http://communications.science.ankara.edu.tr

Commun.Fac.Sci.Univ.Ank.Ser. A1 Math. Stat.
Volume 71, Number 2, Pages 377–394 (2022)
DOI:10.31801/cfsuasmas.826316
ISSN 1303-5991 E-ISSN 2618-6470

Research Article; Received: November 15, 2020; Accepted: November 2, 2021

WEAK SUBGRADIENT METHOD WITH PATH BASED TARGET LEVEL ALGORITHM FOR NONCONVEX OPTIMIZATION

Gulcin DINC YALCIN

Department of Industrial Engineering, Faculty of Engineering,
Eskisehir Technical University, Eskisehir, TURKEY

ABSTRACT. We study a new version of the weak subgradient method, recently developed by Dinc Yalcin and Kasimbeyli for solving nonsmooth, nonconvex problems. This method is based on the concept of using any weak subgradient of the objective of the problem at the currently generated point with a version of the dynamic stepsize in order to produce a new point at each iteration. The target value needed in the dynamic stepsize is defined using a path based target level (PBTL) algorithm to ensure the optimal value of the problem is reached. We analyze the convergence and give an estimate of the convergence rate of the proposed method. Furthermore, we demonstrate the performance of the proposed method on nonsmooth, nonconvex test problems, and give the computational results by comparing them with the approximately optimal solutions.

1. Introduction

In this paper, we focus on nonsmooth problems where the objective function is lower locally Lipschitz but not necessarily convex or smooth. Many real-world application such as control theory, machine learning, optimal shape design are nonsmooth optimization problems.

In nonsmooth convex optimization, a subgradient defines the normal vectors of the supporting hyperplane to the graph of the function at the relevant point. Thus, in nonsmooth convex optimization, the projected subgradient methods are well known and the fundamentals of these methods have been investigated by Polyak [50], Ermoliev [23], Shor [53]. The main purpose of a projected subgradient method

2020 *Mathematics Subject Classification.* 90C26, 90C30, 65K05, 49M37.
Keywords. Weak subgradient, nonconvex optimization, nonsmooth optimization, path based target level algorithm.
gdinc@eskisehir.edu.tr
0000-0001-7696-7507.

is to generate a new point by using a subgradient of the function at the current point and a positive stepsize parameter. The projection is not computationally expensive if the constraint set is easy for example box constraints. For the convergence analysis, the selection of the stepsize parameter is significant. The classical stepsize types are a (fixed positive) constant, diminishing, and dynamic stepsize. With the dynamic stepsize, the target value is an estimate of the optimal value of the problem and it can be defined as a constant or it can be updated throughout the projected subgradient method. The constant target value may be greater or lower than the optimal value. Alternatively, the target value may be calculated by a path based target level (PBTL) algorithm, which guarantees that the target value will converge to the optimal value [14, 27, 45, 56].

When the function is nonsmooth and nonconvex, various definitions of subgradients are used such as Clarke's subgradient [18] and weak subgradient [3, 4]. Clarke's subgradient is used in nonsmooth, nonconvex (unconstrained or only box constrained) optimization problems, and employed in various methods such as bundle-type methods (see, e.g., [24, 29, 30, 36, 41]), gradient sampling algorithm (see, e.g., [16, 19, 39]), variable metric method (see, e.g., [55]), trust region method (see, e.g., [1, 21, 31, 52]), cutting planes (see, e.g., [25]), proximal algorithms (see, e.g., [9, 11, 12, 48]), quasi-Newton method (see, e.g., [20, 40]). In these methods, the descent directions are usually computed by solving a subproblem which may be quadratic.

Besides subgradient based methods, smoothing methods are also proposed in literature to solve some class of nonsmooth optimization problems. In these methods, the nonsmooth function is approximated by a smooth function, then the smooth function is optimized. The nonsmooth function may be convex (see, e.g., [8, 10, 13, 47, 54]), convex composite(see, e.g., [15]), or nonconvex (see, e.g., [10, 17]).

In addition to these methods, for solving nonsmooth, nonconvex optimization problems, the weak subgradient method [22] is the first to use weak subgradients which have vector and scalar parts, corresponding the supporting conic surfaces to the graph of the function at the relevant point. The weak subgradient method is a generalization of projected subgradient methods, and a convergence analysis of it is investigated with various stepsize parameters: constant and diminishing as well as three types of dynamic.

