (e_1, e_2, ..., e_n) \in \mathbb{R}^n : |e_j| = 1, j = 1, \dots, n\}$ be a set of all vertices of the unit hypercube in $\mathbb{R}^n$, and for some $\alpha \in (0, 1]$ and $e = (e_1, e_2, ..., e_n) \in G$, consider the squence of n vectors

$$e^j(\alpha) = (\alpha e_1, \alpha^2 e_2, ..., \alpha^j e_j, 0, ..., 0), j = 1, \dots, n.$$

Remark 6 Note that, by using Theorems 4 and 5, and the definition (14) of the set $\partial_L^w f(\overline{x})$, the directional derivative for a certain class of functions can (approximately or exactly) be represented as the following "max" relation:

$$f'(\overline{x}; h) = \max\{\langle v, h\rangle - c\,\|h\| : \ (v, c) \in \partial_L^w f(\overline{x})\}. \tag{17}$$

By using this relation, the directional derivatives $f'(\overline{x}; e^j(\alpha)), j = 1, \dots n$ can be computed as follows:

$$f'(\overline{x}; e^j(\alpha)) = \max\{\langle v, e^j(\alpha)\rangle - c\left\|e^j(\alpha)\right\| : \ (v, c) \in \partial_L^w f(\overline{x})\}.$$

Assume that there exists a fixed value $c = \overline{c}$, and a set $A = \{a^1, a^2, \dots, a^m\} \subset V_{\overline{c}}$ such that

$$f'(\overline{x}; e^j(\alpha)) = \max\{\langle v, e^j(\alpha)\rangle - \overline{c}\left\|e^j(\alpha)\right\| : \ v \in A\}.$$

Now, for $e \in G$, we introduce the following sets:

$$R_0(e) = A,$$
$$R_j(e) = \{v = (v_1, \ldots, v_n) : v_j e_j = \max\{w_j e_j : w = (w_1, \ldots, w_n) \in R_{j-1}(e)\}\}$$

for all $j = 1, \ldots, n$. It is clear that

$$R_j(e) \neq \emptyset \quad \text{and } R_j(e) \subseteq R_{j-1}(e) \quad \text{for every } j = 1, \ldots, n, \tag{18}$$

and

$$v_r = u_r, \quad \forall u, v \in R_j(e), \quad r = 1, \ldots, j. \tag{19}$$

Consider the following set:

$$R(\overline{x}, e^j(\alpha)) = \{v \in A : \langle v, e^j(\alpha)\rangle = \max\{\langle w, e^j(\alpha)\rangle : w \in A\}\}. \tag{20}$$

By using this construction, we obtain that, for every $e^j(\alpha), j = 1, \ldots, n$, there exists an element $v \in R(\overline{x}, e^j(\alpha))$ such that

$$f'(\overline{x}; e^j(\alpha)) = \langle v, e^j(\alpha)\rangle - \overline{c} \left\| e^j(\alpha) \right\|. \tag{21}$$

□

In the sequel, we will introduce a method for approximately computing a weak subgradient $v, \overline{c}$ which satisfies the relation (21).

Proposition 1 *$R_n(e)$ is singleton set.*

Proof The proof follows immediately from (19). □

The following example explains the construction scheme for sets $R_j(e)$.

Example 1 Let $A = \{(-2.5, -1.7, -0.1, 1.7), (-2.5, -1.08, 3.09, 1.42),$
$(4.97, -4.33, 3.59, -3.53), (-2.5, -1.7, -0.1, 0.28), (-2.21, -4.67, 0.96, 0.28),$
$(3.63, 3.14, 2.84, -3.11), (1.75, -2.29, 0.19, 1.44)\}$ and $e = (-1, -1, 1, 1)$. Then,
$R_0 = A$,
$R_1 = \{v = (v_1, \ldots, v_n) \in R_0(e) : v_1 e_1 = \max\{(-2.5 \times -1), (-2.5 \times -1), (4.97 \times -1), (-2.5 \times -1), (-2.21 \times -1), (3.63 \times -1), (1.75 \times -1)\}$
$R_1 = \{(-2.5, -1.7, -0.1, 1.7), (-2.5, -1.08, 3.09, 1.42), (-2.5, -1.7, -0.1, 0.28)\}$
$R_2 = \{v = (v_1, \ldots, v_n) \in R_1(e) : v_2 e_2 = \max\{(-1.7 \times -1), (-1.08 \times -1), (-1.7 \times -1)\}$
$R_2 = \{(-2.5, -1.7, -0.1, 1.7), (-2.5, -1.7, -0.1, 0.28)\}$
$R_3 = \{v = (v_1, \ldots, v_n) \in R_2(e) : v_3 e_3 = \max\{(-0.1 \times 1), (-0.1 \times 1)\}$
$R_3 = \{(-2.5, -1.7, -0.1, 1.7), (-2.5, -1.7, -0.1, 0.28)\}$
$R_4 = \{v = (v_1, \ldots, v_n) \in R_3(e) : v_4 e_4 = \max\{(1.7 \times 1), (0.28 \times 1)\}$
$R_4 = \{(-2.5, -1.7, -0.1, 1.7)\}$

Remark 7 Let $a \in A$ be such that $a \notin R_n(e)$. Then there exists $r \in \{1, ..., n\}$ such that, $a \in R_t(e)$ for all $t = 0, ..., r-1$ and $a \notin R_r(e)$. It follows from $a \notin R_r(e)$ that $v_r e_r > a_r e_r$ for every $v \in R_r(e)$. Let $v \in R_r(e)$, and define the number

$$\overline{d} = \min\{d(a) = v_r e_r - a_r e_r : a \in A \setminus R_n(e)\} \tag{22}$$

Since A is a finite set and $d(a) > 0$ for all $a \in A \setminus R_n(e)$, it follows that $\overline{d} > 0$. Since $\partial^w_L f(\overline{x})$ is a compact set by Theorem 5, there exist a positive number D such that

$$\|v\| \leq D, \overline{c} \leq L$$

for every $(v, \overline{c}) \in \partial^w f(\overline{x})$. We take any $j \in \{1, ..., n\}$ with $r < j$. Then for every $v, w \in A$, we obtain:

$$\begin{aligned}
|\sum_{t=r+1}^{j} (v_t - w_t)\alpha^{t-r} e_t| &\leq \sum_{t=r+1}^{j} \alpha^{t-r} |(v_t - w_t)||e_t| \\
&\leq \alpha(\sum_{t=r+1}^{j} |v_t| + \sum_{t=r+1}^{j} |w_t|) \\
&\leq \alpha(\|v\| + \|w\|) \leq 2D\alpha.
\end{aligned}$$

Let $\alpha_0 = min\{1, \overline{d}/(4D)\}$. Then, for any $\alpha \in (0, \alpha_0]$ we have:

$$|\sum_{t=r+1}^{j} (v_t - w_t)\alpha^{t-r} e_t| \leq \frac{\overline{d}}{2} \tag{23}$$

Proposition 2 *Let $\overline{x} \in \mathbb{X}$, then there exists a number $0 < \alpha_0 \leq 1$ such that $R(\overline{x}, e^j(\alpha)) \subset R_j(e)$ for all $\alpha \in (0, \alpha_0]$ and $j = 1, \ldots, n$.*

Proof Assume the contrary. Then there exists $y \in R(\overline{x}, e^j(\alpha))$ such that $y \notin R_j(e)$ for some j. Consequently there exists $r \in \{1, ..., n\}$ and $r \leq j$ such that $y \notin R_r(e)$ and $y \in R_t(e)$ for any $t = 0, ..., r-1$. We take any $v \in R_j(e)$. From (19) and (22) we have $v_t e_t = y_t e_t, t = 1, ..., r-1, v_r e_r \geq y_r e_r + \overline{d}$. It follows from (23) that

$$\begin{aligned}
&\langle v, e^j(\alpha)\rangle - \langle y, e^j(\alpha)\rangle \\
&\qquad = \sum_{t=1}^{j} (v_t - y_t)\alpha^t e_t \\
&\qquad = \alpha^r [v_r e_r - y_r e_r + \sum_{t=r+1}^{j} (v_t - y_t)\alpha^{t-r} e_t] \geq \alpha^r [\overline{d} + (-\frac{\overline{d}}{2})] = \alpha^r \frac{\overline{d}}{2}.
\end{aligned}$$

Therefore,

$$\langle v, e^j(\alpha)\rangle \geq \langle y, e^j(\alpha)\rangle + \alpha^r \frac{\overline{d}}{2}.$$

Since $\langle y, e^j(\alpha)\rangle = max\{\langle u, e^j(\alpha)\rangle : u \in A\}$ and $v \in A$, we get

$$\langle y, e^j(\alpha)\rangle \geq \langle v, e^j(\alpha)\rangle \geq \langle y, e^j(\alpha)\rangle + \alpha^r \frac{\overline{d}}{2}.$$

Now it follows from $\alpha^r > 0$ and $\overline{d} > 0$ that, the last relation is a contradiction. □

Corollary 1 *There exist $\alpha_0 > 0$ and $v \in R_j(e)$ such that*

$$f'(\overline{x}; e^j(\alpha)) = f'(\overline{x}; e^{j-1}(\alpha)) + v_j\alpha^j e_j - \overline{c}\alpha^j$$

for all $\alpha \in (0, \alpha_0]$ and for every $j = 1, \ldots, n$.

Proof Proposition 2 implies that $R(x, e^j(\alpha)) \subseteq R_j(e), j = 1, \ldots, n$. Then, as it was explained in Remark 6, there exist $v \in R_j(e), v^0 \in R_{j-1}(e)$ such that

$$f'(\overline{x}; e^j(\alpha)) - f'(\overline{x}; e^{j-1}(\alpha)) = \langle v, e^j(\alpha)\rangle - \overline{c}\|e^j(\alpha)\| - \langle v^0, e^{j-1}(\alpha)\rangle - \overline{c}\|e^{j-1}(\alpha)\|$$

By using the ℓ_1 norm for simplicity, the norm terms in the last relation can be simplified as follows:

$$\begin{aligned}
\|e^j(\alpha)\|_1 &= \alpha|e_1| + \alpha^2|e_2| + ... + \alpha^{j-1}|e_j - 1| + \alpha^j|e_j| + 0 + ... + 0 \\
&= \alpha + \alpha^2 + ... + \alpha^j, \\
\|e^{j-1}(\alpha)\|_1 &= \alpha|e_1| + \alpha^2|e_2| + ... + \alpha^{j-1}|e_{j-1}| \\
&= \alpha + \alpha^2 + ... + \alpha^{j-1}.
\end{aligned}$$

Since $v_i = v_i^0$ for every $i = 1, ..., j-1$, we obtain:

$$f'(x; e^j(\alpha)) - f'(x; e^{j-1}(\alpha)) = v_j\alpha^j e_j - \overline{c}\alpha^j$$

□

Let $e \in G$, and let $\lambda > 0, \alpha > 0$ be given numbers. Consider the following points:

$$x^0 = \overline{x}, \quad x^j = x^0 + \lambda e^j(\alpha), \quad j = 1, \ldots, n.$$

Then obviously $x^j = x^{j-1} + (0, \ldots, 0, \lambda\alpha^j e_j, 0, \ldots)$ for every $j = 1, \ldots, n$. Define the vector $v(e, \alpha, \lambda) \in \mathbb{R}^n$ with the coordinates:

$$v_j(e, \alpha, \lambda) = \frac{f(x^j) - f(x^{j-1})}{\lambda\alpha^j e_j} + \frac{\overline{c}}{e_j}, \quad j = 1, \ldots, n.$$

For any fixed $e \in G$ and $\alpha > 0$, we introduce the set:

$$W(e, \alpha) = \{(w, \overline{c}) \in \mathbb{R}^n \times C : \exists(\lambda_k \to +0, k \to +\infty), w = \lim_{k\to\infty} v(e, \alpha, \lambda_k)\}$$

Proposition 3 *There exists $\alpha_0 > 0$ such that $W(e, \alpha) \subset \partial^w f(\overline{x}) \quad \forall\alpha \in (0, \alpha_0]$*

Proof It follows from the definition of the vector $v = v(e, \alpha, \lambda)$ that

$$v_j(e, \alpha, \lambda) = \frac{f(x^j) - f(x^{j-1})}{\lambda \alpha^j e_j} + \frac{\overline{c}}{e_j}$$
$$= \frac{[f(x^j) - f(\overline{x})] - [f(x^{j-1}) - f(\overline{x})]}{\lambda \alpha^j e_j} + \frac{\overline{c}}{e_j}.$$

Since $f'(\overline{x}, e^j(\alpha)) = \lim_{\lambda \to +0} \frac{f(x^j) - f(\overline{x})}{\lambda}$, we have:

$$v_j(e, \alpha, \lambda) = \frac{\lambda f'(\overline{x}, e^j(\alpha)) - \lambda f'(\overline{x}, e^{j-1}(\alpha)) + o(\lambda, e^j(\alpha)) - o(\lambda, e^{j-1}(\alpha))}{\lambda \alpha^j e_j} + \frac{\overline{c}}{e_j},$$

where $\lambda^{-1} o(\lambda, e^i) \to 0, \lambda \to +0, i = j-1, j$.

Now let $w \in R_n(e)$. By Proposition 1, w is unique. This implies by Proposition 2 that $R_n(e) = R(x, e^n(\alpha))$. The inclusion $w \in R_n(e)$ implies that $w \in R_j(e)$ for all $j \in 1, ..., n$. Then it follows from Corollary 1 that there exists $\alpha_0 > 0$ such that

$$v_j(e, \alpha, \lambda) = \frac{\lambda(w_j \alpha^j e_j - \overline{c} \alpha^j) + o(\lambda, e^j(\alpha)) - o(\lambda, e^{j-1}(\alpha))}{\lambda \alpha^j e_j} + \frac{\overline{c}}{e_j}$$
$$= w_j - \frac{\overline{c}}{e_j} + \frac{o(\lambda, e^j(\alpha)) - o(\lambda, e^{j-1}(\alpha))}{\lambda \alpha^j e_j} + \frac{\overline{c}}{e_j}$$

for all $\alpha \in (0, \alpha_0]$.

Then, for any fixed $\alpha \in (0, \alpha_0]$, we have

$$\lim \|v_j(e, \alpha, \lambda) - w_j\| = 0.$$

Consequently, $\lim_{\lambda \to +0} v(e, \alpha, \lambda) = w \in V_{\overline{c}}$, and the proof is completed.

□

Now we are ready to formulate the algorithm for approximate computing a weak subgradient (v, c) of a given function f at the point $\overline{x} \in \mathbb{X}$.

Algorithm 1 Approximate computing of the weak subgradient $(v, c) \in \partial^w f(\overline{x})$.

1: Let $e \in G = \{e = (e_1, e_2, ..., e_n) \in \mathbb{R}^n : |e_j| = 1, j = 1, \ldots, n\}$ and $\lambda > 0, \alpha \in (0, 1]$, $\overline{x} \in \mathbb{X}$.
2: Define $e^j(\alpha) = (e_1 \alpha, e_2 \alpha^2, ..., e_j \alpha^j, 0, ..., 0), j = 1, \ldots, n$.
3: Choose a number $c > 0$.
4: Let $x^0 = \overline{x}$.
5: $j \leftarrow 1$.
6: **while** $j \leq n$ **do**
7: $\quad x^j = x^0 + \lambda e^j(\alpha)$,
8: $\quad v_j = \frac{f(x^j) - f(x^{j-1})}{\lambda \alpha^j e_j} + \frac{c}{e_j}$,
9: $\quad j \leftarrow j + 1$.
10: **end while**

The next example gives an illustration of Algorithm 1, on the so-called spiral function from [53].

Example 2 Let

$$f(x_1,x_2) = \max\Big\{ \Big(x_1 - \sqrt{x_1^2+x_2^2}\cos\sqrt{x_1^2+x_2^2}\Big)^2 + 0.005(x_1^2+x_2^2),$$
$$\Big(x_2 - \sqrt{x_1^2+x_2^2}\sin\sqrt{x_1^2+x_2^2}\Big)^2 + 0.005(x_1^2+x_2^2)\Big\}.$$

We apply Algorithm 1 to compute a weak subgradient of this function at the point $\overline{x} = (2,0)$.

1: $e = (1,-1)$ and $\lambda = 0.1, \alpha = 0.9, \overline{x} = (2,0)$.
2: $e^1(\alpha) = (0.9,0), e^2(\alpha) = (0.9,-0.81)$
3: $c = 10$
4: $x^0 = (2,0)$
5: $j \leftarrow 1$
6: $x^1 = (2,0) + 0.1 \times (0.9,0) = (2.09,0)$,
7: $v_1 = \frac{f(x^1)-f(x^0)}{0.1\times 0.9^1\times 1} + \frac{10}{1} = \frac{(9.8002-8.04189)}{0.09} + 10 = 29.5367$
8: $j \leftarrow 2$
9: $x^2 = (2,0) + 0.1 \times (0.9,-0.81) = (2.09,-0.081)$,
10: $v_2 = \frac{f(x^2)-f(x^1)}{0.1\times 0.9^2\times -1} + \frac{10}{-1} = \frac{(9.8229-9.8002)}{-0.081} - 10 = -10.2806$

Hence, we conclude that $(v,c) = ((29.5367,-10.2806),10) \in \partial^w f(2,0)$. Geometrically this means that the hypograph of the function $g(x_1,x_2) = 8.04189 + 29.5367(x_1-2) - 10.2806(x_2-0) - 10(|x_1-2|+|x_2-0|)$, where $8.04189 = f(2,0)$, is a supporting cone to the epigraph of f at the point $(\overline{x}, f(\overline{x}))$. Figure 1 illustrates the graph of the spiral function f and the supporting cone $hypo(g)$ at the point $((2,0), f(2,0))$.

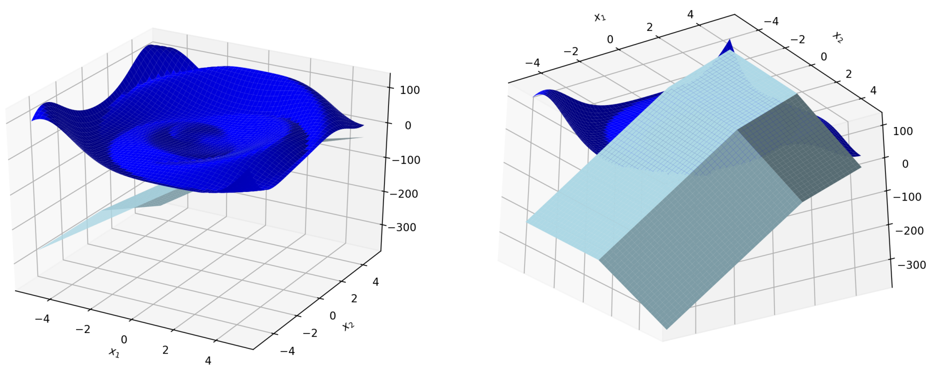

Fig. 1 View of the graph of spiral function (blue) and the supporting cone (light blue) from right and bottom, respectively.

4 Weak subgradient method for nonsmooth nonconvex unconstraint optimization

The weak subgradient algorithm generates the sequence $\{x_k\}_{k=1}^{\infty} \subset \mathbb{X}$, given by

$$x_{k+1} = \mathcal{P}_{\mathbb{X}}[x_k - \alpha_k v_k], \quad (v_k, c_k) \in \partial^w f(x_k),$$

where $\mathcal{P}_{\mathbb{X}}$ denotes the projection on the feasible set

$$\mathbb{X} = \{x = (x_1, \ldots, x_n) : a_j \leq x_j \leq b_j, \quad j = 1, \ldots, n\}$$

and α_k is a positive stepsize.

The notations and assumptions used in this section, are listed below.

- The weak subgradient method, uses weak subgradients of the objective function, calculated at every iteration k, using Algorithm 1. This algorithm is based on the approximate representation $\partial_L^w f(x_k)$ of the weak subdifferential, which assumes that there exist positive numbers D and L such that

$$\|v_k\| \leq D, \tag{24}$$

$$c_k \leq L, \tag{25}$$

  for every $(v_k, c_k) \in \partial^w f(x_k)$ for every $x_k \in \mathbb{X}$.
- x^* and $f^* = f(x^*)$ denote the optimal solution and the optimal value of problem (1), respectively.
- $d_{\mathbb{X}} = diam(\mathbb{X}) = \max_{x_1, x_2 \in \mathbb{X}} \|x_1 - x_2\|$ denotes the diameter of $\mathbb{X}$. Then clearly

$$\|x_k - x^*\| \leq d_{\mathbb{X}}, \tag{26}$$

  where $\|\cdot\|$ denotes the standard Euclidean norm.

The following lemma gives a general estimate that holds for every positive stepsize and will be used in subsequent convergence analysis. Note that, although the convergence theorems given for different versions of subgradient methods in the literaturte are similar, all they have specific properties which depend on the basic philosphy used for each approach that makes them different, see e.g. [2,28,33,44,45,57,58,63].

Lemma 2 *Let $\{x_k\}$ be the sequence generated by the weak subgradient method. Then for all $k \geq 0$, we have*

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 2\alpha_k[f(x_k) - f^* - c_k\|x^* - x_k\|] + \alpha_k^2\|v_k\|^2.$$

Proof Using the nonexpansion ptoperty of the projection, and the weak subgradient inequality

$$\langle v_k, x_k - x^* \rangle \geq f(x_k) - f(x^*) - c_k\|x^* - x_k\|$$

Table 1 Step size parameters

Step Size Name	**Step Size**
Constant stepsize	$\alpha_k \in \mathbb{R}_+$
Diminishing stepsize	$\lim_{k\to\infty} \alpha_k = 0$ $\Sigma_{k=0}^{\infty}\alpha_k = \infty.$
Dynamic stepsize rule for unknown f^* with $f^{lev} > f^*$	$\alpha_k = \gamma_k \frac{f(x_k)-f^{lev}-c_k d_{\mathbb{X}}}{\lVert v_k\rVert^2}$ $0 < \underline{\gamma} < \gamma_k \leq \overline{\gamma} < 2$
Dynamic stepsize rule for unknown f^* with $f^{lev} < f^*$	$\alpha_k = \gamma_k \frac{f(x_k)-f^{lev}-c_k d_{\mathbb{X}}}{\lVert v_k\rVert^2}$ $0 < \underline{\gamma} < \gamma_k \leq \overline{\gamma} < 1$
Dynamic stepsize rule with dynamic f_k^{lev}	$\alpha_k = \gamma_k \frac{f(x_k)-f_k^{lev}-c_k d_{\mathbb{X}}}{\lVert v_k\rVert^2}$ $0 < \underline{\gamma} < \gamma_k \leq \overline{\gamma} < 1$ where $f_k^{lev} = \min_k f(x_k) - \delta_k$, $\delta_{k+1} = \begin{cases} \min\{\beta_1\delta_k, \overline{\delta}\} & if \quad f(x_{k+1}) < f_k^{lev} \\ \max\{\beta_2\delta_k, \underline{\delta}\} & if \quad f(x_{k+1}) > f_k^{lev} \end{cases}$ and $1 < \beta_1, 0 < \beta_2 < 1$ and $\delta_1, \overline{\delta}, \underline{\delta}$ are positive constants

for every $(v_k, c_k) \in \partial^w f(x_k)$, we obtain:

$$\begin{aligned}\|x_{k+1} - x^*\|^2 &= \|\mathcal{P}_{\mathbb{X}}[x_k - \alpha_k v_k] - x^*\|^2 \leq ||x_k - \alpha_k v_k - x^*||^2 \\ &= \|x_k - x^*\|^2 - 2\alpha_k\langle v_k, x_k - x^*\rangle + \alpha_k^2\|v_k\|^2 \\ &\leq \|x_k - x^*\|^2 - 2\alpha_k[f(x_k) - f^* - c_k\|x^* - x_k\|] + \alpha_k^2\|v_k\|^2.\end{aligned}$$

The proof is completed. □

In this paper, we analyze the convergence propoerties and computational results for the weak subgradient algorithm, using the five different types of stepsize rules, which are described in Table 1.