The aim of this paper is to propose a new version of the weak subgradient method that uses a stepsize parameter computed with PBTL algorithm. Then, the convergence properties and the convergence rate of the proposed method are also investigated. We approximately compute the weak subgradient of the function at the relevant point with the algorithm using the theorem [22, Theorem 2.8] which establish the relation between the directional derivative and weak subdifferential. Additionally, we test the performance of the method on nonsmooth, nonconvex test problems from the literature.

The rest of the paper is organized as follows. Section 2 gives the main properties of the weak subdifferentials and the algorithm for the approximate computing of the weak subgradient is presented. In section 3, we give the convergence properties and convergence rate of the weak subgradient method with PBTL algorithm. Section 4 gives the computational results. In section 5 we draw some conclusions.

2. Preliminaries

In this section, we explain the weak subdifferential and the approximate computing of the weak subgradient.

2.1. **Weak Subdifferentials.**

In this section, we give the definition of the weak subdifferentials and some properties related to this study (see [3, 4, 22, 33, 34]).

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

$$f(x) \geq f(\bar{x}) + \langle v, x - \overline{x}\rangle - c\|x - \bar{x}\|, \quad \forall x \in \mathbb{S}. \tag{1}$$

The set

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

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

As a result of the definition of the weak subgradient, a continuous (superlinear) and concave function is obtained as follows

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

where $x \in \mathbb{S}, g(\bar{x}) = f(\bar{x})$, and $(v, c) \in \partial^w_{\mathbb{S}} f(\bar{x})$. In addition, the hypograph of this function $g(x)$ is a cone and thus supports the epigraph of the function $f(x)$ at the point $(\bar{x}, f(\bar{x}))$.

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

$$f(x) - f(\bar{x}) \geq f'(\bar{x}; x - \bar{x}) \quad \textit{for all } x \in \mathbb{S} - \{\bar{x}\}. \tag{2}$$

When Assumption 1 holds, the following equation

$$f'(\bar{x}; h) = \max\{\langle v, h\rangle - c\,\|h\| : \ (v, c) \in \partial^w_{\mathbb{S}} f(\bar{x}), \|v\| + c \leq M\}, \ \forall h \in \mathbb{R}^n$$

explains the relation between the weak subdifferential $\partial^w_{\mathbb{S}} f(\bar{x})$ and the directional derivative $f'(\bar{x}; h)$ (see [22, Theorem 2.8]), where M is a positive number. The relation plays an important role in the approximation of the weak subgradients.

In addition, it is known that the weak subdifferential of a function is convex and closed (see [33, Theorem 2.4]), and also compact (see [22, Theorem 2.9]). The property of compactness is handled by limiting the scalar part of weak subgradient c with an upper bound L and thus the norm of the vector part of the weak subgradient v is also bounded with an uper bound D. It means that $\partial^w_{\mathbb{S}_L} f(\bar{x})$ is nonempty for

$c \leq L$ and with the number $D > 0$, $\|v\| \leq D$ for all $(v, c) \in \partial^w_{\mathbb{S}_L} f(\bar{x})$. This property of the weak subgradient is essential for both the approximation of the weak subgradients and the convergence analysis of the weak subgradient method.

2.2. **Approximation of Weak Subgradients.** Dinc Yalcin and Kasimbeyli [22] presented an algorithm which makes use of the relation between the directional derivative and weak subgradients, and also the compactness property of the weak subdifferential and, in addition, utilizes the discrete gradient method given by [6]. The algorithm numerically computes the weak subgradient of a function at a given point. Note that the approximation is computed more properly when the value of L which is the upper limit of the scalar part of the weak subgradient c is defined large enough. In addition, throughout this work Assumption 1 holds. We briefly explain the method.

Let us consider the set $G = \{e = (e_1, e_2, ..., e_n) \in \mathbb{R}^n : |e_j| = 1, j = \bar{1,n}\}$ and generate the n vectors $e^j(\alpha) = (\alpha e_1, \alpha^2 e_2, ..., \alpha^j e_j, 0, ..., 0)$, $j = \bar{1,n}$ where $e = (e_1, e_2, ..., e_n) \in G$ and $\alpha \in (0, 1]$ is a fixed number. Then, the equation $f'(\bar{x}; e^j(\alpha)) = \langle \bar{v}, e^j(\alpha) \rangle - \bar{c} \left\| e^j(\alpha) \right\|$ is constructed by the relation between the directional derivative and the weak subdifferential. In addition, with using the compactness of the weak subdifferential, the set $V_{\bar{c}} = \{v \in \mathbb{R}^n : (v, \bar{c})\}$ is obtained for the particular $\bar{c} \leq L$. Thus, the weak subgradient $(\bar{v}, \bar{c})$ exists, where $\bar{v} \in V_{\bar{c}}$. Note that L may be defined as the lower Lipschitz constant.

Due to the compactness of the weak subdifferential and the relation with the directional derivative, a weak subgradient $(\bar{v}, \bar{c})$ that satisfies the equation $f'(\bar{x}; e^j(\alpha)) = \langle \bar{v}, e^j(\alpha) \rangle - \bar{c} \left\| e^j(\alpha) \right\|$ exists, where $\bar{v} \in V_{\bar{c}}$ defined as $V_{\bar{c}} = \{v \in \mathbb{R}^n : (v, \bar{c})\}$ for the particular $\bar{c}$. Note that $\bar{c}$ can be taken less or equal to the lower Lipschitz constant L.

Let take any $e \in G$, and let define $\lambda > 0, \alpha > 0$ and given any $\bar{c}$ and generate the points where the zeroth point is the current point $x^0 = \bar{x}$ and the others are obtained as $x^j = x^0 + \lambda e^j(\alpha), \quad j = \bar{1,n}$. Furthermore, the points are easily generated by $x^j = x^{j-1} + (0, \ldots, 0, \lambda \alpha^j e_j, 0, \ldots)$ for every $j = \bar{1,n}$. After that, the vector $v(e, \alpha, \lambda) \in \mathbb{R}^n$ with the coordinates

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

is defined and with the given numbers, we can state the set $W(e, \alpha) = \{(w, \bar{c}) \in \mathbb{R}^n \times C : \exists(\lambda_k \to +0, k \to +\infty), w = \lim_{k \to \infty} v(e, \alpha, \lambda_k)\}$. Finally, the set $W(e, \alpha)$ is a subset of weak subdifferential, $W(e, \alpha) \subset \partial^w_{\mathbb{S}_L} f(\bar{x}) \quad \forall \alpha \in (0, \alpha_0]$ (see [22, Proposition 3.5], also see [22, Proposition 3.1], [22, Proposition 3.3], [22, Corollary 3.4] for more details).

By using the construction given above, Algorithm 1 is constructed in [22] as follows.

Algorithm 1 Approximate computing of the weak subgradient $(v,c) \in \partial^w_{\mathbb{S}_L} f(\bar{x})$.

Let $e \in G = \{e = (e_1, e_2, ..., e_n) \in \mathbb{R}^n : |e_j| = 1, j = \overline{1,n}\}$ and $\lambda > 0, \alpha \in (0,1]$, $\bar{x} \in \mathbb{S}$, and $L > 0$ sufficient large.
Define $e^j(\alpha) = (e_1\alpha, e_2\alpha^2, ..., e_j\alpha^j, 0, ..., 0), j = \overline{1,n}$.
Choose a number $0 < c < L$.
Let $x^0 = \bar{x}$.
$j \leftarrow 1$.
while $j \leq n$ **do**
    $x^j = x^0 + \lambda e^j(\alpha)$,
    $v_j = \frac{f(x^j) - f(x^{j-1})}{\lambda \alpha^j e_j} + \frac{c}{e_j}$,
    $j \leftarrow j + 1$.
end while

3. Weak Subgradient Method with Path Based Target Level (Pbtl) Algorithm

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

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbb{S} \end{array} \tag{3}$$

where $f : \mathbb{S} \to \mathbb{R}$ is a lower locally Lipschitz function not necessarily convex and smooth. $\mathbb{S} \subset \mathbb{R}^n$ defines the box constraints $\mathbb{S} = \{x \in \mathbb{R}^n : l \leq x \leq u\}$, where l and u shows the lower and upper bounds, respectively.