Next, we present a pseudocode of the Weak Subgradient Algorithm.

4.1 Convergence analysis for the constant stepsize.

We start the convergence analysis, with the constant stepsize rule.

Proposition 4 *For the sequence $\{x_k\}$ generated by the weak subgradient method with a constant stepsize α, we have:*

(a) if $f^* = -\infty$ then

$$\liminf_{k\to\infty}(f(x_k)) = -\infty$$

(b) if $f^* > -\infty$ then

$$\liminf_{k\to\infty}(f(x_k)) \leq f^* + \alpha\frac{D^2}{2} + \liminf_{k\to\infty} c_k d_{\mathbb{X}}. \tag{27}$$

Algorithm 2 Weak Subgradient Algortihm

Select a starting initial solution $x_0 \in \mathbb{X}$, let the upper bound be $UB = f(x_0)$, and the incumbent solution be $x^{best} = x_0$.
Define the Iteration Number, and let $k \leftarrow 0$
Select a positive stepsize parameter α_k.
while $k \leq IterationNumber$ **do**
    Apply Algorithm 1 to compute a weak subgradient $(v_k, c_k) \in \partial^w f(x_k)$ of f at x_k.
    Update stepsize parameter α_k if needed
    $x_{k+1} = \mathcal{P}_{\mathbb{X}}[x_k - \alpha_k v_k]$.
    if $f(x_{k+1}) < UB$ **then**
        $UB = f(x_{k+1})$
        $x^{best} = x_{k+1}$.
    end if
    $k \leftarrow k + 1$
end while

Proof (a) and (b) are proven simultaneously. If the result does not hold, there must exist an $\varepsilon > 0$ such that

$$\liminf_{k\to\infty}(f(x_k)) > f^* + \alpha\frac{D^2}{2} + \liminf_{k\to\infty} c_k d_{\mathbb{X}} + \varepsilon,$$

and let k_0 be large enough so that for all $k \geq k_0$ we have

$$f(x_k) - f^* - c_k d_{\mathbb{X}} > \alpha\frac{D^2}{2} + \varepsilon. \tag{28}$$

By using Lemma 2, and the relations (24)-(28), we obtain:

$$\begin{aligned}
\|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 - 2\alpha[f(x_k) - f^* - c_k\|x^* - x_k\|] + \alpha^2\|v_k\|^2 \\
&\leq \|x_k - x^*\|^2 - 2\alpha[f(x_k) - f^* - c_k d_{\mathbb{X}}] + \alpha^2\|v_k\|^2 \\
&\leq \|x_k - x^*\|^2 - 2\alpha\big[\alpha\frac{D^2}{2} + \varepsilon\big] + \alpha^2\|v_k\|^2 \\
&= \|x_k - x^*\|^2 - \alpha^2 D^2 - 2\alpha\varepsilon + \alpha^2 D^2 \\
&= \|x_k - x^*\|^2 - 2\alpha\varepsilon \leq \|x_{k-1} - x^*\|^2 - 4\alpha\varepsilon \\
&\leq \cdots \leq \|x_{k_0} - x^*\|^2 - 2(k + 1 - k_0)\alpha\varepsilon,
\end{aligned}$$

which cannot hold for k sufficiently large, so it is a contradiction. □

4.2 Convergence analysis for the diminishing stepsize

Proposition 5 *Let the stepsize α_k be such that*

$$\lim_{k\to\infty}\alpha_k = 0 \quad \Sigma_{k=0}^{\infty}\alpha_k = \infty.$$

Then for sequence $\{x_k\}$ generated by the weak subgradient method with the diminishing stepsize α_k, we have:

$$\liminf_{k\to\infty} f(x_k) \leq f^* + \liminf_{k\to\infty} c_k d_{\mathbb{X}} \tag{29}$$

Proof If the result does not hold, there must exist an $\varepsilon > 0$ such that

$$\liminf_{k\to\infty} f(x_k) > f^* + \liminf_{k\to\infty} c_k d_{\mathbb{X}} + \varepsilon.$$

Let k_0 be large enough so that for all $k \geq k_0$ we have

$$f(x_k) - f^* - c_k d_{\mathbb{X}} > \varepsilon. \tag{30}$$

Again by using Lemma 2 and relations (26) and (30), in a similar way used in the proof of Proposition 4, we obtain:

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 2\alpha_k\varepsilon + \alpha_k(\alpha_k\|v_k\|^2).$$

Because $\alpha_k \to 0$, and $\{v_k\}$ is bounded by (24), without loss of generality we may assume that k_0 is large enough so that $\varepsilon > \alpha_k\|v_k\|^2$ for all $k \geq k_0$, implying that

$$\begin{aligned}\|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 - 2\alpha_k\varepsilon + \alpha_k\varepsilon \\ &= \|x_k - x^*\|^2 - \alpha_k\varepsilon \leq \|x_{k-1} - x^*\|^2 - (\alpha_{k-1} + \alpha_k)\varepsilon \\ &\leq \cdots \leq \|x_{k_0} - x^*\|^2 - \varepsilon\Sigma_{j=k_0}^{k}\alpha_j.\end{aligned}$$

Since $\Sigma_{k=0}^{\infty}\alpha_k = \infty$, this relation cannot hold for k sufficient large, so it is a contradiction. □

4.3 Convergence analysis for the dynamic stepsize for unknown f^*, with $f^{lev} > f^*$.

This stepsize rule uses an estimate f^{lev} of f^*, often referred to as target level. This value f^{lev} may be greater or smaller than f^*. First, we investigate the case $f^{lev} > f^*$ in this subsection.

The dynamic stepsize rule for unknown f^* with $f^{lev} > f^*$, is defined in the following form:

$$\alpha_k = \gamma_k \frac{f(x_k) - f^{lev} - c_k d_{\mathbb{X}}}{\|v_k\|^2}, \tag{31}$$

where $0 < \underline{\gamma} \leq \gamma_k \leq \overline{\gamma} < 2$.

As a matter of course, the value of c_k must be selected less than $\frac{f(x_k)-f^{lev}}{d_{\mathbb{X}}}$, to ensure positiveness of the stepsize. The following proposition shows that the distance between an optimal solution and the x_k generated by the weak subgradient method using the dynamic stepsize for unknown f^* with $f^{lev} > f^*$, decreases at each iteration.

Proposition 6 *Assume that $f(x_k) > f^{lev} + c_k d_{\mathbb{X}}$ for every $\{x_k\}$ generated by the weak subgradient method using the dynamic stepsize rule* (31) *for unknown f^* with $f^{lev} > f^*$. Then we have:*

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2.$$

Proof By Lemma 2 and by (26), we have:

$$\begin{aligned}\|x_{k+1}-x^*\|^2 &\le \|x_k-x^*\|^2 - 2\alpha_k(f(x_k)-f^*-c_k\|x^*-x_k\|)+\alpha_k^2\|v_k\|^2\\ &\le \|x_k-x^*\|^2 - 2\alpha_k(f(x_k)-f^{lev}-c_kd_{\mathbb{X}})+\alpha_k^2\|v_k\|^2.\end{aligned}$$

It is straightforward to verify that for the range of the stepsize α_k defined by (31), the sum of the last two terms in the above relation is negative. □

The following convergence result for the sequence of values $\{f(x_k)\}$ generated using the dynamic stepsize rule (31) for unknown f^* with $f^{lev} > f^*$ is similar to the convergence result obtained for diminishing stepsize rule given in Proposition 5.

Proposition 7 *For the sequence $\{x_k\}$ generated by the weak subgradient method using the dynamic stepsize rule* (31) *for unknown f^* with $f^{lev} > f^*$, we have:*

$$\liminf_{k\to\infty} f(x_k) \le f^{lev} + \liminf_{k\to\infty} c_kd_{\mathbb{X}}$$

Proof Assume to contrary that

$$\liminf_{k\to\infty}(f(x_k)) > f^{lev} + \lim_{k\to\infty} c_kd_{\mathbb{X}} + \varepsilon$$

for some $\varepsilon > 0$.

Let k_0 be large enough so that

$$f(x_k) - f^{lev} - c_kd_{\mathbb{X}} \ge \varepsilon \tag{32}$$

for all $k \ge k_0$. Then by using Lemma 2 and relations $f^{lev} > f^*$ and (26), we obtain:

$$\|x_{k+1}-x^*\|^2 \le \|x_k-x^*\|^2 - 2\alpha_k(f(x_k)-f^{lev}-c_kd_{\mathbb{X}})+\alpha_k^2\|v_k\|^2.$$

Using the definition of the dynamic stepsize (31), the preceding inequality implies for all $k \ge k_0$:

$$\|x_{k+1}-x^*\|^2 \le \|x_k-x^*\|^2 - \gamma_k(2-\gamma_k)\frac{(f(x_k)-f^{lev}-c_kd_{\mathbb{X}})^2}{\|v_k\|^2}.$$

Then by combining with (32) and (24), this leads to

$$\|x_{k+1}-x^*\|^2 \le \|x_k-x^*\|^2 - \gamma_k(2-\gamma_k)\frac{\varepsilon^2}{D^2}.$$

Since $0 < \underline{\gamma} < \gamma_k \le \overline{\gamma} < 2$, the last relation gives

$$\|x_{k+1}-x^*\|^2 \le \|x_k-x^*\|^2 - \underline{\gamma}(2-\overline{\gamma})\frac{\varepsilon^2}{D^2},$$

from which we obtain for all $k \ge k_0$,

$$\begin{aligned}\|x_{k+1}-x^*\|^2 &\le \|x_k-x^*\|^2 - \underline{\gamma}(2-\overline{\gamma})\frac{\varepsilon^2}{D^2}\\ &\le \cdots \le \|x_{k_0}-x^*\| - \underline{\gamma}(2-\overline{\gamma})\frac{(k+1-k_0)\varepsilon^2}{D^2}.\end{aligned}$$

This relation cannot hold for k sufficient large, so it is a contradiction. □

4.4 Convergence analysis for the dynamic stepsize for unknown f^*, with $f^{lev} < f^*$.

As it is explained in the previous section, f^{lev} may be greater or smaller than f^*. In this subsection we investigate the case $f^{lev} < f^*$. The dynamic stepsize rule for unknown f^* with $f^{lev} < f^*$, is defined in the following form:

$$\alpha_k = \gamma_k \frac{f(x_k) - f^{lev} - c_k d_{\mathbb{X}}}{||v_k||^2}, \tag{33}$$

where $0 < \underline{\gamma} \leq \gamma_k \leq \overline{\gamma} < 1$. As a matter of course, the value of c_k must be selected less than $\frac{f(x_k)-f^{lev}}{d_{\mathbb{X}}}$, to ensure positiveness of the stepsize.