We present the weak subgradient method with the PBTL algorithm for solving Problem (3). The process of weak subgradient method at every iteration k is as follows:

$$x_{k+1} = P_{\mathbb{S}}(x_k - \alpha_k v_k). \tag{4}$$

Here, $P_{\mathbb{S}}$ denotes projection on the set $\mathbb{S}$, $(v_k, c_k) \in \partial^w_{\mathbb{S}_L} f(x_k)$ is the weak subgradient and the parameter α_k is a positive stepsize. Since the set consists of box constraints, the projection is simple.

Some notations is used through this section. x^* and f^* denote a critical point and the critical value of the problem (3) in the sense of weak subdifferential, respectively. We assume that positive numbers D and L exists satisfying

$$\|v_k\| \leq D, \tag{5}$$

$$c_k \leq L, \tag{6}$$

for all $(v_k, c_k) \in \partial^w_{\mathbb{S}_L} f(x_k)$ for all $x_k \in \mathbb{S}$. The diameter of $\mathbb{S}$ is denoted by the notion $d_{\mathbb{S}} = diam(\mathbb{S}) = \max_{x_1, x_2 \in \mathbb{S}} \|x_1 - x_2\|$. Then

$$\|x_k - x^*\| \leq d_{\mathbb{S}}, \tag{7}$$

where $\|\cdot\|$ is the Euclidean norm.

The dynamic stepsize is generally defined as

$$\alpha_k = \gamma_k \frac{f(x_k) - f_k^{lev} - c_k d_{\mathbb{S}}}{\|v_k\|^2}, 0 < \underline{\gamma} \leq \gamma_k \leq \bar{\gamma} < 2, \tag{8}$$

where the target value f_k^{lev} is an estimate of f^*. The convergence analysis for the various selections of f_k^{lev} is given in [22]. When these selections of f^{lev} are defined constantly, greater or lower value of f^{lev} than the optimal value f^* occurs. In this circumstances, the convergence depends on f^{lev} and the difference $(f^* - f^{lev})$, respectively. When f_k^{lev} is updated during the algorithm with the procedure $f_k^{lev} = \min_k\{f(x_k)\} - \delta_k$ and the parameter δ_k is computed, regardless of whether or not the current iteration is better than f_k^{lev}, the upper limit of δ_k has an impact on the convergence.

In this paper, we analyze the weak subgradient method with a new dynamic stepsize (8), where f_k^{lev} is defined by the PBTL algorithm given in [14, 27, 45, 56] to ensure $f_k^{lev} \to f^*$. The pseudocode is given in Algorithm 2.

The algorithm decreases the δ_l parameter only in Steps 14-16 if the length of the path σ_k travelled by iterates for all $k < k_{l+1}$ exceeds the prescribed upper bound R; otherwise, the parameter remains the same. Decreasing δ_l means increasing the target level f_k^{lev}. σ_k is reset when a new point is generated with sufficient descent of the objective function.

We begin the convergence analysis with the following lemma without proof which gives a general inequality between the generated points and the critical point that is true for all stepsizes (also, see e.g. [2, 26, 32, 37, 38, 45, 46, 51] for other subgradient methods) This lemma is essential for the subsequent convergence analysis.

Lemma 1. [22, Lemma 2] Let $\{x_k\}$ be the sequence generated by the weak subgradient method. Then for all $k \geq 0$, we have

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

We start with a lemma which explains that if δ_l is nondiminishing, then the target values f_k^{lev} are updated infinitely through iterations which means $\inf_{k\geq 0} f(x_k) = -\infty$. The lemma holds true regardless of whether the computation of the weak subgradient is exact or approximate.

Lemma 2. *Algorithm 2 generates infinitely many values of l which means $l \to \infty$. Thus we have either $\inf_{k\geq 0} f(x_k) = -\infty$ or $\lim_{l\to\infty} \delta_l = 0$ for the sequence $\{x_k\}$ generated by the weak subgradient method with the PBTL algorithm.*