The following proposition gives a convergence result for the sequence of values $\{f(x_k)\}$, generated using the dynamic stepsize rule (33) for unknown f^* with $f^{lev} < f^*$.

Proposition 8 *Assume that $f(x_k) > f^{lev} + c_k d_{\mathbb{X}}$ for every $\{x_k\}$ generated by the weak subgradient method using the dynamic stepsize rule* (33) *for unknown f^* with $f^{lev} < f^*$. Then we have:*

$$\liminf_{k\to\infty} f(x_k) \leq f^* + (f^* - f^{lev}) + \liminf_{k\to\infty} c_k d_{\mathbb{X}}.$$

Proof Assume to contrary that

$$\liminf_{k\to\infty}(f(x_k)) > f^* + (f^* - f^{lev}) + \liminf_{k\to\infty} c_k d_{\mathbb{X}} + \varepsilon$$

for some $\varepsilon > 0$. Let k_0 be large enough so that

$$f(x_k) \geq f^* + (f^* - f^{lev}) + c_k d_{\mathbb{X}} + \varepsilon \tag{34}$$

for all $k \geq k_0$. Then by using Lemma 2 and relations $f^{lev} < f^*$ and (26), we obtain:

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 2\alpha_k(f(x_k) - f^* - c_k d_{\mathbb{X}}) + \alpha_k^2\|v_k\|^2$$

Using the definition of the dynamic stepsize (33), the preceding inequality implies for all $k \geq k_0$:

$$\begin{aligned}\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 &- 2\gamma_k \frac{f(x_k) - f^{lev} - c_k d_{\mathbb{X}}}{\|v_k\|^2}(f(x_k) - f^* - c_k d_{\mathbb{X}}) \\ &+ \gamma_k^2 \frac{(f(x_k) - f^{lev} - c_k d_{\mathbb{X}})^2}{\|v_k\|^2}.\end{aligned}$$

Since $\gamma_k < 1$, then $\gamma_k^2 \leq \gamma_k$, hence

$$\begin{aligned}\|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 - 2\gamma_k \frac{f(x_k) - f^{lev} - c_k d_{\mathbb{X}}}{\|v_k\|^2}(f(x_k) - f^* - c_k d_{\mathbb{X}}) \\ &+ \gamma_k \frac{(f(x_k) - f^{lev} - c_k d_{\mathbb{X}})^2}{\|v_k\|^2} \\ &= \|x_k - x^*\|^2 \\ &- \gamma_k(f(x_k) - f^{lev} - c_k d_{\mathbb{X}}) \frac{[f(x_k) - f^* - (f^* - f^{lev}) - c_k d_{\mathbb{X}}]}{\|v_k\|^2}.\end{aligned}$$

By using assumptions (34) and (24) we obtain:

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 2\gamma_k \frac{[2(f^* - f^{lev}) + \varepsilon]\varepsilon}{D^2}$$

Since $0 < \underline{\gamma} < \gamma_k \leq \overline{\gamma} < 1$, the last relation gives

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \underline{\gamma} \frac{[2(f^* - f^{lev}) + \varepsilon]\varepsilon}{D^2},$$

from which we obtain for all $k \geq k_0$,

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 - \underline{\gamma} \frac{[2(f^* - f^{lev}) + \varepsilon]\varepsilon}{D^2} \leq ... \\ &\leq \|x_{k_0} - x^*\| - (k + 1 - k_0)\underline{\gamma} \frac{[2(f^* - f^{lev}) + \varepsilon]\varepsilon}{D^2}. \end{aligned}$$

Since $\mathbb{X}$ is a compact set, the last relation cannot hold for k sufficiently large, so it is a contradiction. □

4.5 Convergence analysis for the dynamic stepsize with dynamic f_k^{lev}.

Instead of choosing a constant value for f^{lev} at the beginning of the weak subgradient algorithm, a dynamic value f_k^{lev} which changes at every iteration according to some rule, is investigated in this subsection. The dynamic stepsize rule with dynamic f_k^{lev} is defined in the following form:

$$\alpha_k = \gamma_k \frac{f(x_k) - f_k^{lev} - c_k d_{\mathbb{X}}}{\|v_k\|^2}, \tag{35}$$

where $0 < \underline{\gamma} \leq \gamma_k \leq \overline{\gamma} < 1$, and $f_k^{lev} = \min_k\{f(x_k)\} - \delta_k$. For the next iteration, the value of δ_{k+1} is calculated as follows:

$$\delta_{k+1} = \begin{cases} \min\{\beta_1 \delta_k, \overline{\delta}\} & if \quad f(x_{k+1}) < f_k^{lev} \\ \max\{\beta_2 \delta_k, \underline{\delta}\} & if \quad f(x_{k+1}) > f_k^{lev}, \end{cases}$$

where $1 < \beta_1, 0 < \beta_2 < 1$ and $\delta_0, \overline{\delta}, \underline{\delta}$ are positive constants. As a matter of course, the c_k must be selected such that as $f(x_k) - f_k^{lev} - c_k d_{\mathbb{X}} > 0$, to ensure positiveness of the stepsize.

The following proposition gives a convergence result for the sequence of values $\{f(x_k)\}$, generated using the dynamic stepsize rule (35) with dynamic f^{lev}.

Proposition 9 *Assume that there exists a number ρ such that $f(x_k) - f_k^{lev} - c_k d_{\mathbb{X}} > \rho > 0$ for every $\{x_k\}$ generated by the weak subgradient method using the dynamic stepsize rule* (35) *with dynamic f_k^{lev}. Then we have:*

$$\liminf_{k\to\infty} f(x_k) \leq f^* + \overline{\delta} + \liminf_{k\to\infty} c_k d_{\mathbb{X}}.$$

Proof Assume to contrary that

$$\liminf_{k\to\infty}(f(x_k)) > f^* + \overline{\delta} + \liminf_{k\to\infty} c_k d_{\mathbb{X}} + \varepsilon$$

for some $\varepsilon > 0$. Let k_0 be large enough so that

$$f(x_k) - f^* - c_k d_{\mathbb{X}} \geq \overline{\delta} + \varepsilon \tag{36}$$

for all $k \geq k_0$. By applying Lemma 2 and the similar procedure used in the proof of Proposition 8, we obtain (just by taking f_k^{lev} instead of f^{lev}) :

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 2\gamma_k \frac{f(x_k) - f_k^{lev} - c_k d_{\mathbb{X}}}{\|v_k\|^2}(f(x_k) - f^* - c_k d_{\mathbb{X}})$$
$$+ \gamma_k \frac{(f(x_k) - f_k^{lev} - c_k d_{\mathbb{X}})^2}{\|v_k\|^2}$$
$$= \|x_k - x^*\|^2$$
$$- \gamma_k(f(x_k) - f_k^{lev} - c_k d_{\mathbb{X}}) \frac{[2(f(x_k) - f^* - c_k d_{\mathbb{X}}) - (f(x_k) - f_k^{lev} - c_k d_{\mathbb{X}})]}{\|v_k\|^2}$$
$$= \|x_k - x^*\|^2$$
$$- \gamma_k(f(x_k) - f_k^{lev} - c_k d_{\mathbb{X}}) \frac{[(f(x_k) - f^* - c_k d_{\mathbb{X}}) - (f^* - f_k^{lev})]}{\|v_k\|^2}.$$

Now by using assumptions (36)) and (24), for all $k \geq k_0$ we obtain:

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \gamma_k(f(x_k) - f_k^{lev} - c_k d_{\mathbb{X}}) \frac{[\overline{\delta} + \varepsilon - (f^* - f_k^{lev})]}{D^2}. \tag{37}$$

According to definition (35) of α_k, we have $f_k^{lev} = \min_k f(x_k) - \delta_k$. Obviously, $\min_k f(x_k) \geq f^*$, for all k, hence, $f_k^{lev} \geq f^* - \delta_k$. Additionally, by the hypothesis $\delta_k \leq \overline{\delta}$. Thus, we have $f_k^{lev} \geq f^* - \overline{\delta}$ and finally

$$\overline{\delta} \geq f^* - f_k^{lev}. \tag{38}$$

Now, (37) and (38) lead to

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \gamma_k(f(x_k) - f_k^{lev} - c_k d_{\mathbb{X}}) \frac{\varepsilon}{D^2},$$

for all $k \geq k_0$. Let ρ be a positive number with $0 < \rho < f(x_k) - f_k^{lev} - c_k d_{\mathbb{X}}$. Then we obtain:

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \gamma_k \rho \frac{\varepsilon}{D^2}.$$

Since $0 < \underline{\gamma} < \gamma_k \leq \overline{\gamma} < 1$, the last relation gives

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \underline{\gamma} \rho \frac{\varepsilon}{D^2},$$

from which we obtain for all $k \geq k_0$,

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \|x_k - x^*\|^2 - \underline{\gamma}\rho\frac{\varepsilon}{D^2} \\ &\leq \|x_{k-1} - x^*\|^2 - 2\underline{\gamma}\rho\frac{\varepsilon}{D^2} \\ &\leq \cdots \leq \|x_{k_0} - x^*\| - (k+1-k_0)\underline{\gamma}\rho\frac{\varepsilon}{D^2}. \end{aligned}$$

Since $\mathbb{X}$ is a compact set, the last relation cannot hold for k sufficiently large, so it is a contradiction. □

The following remark subsums convergence analysis for all the stepsizes used in this paper.

Remark 8 Lemma 2 shows that, by using the assumptions (24), (25) and the compactness assumption for the feasible set $\mathbb{X}$, and by choosing decreasing sequences for $\{c_k\}$ and $\{\alpha_k\}$, it is possible to generate sequence of feasible solutions x_k with decreasing sequence of distances to an optimal solution x^*, for all stepsizes used. On the other hand, as it follows from Proposition 6, the sequence of feasible solutions generated using the dynamic stepsize for unknown f^* with $f^{lev} > f^*$, satisfies this property.

The propositions 5, 7, 8, and 9 show that, the convergence quality of the sequences of values $f(x_k)$, generated using the diminishing and the dynamic stepsizes, depends on the weak subgradient (v_k, c_k) and the chosen value f^{lev} and/or f_k^{lev} for the target level, at every iteration.

It follows from Proposition 5 that, if the sequence of weak subgradients $\{(v_k, c_k)\}$ can be chosen so that $(v_k, c_k) \to (0, 0)$ as $k \to \infty$, then for the sequence of values $f(x_k)$, generated using the diminishing stepsize, we have $\liminf_{k\to\infty} f(x_k) = f^*$. The reason for choosing such a sequence $\{c_k\}$, is the optimality of point x, for which $(0, 0) \in \partial^w f(x)$, see Remark 2.

5 Computational Results

This section presents computational results which demonstrate the performance and efficiency of the weak subgradient algorithm. These computational results are obtained solving 71 academic test problems from the literature. The analytic problem taken from [70] (see also [3]), demonstrates the ability of the method, to escape local optimal points to find global solution. 21 nonsmooth convex test problems are taken from [51], 19 small scale nonsmooth nonconvex test problems having from 2 to 10 decision variables, are taken from [9, 40, 50, 54, 61], and totally 15 + 15 large scale nonsmooth nonconvex test problems having 50 and 200 decision variables respectively, are taken from [9, 14, 30, 32, 53].

The weak subgradient method is coded with Python programming language and run with MacBook Pro which has 2.5GHz Intel Core i7 processor and 16 GB 1600 MHz DDR3 RAM. In all problems, the box constraints are chosen as $[-5, 5]$, for all variables. If any component of optimal point for some problem, is not in this interval, then the box constraint in the relevant problem, is updated as $[x_i^* - 5, x_i^* + 5] \quad i \in \{1, \dots, n\}$.

We use the success percentage evaluation criteria (39) given below, where the values for ε are taken 5×10^{-5}, 10^{-3} and 10^{-2}, for all stepsizes, except for dynamic stepsize α_k^{dyn1} with $f^{lev} > f^*$. For this stepsize, the evaluation is computed by taking f^{lev} instead of f^* :

$$\frac{f_{wsa} - f^*}{1 + |f^*|} \leq \varepsilon \tag{39}$$

The notations used, are listed below:

- n: The number of variables;
- $\overline{c_k}$: The upper bound of c_k for the dynamic stepsizes;
- f_{wsa}: The best objective function value calculated by the weak subgradient algorithm;
- f_{wsa}^c: The objective function value calculated by the weak subgradient algorithm with constant stepsize;
- f_{wsa}^{dim}: The objective function value calculated by the weak subgradient algorithm with diminishing stepsize;
- f_{wsa}^{dyn1}: The objective function value calculated by the weak subgradient algorithm with dynamic stepsize with $f^{lev} > f^*$;
- f_{wsa}^{dyn2}: The objective function value calculated by the weak subgradient algorithm with dynamic stepsize with $f^{lev} < f^*$;
- f_{wsa}^{dyn3}: The objective function value calculated by the weak subgradient algorithm with dynamic stepsize with f_k^{lev};
- α^c : Constant stepsize;
- α_k^{dim} : Diminishing stepsize at kth iteration;
- α_k^{dyn1} : Dynamic stepsize with with $f^{lev} > f^*$ at kth iteration;
- α_k^{dyn2} : Dynamic stepsize with with $f^{lev} < f^*$ at kth iteration;
- α_k^{dyn3} : Dynamic stepsize with with f_k^{lev} at kth iteration;
- ε : The accuracy level for the evaluation criteria given in (39).

5.1 An analytic problem from [70] (see also [3]).

An analytic problem given below, has too many local optimum points. The graph of this function is given in Figure 2, which is taken from [3].

$$f(a,b) = e^{\sin 50a} + \sin(60e^b) + \sin(70\sin(a)) + \sin(\sin(80b)) - \sin(10(a+b) + \frac{1}{4}(a^2+b^2). \tag{40}$$

The global optimal solution of the problem is $x^* = [-0.024, 0.211]$. The parameters used in and the results obtained by the weak subgradient algortihm, are given in Tables 2 and 3, respectively. The initial point for applying the weak subgradient algorithm, is taken as $x^1 = [3, 3]$, which is far from the

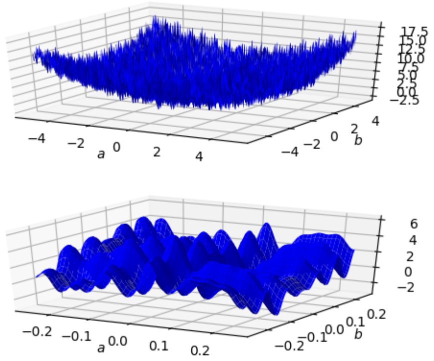

Fig. 2 Graph of analytic problem with bounds $[-5, 5]$ (above) and $[-0.25, 0.25]$ (below)

optimal point. The results represent that the algortihm successfully escapes from local optimums and finds global optimal solution as in Mash Adaptive Direct Search with Variable Neighborhood Search (MADS with VNS) [3].

Table 2 Parameters used in the weak subgradient algorithm with stepsizes for the analitic problem

	α^c	α_k^{dim}	α_k^{dyn1}	α_k^{dyn2}	α_k^{dyn3}
α_k	0.01	$1-\frac{1}{IterNum}*k$	–	–	–
c_k	$1-\frac{1}{IterNum}*k$	$1-\frac{1}{IterNum}*k$	$\overline{c}_k * 0.85$	$\overline{c}_k * 0.4$	$\overline{c}_k * 0.4$
λ	0.1	1	0.1	0.001	0.1
α	1	1	1	1	0.5
f^{lev}	–	–	-2.807	-3.807	–
δ_1	–	–	–	–	$f(x_1) * 0.15$
$\underline{\delta}$	–	–	–	–	$\delta_1 * 1.15$
β_1	–	–	–	–	1.5
β_2	–	–	–	–	0.5

5.2 Test problems

Brief descriptions of nonsmooth test problems which include names, references, numbers of variables and optimal values of problems (some of them are approximate values), are given in Table 4, and the parameters used in the weak

Table 3 Results of the analitic problem obtained by the weak subgradient algorithm

f^*	$f_{mads-vns}$	f_{wsa}^c	f_{wsa}^{dim}	f_{wsa}^{dyn1}	f_{wsa}^{dyn2}	f_{wsa}^{dyn3}
−3.307	−3.307	−3.305	−3.293	−3.178	−3.305	−3.307
CPU (in seconds)	–	0.621	0.739	0.409	0.738	17.556
$\frac{f_{WSA}-f*}{1+\|f*\|}$	0	0.0005	0.0107	0	0.0005	0

subgradient algorithm for these test problems are shown in Tables 5 and 6. The starting points are also taken equal to the same values given in the reference of the corresponding problem. The iteration limit is taken equal to 40000, for all test problems. Note that, the weak subgradients (v_k, c_k) are computed using the Algorithm 1, given in Section 3. In accordance with the explanations given in Remarks 2 and 8, the sequence $\{c_k\}$ are generated as a decreasing sequence, see Tables 2 and 5.

The computational results and the CPU times obtained for nonsmooth (convex, nonconvex small scale, nonconvex large scale with 50 variables and nonconvex large scale with 200 variables) test problems are given in Tables 7 and 8, respectively. The success percentage values for every stepsize parameter, are presented in Table 9. Additionally, we present detailed solutions obtained using the weak subgradient method with different stepsizes, for problem $P35$ (or EXP) in a separate table (see Table 10), for which the proposed method essentially improves the best known solution.

Below, we present interpretations made on (and conclusions drawn from) the computational results obtained by applying the weak subgradient method, which are depicted in tables 4 - 10.

- Computational results presented in Table 7 demonstrate that the proposed method is able to find (approximate) optimal solutions in all cases. The weak subgradient algoirthm found better solution for problem $P35$ (or EXP) (see Table 10). The solutions found for problems P1, P5, P6 and P16, are slightly better than the solutions reported in the literature, and the result obtained for problems P7 and P39, are slightly worse than the reported one.
- The best success percentange values for the proposed method, are obtained as %81, %100, %80 and %60 for each type of test problems, respectively, using $\varepsilon = 5 \times 10^{-5}$ and $\varepsilon = 10^{-3}$. All these values are obtained for the dynamic stepsize with $f^{lev} > f^*$.
- For $\varepsilon = 10^{-2}$, the best success percentange values obtained, are %95, %100, %94 and %87 for each type of test problems, respectively. All these values are again obtained for the dynamic stepsize with $f^{lev} > f^*$.
- In the cases, if $f^{lev} > f^*$, the weak subgradient method works well and reaches at worst the value f^{lev}.

- For $\varepsilon = 10^{-2}$, the average success percentage value for the proposed method, is obtained %92 over all test problems.
- For $\varepsilon = 10^{-3}$, the average success percentage value for the proposed method, is obtained %77 over all test problems.
- For $\varepsilon = 5 \times 10^{-5}$, the average success percentage value for the proposed method, is obtained %59 over all test problems.
- We have observed that one of the main reasons for the worst result obtained using the proposed method, is the iteration limit. While the number of variables increases, the method needs more iteration number. For example, we tested that, by using the dynamic stepsize with $f^{lev} < f^*$ and the iterariton limit 250000, the reported best solution for problem P39 (or PBC1) is obtained. The another reason is that, the assumptions made for the sup and max relations between the directional derivative and the weak subdifferential, may not be satisfied.
- The CPU times are in accepteble limits for all test problems. It varies between 1 second and 1 minute for small scale problems, between 1 minute and 15 minutes for large scale test problems with 50 variables, and between 15 minutes and 4 hours for large scale test problems with 200 variables. The most of the CPU time probably, is needed for approximate evaluation of the weak subgradinets, which increases with the number of variables.

5.2.1 Comparision with other algorithms

In this section we compare the results obtained by the proposed method (WSA), with results of the following methods taken from the literature: Discreate Gradient Method (DGM) [11], Nonconvex via Cutting Planes (NCVX) [26], Approximate Subgradient Method (ASM) [10], Subgradient Algorithm 1 (SUB1) [10], Subgradient Algorithm 2 (SUB2) [10], DC-NCVX [27], Variable Matric Nonconvex (VMNC) [71]. The results are given in Table 11. The methods denoted by SUB 1 and SUB 2, are subgradient method and use constant and diminishing stepsizes, respectively. The notations used in this table, are listed below:

- f_{dgm}: The best value of the objective function, computed using the discrete gradient method.
- f_{ncx}: The best value of the objective function, computed using the "nonconvex via cutting planes" method.
- f_{asm}: The best value of the objective function, computed using the approximate subgradient method.
- f_{sub1}: The best value of the objective function, computed using the subgradient algorithm 1.
- f_{sub2}: The best value of the objective function, computed using the subgradient algorithm 2.
- $f_{dc-vxnc}$: The best value of the objective function, computed using "DC-NCVX" method.

– f_{vmnc}: The best value of the objective function, computed using the "variable matric nonconvex" method.

The weak subgradient algorithm generates better results for 6 test problems, which are P2, P7, P16, P34, P35 and P38. In Table 11 the results corresponding to these problems in the column of WSA, are shown by bold numbers.

6 Conclusion

In this paper, the weak subgradient method for solving nonsmooth nonconvex unconstrained optimization problems, is proposed. The method, iteratively calculates feasible solutions, using weak subgradients at every iteration. To compute the weak subgradients, the paper presents a special approximate algorithm. This algorithm is based on the sup relation between the directional derivative and the weak subdifferential, which requires some conditions on the directional derivtive. One of the main problems for future works, is to investigate the ways of weakining these conditions. On the other hand, the paper investigates and presents conditions, under which the sup relation can be written as a max relation, which is very important from point of view of develpoing more accurate methods for computing weak subgradients. We beleive that the problem of investigating the class of functions for which the sup relation can be written as a max relation, is an independent and interesting problem, which can be considered in the future. The proposed method employs the constant, the diminishing and the three types of dynamic stepsizes, and analyses convergence properties for all of the stepsize parameters used in the presented method. The performance of the proposed weak subgradient method is demonstrated on well-known test problems from the literature, and the detailed analysis of the stepsizes and the obtained computational results, are presented.

7 Acknowledgements

The authors thank Prof. Adil Bagirov for his insightful comments on the manuscript.

This study was supported by Anadolu University Scientific Research Projects Commission under the grant no: 1602F082 and the Scientific and Technological Research Council of Turkey (Tubitak) as PhD fellowship of the first author.

References

1. Akbari, Z., Yousefpour, R., Peyghami, M. R.: A new nonsmooth trust region algorithm for locally lipschitz unconstrained optimization problems. J optim Theory Appl, **164**, 733–754 (2015)
2. Allen, E., Helgason, R., Kennington, J., Shetty, B.: A generalization of Polyak's convergence result for subgradient optimization. Mathematical Programming, **37**, 309–317 (1987)

3. Audet, C., Bechard, V., Digabel, S.: Nonmooth optimization through Mesh Adaptive Direct Search and Variable Neighborhood Search, Journal of Global Optimization. **41**, 299–318 (2008)
4. Auslender, A., Teboulle, M.: Interior gradient and $\varepsilon-$subgradient descent methods for constrained convex minimization. Mathematics of Operations Research. **29**, 1–26 (2004)
5. Azimov, A., Gasimov, R: On weak conjugency, weak subdifferentials and dualty with zero gap in nonconvex optimiztion. International Journal of Applied Mathematics. **1**, pp. 171–192 (1999)
6. Azimov, A., Gasimov, R: Stability and dualty of nonconvex problems via augmented Lagrangian. Cybernetics and System Analysis. **3**, 403–417 (2002)
7. Bagirov, A: A method for approximating subdifferential. Computational Mathematics and Mathematical Physics. **32**(4), 561–566 (1992)
8. Bagirov, A.: A method for minimizing convex functions based on continuous approximations to subdifferential. Optimization Methods and Software. **9**, 1–17 (1998)
9. A. Bagirov, A.: A method for minimization of quasidifferentiable functions. Optimization Methods and Software. **17**(1), 31–60 (2002)
10. Bagirov, A., Ganjehlou, A. N. : An approximate subgradient algorithm for unconstrained nonsmooth, nonconvex optimization. Math Meth Oper Res. **67**, 187–206 (2008)
11. Bagirov, A., Karasozen, B., Sezer, M.: Discrete gradient method: derivative-freee method for nonsmooth optimization. Journal of Optimization Theory and Applications. **137**, 317–334 (2008)
12. Bagirov, A., Karmitsa, N., Makela, M.: Introduction to Nonsmooth Optimization, Springer, Cham (2014)
13. Bagirov, A., Ozturk, G., Kasimbeyli, R: A sharp augmented Lagrangian-based method in constrained nonconvex optimization. Optimization Methods and Software (to appear), DOI: 10.1080/10556788.2018.1496431
14. Bagirov, A., Ugon, J.: Coodifferential method for minimizing nonsmooth dc functions. Journal of Global Optimization. **50**(1), 3–22 (2011)
15. Bazaraa, M., Sherali, H.: On the coice of the stepsize in subgradient optimization. European Journal of Operational Research. **7**, 380–388 (1981)
16. Bazaraa, M., Sherali, H.,Shetty, C.: Nonlinear programming Theory and Algorithms. 4th Edition. New Jersey: Wiley (2006)
17. Bertsekas, D. P.: Nonlinear Programming. Belmont, Massachusetts: Athena Scientific (1999)
18. Bertsekas, D.P., Nedić, A, Ozdaglar, A.: Convex Analysis and Optimization, Athena Scientific, Cambridge, 2003
19. Burachik, R., Gasimov, R., Ismayilova, N., Kaya, C.: On a modified subgradient algorithm for dual problems via sharp augmented lagrangian. Journal of Global Optimization. **34**(1), 55–78 (2006)
20. Burke, J., Lewis, A., Overton, M.: A robust gradient sampling algorithm for nonsmooth nonconvex optimization. SIAM Journal on Optimization, **15**, 571–779 (2005)
21. Clarke, F.: Optimization and nonsmooth analysis. New York: Wiley-Interscience (1983)
22. Dennis, J., Li, S.,Tapia, R.: A unified approach to global convergence of trust region methods for nonsmooth optimization. Math. Program. **68**, 319—346 (1995)
23. DAntonio, G., Frangioni, A.: Convergence analysis of deflected conditional approximate subgradient methods. SIAM Journal on Optimization. **20**, 357–386, (2009)
24. Ermoliev, Y.M.: Methods of solution of nonlinear extremal problems, Cybernetics and Systems Analysis, **2**, 1–14 (1966)
25. Ermoliev, Y. M.: Stochastic Programming Methods, Nauka, Moscow, (1976) (in Russian)
26. Fuduli, A., Gaudioso, M., Giallombardo, G.: Minimizing nonconvex nonsmooth functions via cutting planes and proximity control. SIAM Journal on optimization. **14**(3), 743–756 (2004)
27. Fuduli, A., Gaudioso, M., Giallombardo, G.: A DC piecewise affine model and a bundling technique in nonconvex nonsmooth minimization. Optimization Methods and Software, **19**(1), 89–102 (2004)
28. Gasimov, R.: Augmented Lagrangian duality and nondifferentiable optimization methods in nonconvex programming. J. Global Optimization. **24**, 187–203 (2002)

29. Gasimov, R., Rubinov, A. M.: On augmented Lagrangians for optimization problems with a single constraint. Journal of Global Optimization. **28**(5), 153–173 (2004)
30. Grothey, A.: Decomposition methods for nonlinear nonconvex optimization problems. PhD thesis, University of Edinburgh (2001)
31. Goffin, J. L.: On Convergent Rates of Subgradient Optimization Methods. Math. Programming. **13**, 329–347 (1977)
32. Haarala, M., Miettinen, K. Mäkelä, M. M.: New limited bundle method for large scale nonsmooth optimization. Optimization Methods and Software. **19**(6), 673–692 (2004)
33. Hu, Y., Yang, X., Sim, K.C.: Inexact subgradient methods for quasi-convex optimization problems. European Journal of Operational Research. **240**(2), 315–327 (2015)
34. Kasimbeyli, R.: Radial Epiderivatives and Set-Valued Optimization. Optimization. **58**, 521–534 (2009)
35. Kasimbeyli, R.: A Nonlinear Cone Separation Theorem and Scalarization in Nonconvex Vector Optimization, SIAM J. on Optimization, **20**, 1591–1619 (2010)
36. Kasimbeyli, R.: A conic scalarization method in multi-objective optimization. Journal of Global Optimization. **56**(2), 279–297 (2013)
37. Kasimbeyli, R., Mammadov, M.: On weak subdifferentials, directional derivatives and radial epiderivatives for nonconvex functions. SIAM Journal on Optimization. **20**(2), 841–855 (2009)
38. Kasimbeyli, R., Mammadov, M.: Optimality conditions in nonconvex optimization via weak subdifferentials, Nonlinear Analysis: Theory, Methods and Applications, **74**(7), 2534–2547 (2011)
39. Kasimbeyli, R., Ustun, O., Rubinov, A.M. : The Modified Subgradient Algorithm Based on Feasible Values. Optimization. **58**(5), 535–560 (2009)
40. Kiwiel, K.: Methods of descent for nondifferentiable optimization Lecture Notes in Mathematics. Berlin: Springer (1985)
41. Kiwiel, K.: The efficiency of subgradient projection methods for convex optimzation, Part I: General level methods. SIAM Journal on Control and Optimization, **34**(2), 660–676 (1996)
42. Kiwiel, K.: The efficiency of subgradient projection methods for convex optimzation, Part II: Implementationas and extensions. SIAM Journal on Control and Optimization. **34**(2), 677–697 (1996)
43. Kiwiel, K.: Restricted step and Levenberg-Marquardt techniques in proximal bundle methods for nonconvex nondifferentiable optimization. SIAM Journal on Optimization. **6**, 227-249 (1996)
44. Kiwiel, K.: Convergence and efficiency of subgradient methods for quasiconvex minimization. Mathematical Programming Series A. **90**, 1–25 (2001)
45. Kiwiel, K.: Convergence of approximate and incremental methods for convex optimization. SIAM Journal of Optimization. **14**(3), 807–840 (2004)
46. Kiwiel, K.: Convergence of the gradient sampling algorithm for nonsmooth nonconvex optimization. SIAM Journal of Optimization. **18**(2), 379-388 (2007)
47. Larsson, T., Patriksson, M., Stromberg, A.-B.: Conditional subgradient optimization theory and applications, European Journal of Operational Research, **88**, 382–403. (1996)
48. Larsson, T., Patriksson, M., Stromberg, A.-B.: On the convergence of conditional $\varepsilon-$subgradient methods for convex programs and convex-concave saddle-point problems. European Journal of Operational Research. **151**, 461– 473 (2003)
49. Lukšan, L., Vlček, J.: A Bundle-Newton method for nonsmooth unconstrained minimizaiton. Mathematical Programming, 83, 373–391 (1998)
50. Lukšan, L., Vlček, J.: NDA: Algorithms for nondifferentiable optimization. Technical Report 797, Institute of Computer Science, Academy of Sciences of the Czech Rebuplic, Prague (2000)
51. Lukšan, L., Vlček, J.: Test problems for nonsmooth unconstrained and linearly constrained optimization. Technical Report 790, Institute of Computer Science, Academy of Sciences of the Czech Rebuplic, Prague (2000)
52. Lukšan, L., Tčma, M., Šiška, M., Vlček, J., Ramešová, N.: UFO 2002. Interactive system for universal functional optimization. Technical Report 883, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague (2002)

53. Lukšan, L., Tčma, M., Šiška, M., Vlček, J., Ramešová, N.: UFO 2002. Interactive system for universal functional optimization. Technical Report 883, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague (2002)
54. Mäkelä, M. M., Neittaanmäki, P.: Nonsmooth optimization: Analysis and algoirthms with applications to optimal control. Singapore: World Scientific Publishing Co. (1992)
55. Mikhalevich, V. S., Gupal, A. M., Norkin, V. I.: Methods of Nonconvex Optimization, Nauka, Moscow, 1987 (in Russian)
56. Haarala, N., Miettinen, K., Mäkelä, M. M.: Globally convergent limited memory bundle method for large-scale nonsmooth optimization. Mathematical Programming Series: A. **109**, 181–205 (2007)
57. Nedić, A., Bertsekas, D. P.: Incremental subgradient methods for nondifferentiable optimization. SIAM Journal on Optimization. **12**, 109–138 (2001)
58. Nedić, A., Bertsekas, D. P.: The effect of deterministic noise in subgradient methods. Mathematical Programming Series: A, **125**, 75–99 (2010)
59. Nedić, A., Ozdaglar, A.: Approximate primal solutions and rate analysis for dual subgradients methods. SIAM Journal on Optimization. **19**(4), 1757–1780 (2009)
60. Nurminski, E. A.: Subgradient method for minimizing weakly convex functions and epsilon-subgradient methods of convex optimization, in Progress in Nondifferentiable Optimization, E. A. Nurminski, ed., CP–82–S8, International Institute for Applied Systems Analysis, Laxenburg, Austria, , 97–123 (1982)
61. Polak, E., Mayne, D. Q., Higging, J. E.: Superlinearly convergent algorithm for min-max problems. Journal of Optimization Theory and Applications. **69**, 407–439 (1991)
62. Polyak, B. T.: A general method for solving extremum problems, Soviet Mathematics Doklady, **8**, 593–597 (1967)
63. Polyak, B. T.: Minimization of Unsmooth Functionals. Z. Vychislitelnoy Matematiki i Matematicheskoy Fiziki, **9**, 14–29 (1969)
64. Polyak, B.T.: Nonlinear programming methods in the presence of noise. Math. Program. **1**(4), 87--97 (1978)
65. Polyak, B.T.: Introduction to Optimization. Optimization Software Inc., New York (1987)
66. Qivand, L., Sun, J.: A trust region algorithm for minimizing of locally Lipschitzian functions. Math. Program. **66**, 25–43 (1994)
67. Rockafellar, T: Convex analysis, Princeton, New Jersey (1970)
68. Shor, N. Z.: Minimizing Methods for Non-Differentiable Functions. (K. Kiwiel and A. Ruszcynski, translate) Heidelberg:Springer-Verlag Berlin (1985)
69. Solodov, M.V., Zavriev, S.K.: Error stability properties of generalized gradient-type algorithms. J. Opt. Theory Appl. **98**(3), 663–680 (1998)
70. Trefethen, N.: The hundred dollar hundredigit challenge. SIAM News. **35**, No.1 January–February (2002). www.siam.org/news/news.php?id=388.
71. Vlček, J., Lukšan, L.: Global convergent variable metric method for nonconvex nondifferentiable unconstrained minimization. Journal of Optimization Theory and Applications. **111**, 407–430 (2001)

Table 4 Brief description of nonsmooth test problems

Type of Prob.	**Number of Prob.**	**Name of Prob.**	**n**	f^*
Convex [51]	P1	Davidon 2	4	115.70644
	P2	Filter	9	0.0061852848
	P3	Wong 1	7	680.63006
	P4	Wong 2	10	24.306209
	P5	Wong 3	20	133.72828
	P6	Polak 3	11	3.7034924
	P7	Watson	20	0
	P8	Rosen-Suziki	4	-44
	P9	CB2	2	1.9522245
	P10	CB3	2	2
	P11	DEM	2	-3
	P12	QL	2	7.2
	P13	LQ	2	-1.4142136
	P14	Miffin 1	2	-1
	P15	Shor	5	22.600162
	P16	Maxquad	10	-0.8414083
	P17	Maxq	20	0
	P18	Maxl	20	0
	P19	Goffin	50	0
	P20	Wolfe	2	-8
	P21	MXHILB	50	0
Nonconvex Small Scale	P22	Crescent[40]	2	0
	P23	Mifflin 2 [54]	2	-1
	P24	WF [50]	2	0
	P25	SPIRAL [50]	2	0
	P26	EVD52 [50]	3	3.5991193
	P27	PBC3 [50]	3	0.0042021427
	P28	Bard [50]	3	0.050816327
	P29	Polak 6 [61]	4	-44
	P30	El-Attar [50]	6	0.5598131
	P31	Gill [50]	10	9.7857721
	P32	Problem 1[9]	2	2
	P33	Rosenbrock function (L_1) [9]	2	0
	P34	Wood function (L_1) [9]	4	0
	P35	EXP [50]	5	0.00012237125
	P36	Kowalik-Osborne [50]	4	0.0080843684
	P37	OET5 [50]	4	0.0026359735
	P38	OET6 [50]	4	0.0020160753
	P39	PBC1 [50]	5	0.022340496
	P40	EVD61 [50]	6	0.034904926
Nonconvex Large Scale	P41	Number of active faces [30]	any	0
	P42	Brown function (Nonsmooth) [32]	any	0
	P43	Chained crescent I [32]	any	0
	P44	Chained crescent II [32]	any	0
	P45	Problem 6 in Test29 [53]	any	0
	P46	Problem 17 in Test29 [53]	any	0
	P47	Problem 19 in Test29 [53]	any	0
	P48	Problem 20 in Test29 [53]	any	0
	P49	Problem 22 in Test29 [53]	any	0
	P50	Problem 24 in Test29 [53]	any	0
	P51	DC Maxl [9]	any	0
	P52	DC Maxlq [14]	any	0
	P53	Problem 6 in [14]	any	0
	P54	Problem 7 in [14]	any	0
	P55	Chained Mifflin 2 [32]	50 200	-34.795 -140.86

Table 5 Parameters used in the weak subgradient algorithm

Parameter Set Name	α_k^c	α_k^{dim}	c_k for α_k^c and α_k^{dim}	c_k for all α_k^{dyn}
a	0.001	$5/(2 * k)$	$10/k$	$\bar{c}_k * 0.9$
b	0.001	$2 - \frac{2}{IterNum} * k$	$10 - \frac{10}{IterNum} * k$	$\bar{c}_k * 0.8$
c	0.001	$1 - \frac{1}{IterNum} * k$	$5 - \frac{5}{IterNum} * k$	$\bar{c}_k * 0.85$
d	0.001	$10 - \frac{10}{IterNum} * k$	$1 - \frac{1}{IterNum} * k$	$\bar{c}_k * 0.5$
e	0.01	$1 - \frac{1}{IterNum} * k$	$1 - \frac{1}{IterNum} * k$	$\bar{c}_k * 0.4$

Table 6 The other parameters for all parameter set used in the weak subgradient algorithm

λ	α	δ_1	$\underline{\delta}$	β_1	β_2	f^{lev}
0.01/0.001/0.0001	1	$f(x_1) * 0.15$	$\delta_1 * 1.15$	1.5	0.5	$f^* \pm 0.5$

Table 7 Computational results for nonsmooth test problems obtained using weak subgradient algorithm

Type of Prob.	Num. of Prob.	f^*	f^c_{wsa}	f^{dim}_{wsa}	f^{dyn1}_{wsa}	f^{dyn2}_{wsa}	f^{dyn3}_{wsa}
Convex	P1	115.70644	115.582006	115.666549	116.20644	115.704403	115.3705
	P2	0.0061852848	0.01376181	0.01385348	0.01385349	0.00857663	0.00798035
	P3	680.63006	681.0100	680.6401	681.9702	682.1301	680.6301
	P4	24.306209	24.338133	24.318913	24.832245	24.523102	24.407890
	P5	133.72828	133.76411	168.74563	134.22828	133.72810	132.768464
	P6	3.7034924	3.694879	23.153547	4.20349	3.694605	3.6585028
	P7	0	0.484044	0.493821	0.434264	0.340174	0.361506
	P8	-44	-44.002375	-44.007696	-38.736398	-44.000197	0
	P9	1.9522245	1.95222562	1.9522272	2.4522245	1.952236	1.9522284
	P10	2	2	2	2.5	2	2
	P11	-3	-2.99999	-2.99999	-3.5	-2.99999	-2.99999
	P12	7.2	7.20001	7.200007	7.70987	7.2000231	7.200107
	P13	-1.4142136	-1.4141598	1.4141316	-0.9142136	-1.4142077	-1.4142025
	P14	-1	-1	-1	-0.8	-1	-0.8
	P15	22.600162	22.602604	22.6005197	23.099456	22.600620	22.604247
	P16	-0.8414083	-0.883092	-0.96222	-0.3414083	-0.84723	0
	P17	0	0	0	0.4928	0.0006471	0.08451001
	P18	0	0.0004798	0.000167	0.5	0.000912	0.019951
	P19	0	0.375037	0	0.5	0.006564	0.347126
	P20	-8	-8	-8	-7.5	-8	-8
	P21	0	0.002684	0.005827	0.5	0.000204	0
Nonconvex Small Scale	P22	0	0	0	0.5	0	0
	P23	-1	-1	-1	-0.5	-1	-1
	P24	0	0	0	0.5	0	0
	P25	0	0	0	0.5	0	0
	P26	3.5991193	3.5991193	3.59972865	4.0997193	3.59972443	3.59984305
	P27	0.0042021427	0.0042075	0.006310416	0.250397	0.00420867	0.00421077
	P28	0.050816327	0.05087519	0.05089106	0.5	0.05086314	0.0508552
	P29	-44	-43.99	-43.99	-43.5	-43.99	-43.99
	P30	0.5598131	0.55984302	0.55990149	1.0598131	0.55981367	0.56171104
	P31	9.7857721	9.79081678	9.78893903	10.2857721	9.789556	9.813723
	P32	2	2	2	2.38780222	2	2
	P33	0	0.00631878	0.00128238	0.49641875	0.00012007	0.00015433
	P34	0	0.00301524	0.01805092	0.5	0.00220735	0.0090316
	P35	0.00012237125	**-0.0121952**	**-5.7516929**	0.13212056	**-0.0622914**	**-0.0024076**
	P36	0.0080843684	0.00816606	0.01216215	0.0475133	0.00816741	0.00815057
	P37	0.0026359735	0.00270181	0.00309279	0.5026359	0.00290543	0.00325996
	P38	0.0020160753	0.00316036	0.01109648	0.181600722	0.00336216	0.00317971
	P39	0.022340496	0.19568146	0.02734077	0.5223405	0.08501973	0.11826176
	P40	0.034904926	0.03515377	0.03549639	0.53490493	0.03606183	0.03578041
Nonconvex Large Scale with 50 variables	P41	0	0.00319693	0.02869864	0.499999988	0.004078518	0.004235249
	P42	0	0.03272668	0.003179896	0.5	0.00971032	0.01909278
	P43	0	0.00083241	0.00048519	0.5	0.00116549	0.045976722
	P44	0	0.0143396	0.02202778	0.5	0	0.004872756
	P45	0	0.003469075	0.01176838	0.5	0.004783836	0.004071199
	P46	0	0	0.00034937	0.00034937	0.00034937	0
	P47	0	0.00040046	0.074161531	0.50416224	0.000933674	0.002417618
	P48	0	0.00543769	0.652707923	0.504220913	0.006593647	0.0073205
	P49	0	0.000681087	0.000681087	0.000681087	0.000681087	0.000680983
	P50	0	0.0155781	0.06082041	0.5	0.01266039	0.012916485
	P51	0	0.733963734	0	0.5	0.013277907	2.575630571
	P52	0	0	0	1	1	1
	P53	0	0.002896596	0.004477454	0.5	0.001401189	0.028024848
	P54	0	0	0.00034937	0.00034937	0.00034937	0
	P55	-34.795	-34.767706	-34.780171	-34.295	-34.767043	-34.70324
Nonconvex Large Scale with 200 variables	P41	0	0.09010911	0.97915948	0.5	0.009383143	0.01170317
	P42	0	0.1346936	199	0.5	0	0.096967
	P43	0	0.002835493	0.000968973	0.511846284	0.00180416	0.662850133
	P44	0	0.05358298	0	0.5	0	0.08312041
	P45	0	0.0052669	0.00838916	0.522376989	0.006795878	0.0093181
	P46	0	0	0	0	0	0
	P47	0	0.00719051	0.556159487	0.51424439	0.00378467	0.00852129
	P48	0	0.48974114	0.77293164	0.504458057	0.027131183	0.505147266
	P49	0	0	0	0	0	0
	P50	0	0.01485382	0.15269851	0.5	0.01171833	0.01205072
	P51l	0	6.27698523	0	0.48416392	0.39277116	9.54184555
	P52	0	0	0	1	1	1
	P53	0	0.013907926	0.132773358	0.5	0.00645211	0.061375149
	P54	0	6.75767563	0.53149851	0.5	0.120413022	1.170935921
	P55	-140.86	-140.72646	-140.79148	-139.9791	-140.73536	-139.8939

Table 8 CPU values (in seconds) for solutions of nonsmooth test problems obtained using weak subgradient algorithm

Type of Prob.	**Num. of Prob.**	f^{c}_{wsa}	f^{dim}_{wsa}	f^{dyn1}_{wsa}	f^{dyn2}_{wsa}	f^{dyn3}_{wsa}
Convex	P1	1.345	1.321	1.652	1.765	3.248
	P2	137.17	179.59	108.079	110.88	384.38
	P3	7.09	7.14	7.20	7.19	21.22
	P4	6.705	13.69	6.866	6.843	24.003
	P5	0.13	21.80	44.27	0.39	0.40
	P6	1.703	50.51	43.466	0.333	0.350
	P7	415.36	419.94	420.254	421.74	441.98
	P8	0.009	0.014	1.732	0.013	1.876
	P9	0.526	0.019	0.565	0.582	1.804
	P10	0.508	0.519	0.566	0.555	18
	P11	0.345	0.111	0.521	0.365	1.456
	P12	0.772	0.763	0.826	0.812	18.250
	P13	0.435	0.442	0.488	0.249	17.91
	P14	0.455	0.449	0.552	0.480	17.91
	P15	7.951	7.979	7.971	8.087	25.817
	P16	0.234	0.126	250.33	0.164	500.16
	P17	7.490	7.430	7.423	7.480	25.039
	P18	5.67	5.34	5.060	5.076	8.76
	P19	24.93	24.652	25.408	104.81	48.47
	P20	0.338	0.055	0.676	0.010	18.219
	P21	911.12	923.14	923.13	933.18	944.972
Nonconvex Small Scale	P22	0.01948595	0.014942884	0.631707907	0.434367895	3.1726861
	P23	0.034101009	0.014023066	0.566859007	0.061650991	0.462354898
	P24	0.547713041	0.533092976	0.663811922	0.622557878	20.59515786
	P25	0.281521082	0.86281395	0.912091017	0.952101946	20.95985794
	P26	1.022166967	1.020539999	1.148584127	1.108290911	20.95801091
	P27	9.269674063	9.183539152	9.327889919	9.248991013	26.57418299
	P28	1.943536043	1.943570137	2.056656122	2.07536602	22.84766889
	P29	2.697446108	2.847462893	2.990446091	2.97774601	20.02642322
	P30	32.12747598	36.26708293	36.0158658	36.07827997	53.48014712
	P31	140.1797779	140.2673838	139.792407	139.6418769	157.4582298
	P32	0.164664984	0.095170975	1.152608156	1.149125099	18.26833701
	P33	0.302460909	0.320847988	0.422068834	0.422070026	35.27688599
	P34	0.765971899	0.759523869	0.898361921	0.870677948	18.93293786
	P35	0.008785963	0.006563902	0.012379169	0.000733852	0.020792961
	P36	2.661329985	2.677920103	2.881727934	2.805956841	20.92395306
	P37	5.282094955	5.264348984	5.343361139	5.365101099	22.59601593
	P38	4.073904037	4.05181694	4.300542116	4.179040909	21.37431383
	P39	18.76057696	18.81023693	18.857306	18.71514201	36.15938878
	P40	36.879673	36.7343452	37.06612515	36.97888398	54.38752317
Nonconvex Large Scale with 50 variables	P41	60.89470315	61.32440186	61.43170595	61.22593093	80.11220503
	P42	118.402369	119.1848428	119.8790069	119.8271601	137.5036418
	P43	150.960669	151.9640241	152.103076	151.76353	169.998395
	P44	169.9192271	170.2918479	170.9188769	170.923506	189.8953011
	P45	57.08668494	55.80409384	55.75443602	55.6193831	73.220474
	P46	358.947691	340.8179998	38.76913595	361.896925	9.850687981
	P47	75.80592895	75.24208403	72.29903698	72.37821198	90.01108479
	P48	57.06797409	57.5266881	57.39999795	57.69292903	75.31407619
	P49	120.7020991	122.7205071	122.7205071	111.9200342	129.3395798
	P50	93.43427801	94.0051198	93.92194915	93.91225505	113.7364788
	P51	33.23734617	33.57928205	33.78634214	33.81907105	50.96383381
	P52	60.98545122	61.79449296	61.49921107	61.46381116	78.77859783
	P53	917.8916519	915.804148	918.0848951	915.8455608	935.6233439
	P54	358.947691	340.8179998	38.76913595	361.896925	9.850687981
	P55	140.248518	139.902606	139.7290549	140.1551671	158.2280879
Nonconvex Large Scale with 200 variables	P41	857.0965228	858.453763	862.0762868	859.221395	879.2447381
	P42	1821.448451	1936.892851	1825.738936	1826.521681	1849.44084
	P43	2156.047382	2156.041017	2157.751231	2156.760214	2185.62943
	P44	2691.623174	2692.723938	2698.771745	2691.550648	2708.264858
	P45	784.9884422	787.202239	787.8259768	787.801333	806.652524
	P46	5579.026889	5568.732214	1471.936929	5672.939605	251.4971209
	P47	1059.023876	1060.836025	1078.361139	1084.608362	1076.036092
	P48	805.6136961	807.4834359	810.8531868	818.7503982	833.4470501
	P49	1666.825887	1686.82367	1686.82367	1669.217375	1694.619697
	P50	1609.523945	1621.387023	1611.306507	1613.287288	1633.667805
	P51	443.4699359	439.9732959	444.5235229	442.949214	461.9291141
	P52	873.3499069	901.0368209	877.4669721	876.510994	899.1966832
	P53	13683.90677	13565.5557	13637.44187	13591.0403	13698.78747
	P54	15232.01185	15144.22455	15173.45608	15130.35306	15150.1776
	P55	2156.047382	2156.041017	2157.751231	2156.760214	2185.62943

Table 9 Success percentage for nonsmooth test problems obtained using the weak subgradient algorithm

Type of Prob.	**Criteria** $\frac{f_{WSA}-f*}{1+\lvert f*\rvert}$	f^{c}_{wsa}	f^{dim}_{wsa}	f^{dyn1}_{wsa}	f^{dyn2}_{wsa}	f^{dyn3}_{wsa}	f_{wsa}
Convex	$< 5 \times 10^{-5}$	%57	%62	%81	%62	%52	%76
	$< 10^{-3}$	%76	%71	%81	%76	%57	%86
	$< 10^{-2}$	%90	%81	%95	%95	%67	%91
Nonconvex Small Scale	$< 5 \times 10^{-5}$	%79	%63	%100	%79	%63	%58
	$< 10^{-3}$	%79	%69	%100	%79	%74	%90
	$< 10^{-2}$	%95	%95	%100	%95	%95	%100
Nonconvex	$< 5 \times 10^{-5}$	%27	%40	%80	%20	%14	%47
Large Scale	$< 10^{-3}$	%47	%47	%80	%40	%20	%74
with 50	$< 10^{-2}$	%74	%60	%94	%80	%60	%94
Nonconvex	$< 5 \times 10^{-5}$	%20	%40	%60	%27	%14	%47
Large Scale	$< 10^{-3}$	%27	%47	%60	%34	%14	%54
200	$< 10^{-2}$	%47	%54	%87	%66	%34	%80

Table 10 Solutions obtained for EXP problem using the weak subgradient algorithm with different stepsizes

Step Size	x	f_{WSA}
α^{c}	$(0.441949, 0.05803418, -0.01980376, 0.01978516, -0.0397822715)$	-0.0121952
α^{dim}_{k}	$(-5, -0.21066956, 0.471328389, 0.034778302, -0.00888853819)$	-5.7516929
α^{dyn2}_{k}	$(0.41954116, 0.081719557, -0.03385298, 0.03549992, -0.0361275123)$	-0.0622914
α^{dyn3}_{k}	$(0.4456131, 0.05439745, -0.02348399, 0.023480678, -0.0234753861)$	-0.0024076

Table 11 Comparison of the weak subgradient algorithm with other algorithms.

Type of Prob.	**Name of Prob.**	f_{wsa}	f_{dgm} **[11]**	f_{ncvx} **[26]**	f_{asm} **[10]**	f_{sub1} **[10]**	f_{sub2} **[10]**	$f_{dc-ncvx}$ **[27]**	f_{vmnc} **[71]**
Convex	P1	115.3705	115.7064	NA	115.70646	231.41148	197.90843	NA	NA
	P2	**0.00798035**	0.0356	NA	0.14117	0.52659	0.47843	NA	NA
	P3	680.6206	680.6301	NA	680.63056	1,283.63922	738.18848	NA	680.63011
	P4	24.318913	24.3062	NA	24.30702	1,083.73783	242.21861	NA	24.306706
	P5	132.768464	93.9073	NA	93.93033	6,353.92360	3,597.49593	NA	133.73418
	P6	3.694879	3.7035	NA	3.70348	1.106.25628	3.72973	NA	NA
	P7	**0.340174**	0.3987	NA	0.38112	749.33916	172.41626	NA	NA
	P8	-43.9999	-44	-44	-43.99997	-43.94407	-43.99973	NA	-43.999975
	P9	1.9522562	1.9522	1.9522245	1.95222	1.95236	1.95223	NA	1.9522250
	P10	2	NA	2.0000001	NA	NA	NA	NA	2
	P11	-2.99999	NA	-2.9999999	NA	NA	NA	NA	-2.999997
	P12	7.200007	NA	7.2000005	NA	NA	NA	NA	7.2000023
	P13	-1.4142077	NA	-1.4142135	NA	NA	NA	NA	-1.4142133
	P14	-1	NA	-0.9999977	NA	NA	NA	NA	-0.9999925
	P15	22.6005197	NA	22.600162	NA	NA	NA	NA	22.600186
	P16	**-0.96222**	NA	-0.8414078	NA	NA	NA	NA	-0.8414057
	P17	0	NA	1.660e-07	NA	NA	NA	NA	0.898E-05
	P18	0	NA	1.110e-15	NA	NA	NA	NA	0
	P19	0	NA	1.142e-13	NA	NA	NA	NA	0.332E-05
	P20	-8	NA	-7.9999998	NA	NA	NA	-8	-7.999998
	P21	0	NA	1.768e-05	NA	NA	NA	NA	0.201E-05
Nonconvex	P22	0	NA	8.022e-06	NA	NA	NA	9.989E-07	0.949E-10
	P23	-1	NA	-1	NA	NA	NA	-1	-0.9999998
	P24	0	0	NA	0.00492	2.72430	3.48299	NA	NA
	P25	0	0	NA	0.26208	17.49696	0.08947	NA	NA
	P26	3.5991193	3.5997	NA	3.59972	3.60367	3.59974	NA	NA
	P27	0.0042075	0.0042	NA	0.03891	0.04810	0.04761	NA	NA
	P29	-43.99	-44	NA	12.17732	-11.53027	-29.89826	NA	NA
	P30	0.55981367	NA	0.5598163	NA	NA	NA	0.5598159	0.5598184
	P31	9.78893903	NA	9.7857746	NA	NA	NA 6	9.7857727	9.7862324
	P32	2	NA	NA	2	2	2	NA	NA
	P33	0.00012007	NA	5.009e-07	0.4	2.12330	2.13230	9.599E-07	0.320E-07
	P34	**0.00301524**	NA	NA	0.9	54.55650	44.29527	NA	NA
	P35	**-5.75169**	0.0011	NA	0.10373	4.85194	1.75332	NA	0.0001224
	P36	0.00815057	0.0081	NA	0.04273	1.59493	0.09259	NA	NA
	P37	0.00270181	0.0029	NA	0.00318	0.79443	0.21997	NA	NA
	P38	**0.00316036**	0.0125	NA	0.06862	0.10392	0.06876	NA	NA
	P39	0.02734077	0.0223	NA	0.33172	12.30531	5.94889	NA	NA
	P40	0.03515377	0.0349	NA	0.04008	0.59012	0.50676	NA	NA