Proof. Assume that l takes only a finite number of values, let $\bar{l} > 0$ be the upper bound of l. In this case, we have

$$\sigma_k + \alpha_k\|v_k\| \leq \sigma_k + \alpha_k D = \sigma_{k+1} \leq R$$

Algorithm 2 Weak subgradient method with PBTL algorithm

1: Select a starting initial solution $x_0 \in \mathbb{X}$, and $\delta_0 > 0, R > 0$, let the incumbent solution be $x^{best} = x_0$, and $\sigma_0 = 0, f_{-1}^{rec} = \infty$.
2: Define tolerance tol, and let iteration counter $k = 0$ and $l = 0, k_l = 0$.
3: **while** $\delta_l > tol$ **do**
4:     Calculate $f(x_k)$.
5:     **if** $f(x_k) < f_{k-1}^{rec}$, **then**
6:         Set $f_k^{rec} = f(x_k)$, $x^{best} = x_k$
7:     **else**
8:         Set $f_k^{rec} = f_{k-1}^{rec}$.
9:     **end if**
10:     **if** $f(x_k) < f_{k_l}^{rec} - \frac{\delta_l}{2}$, **then**
11:         Set $k_{l+1} = k, \sigma_k = 0, \delta_{l+1} = \delta_l, l = l + 1$,
12:         Go to 17.
13:     **end if**
14:     **if** $\sigma_k > R$, **then**
15:         Set $k_{l+1} = k, \sigma_k = 0, \delta_{l+1} = \frac{\delta_l}{2}, l = l + 1$.
16:     **end if**
17:     Set $f_k^{lev} = f_{k_l}^{rec} - \delta_l$.
18:     Compute a weak subgradient $(v_k, c_k) \in \partial_{\mathbb{S}_L}^{w} f(x_k)$ of f at x_k via Algorithm 1 in Sect. 2.2.
19:     Calculate x_{k+1} via (4) and (8).
20:     $\sigma_{k+1} = \sigma_k + \alpha_k \|v_k\|$.
21:     $k \leftarrow k + 1$.
22: **end while**

from (5) and Step 20 for all $k \geq k_{\bar{l}}$. This would mean that $\lim_{k\to\infty} \alpha_k = 0$, which is impossible. Since for all $k \geq k_{\bar{l}}$, from Step 17, we have

$$f(x_k) - f_k^{lev} \geq \delta_{\bar{l}}. \tag{9}$$

Furthermore, c_k is chosen less than $\frac{f(x_k) - f_k^{lev}}{d_{\mathbb{S}}}$ since the stepize is a positive parameter. Thus, with (9) and the way of choosing the vale of c_k, we have

$$\alpha_k = \gamma_k \frac{f(x_k) - f_k^{lev} - c_k d_{\mathbb{S}}}{\|v_k\|^2} > 0.$$

This implies that for all $k \geq k_{\bar{l}}$, the stepsize α_k is bounded below with a positive value which means $\lim_{k\to\infty} \alpha_k > 0$. As a consequence l cannot be finite: $l \to \infty$.

Since l goes to infinite, there should be a limit $\delta = \lim_{l\to\infty} \delta_l$. If $\delta = 0$, then $\lim_{l\to\infty} \delta_l = 0$. Otherwise, let l_0 is large enough so that for all $l \geq l_0$, we have $\delta_l = \delta$ from 10-13 and 14-16 and

$$f_{k_{l+1}}^{rec} - f_{k_l}^{rec} \leq -\frac{\delta}{2},$$

implying that $\inf_{k\geq 0} f(x_k) = -\infty$. □

Remark 1. *Algorithm 2 is terminated when δ_l is less than tol. According to Lemma 2, if the function f goes to negative infinity, then while l goes to infinity, δ_l has a limit point δ. In this case, Algorithm 2 runs infinite iterations since the stopping condition $\delta_l \leq tol$ cannot be hold. Therefore, another termination rule such as a time limit or an iteration limit may be used to prevent this situation.*

The convergence property of the weak subgradient method with the PBTL algorithm is given in the following proposition.

Proposition 1. *For the sequence $\{x_k\}$ generated by the weak subgradient method with the PBTL algorithm, we have*

- (a) *if* $\lim_{l\to\infty} \delta_l > 0$, *then*
$$\inf_{k\geq 0} f(x_k) = -\infty,$$
- (b) *if* $\lim_{l\to\infty} \delta_l = 0$, *then*
$$\inf_{k\geq 0} f(x_k) = f^*.$$

Proof. If $\lim_{l\to\infty} \delta_l > 0$, according to Lemma 2, we have $\inf_{k\geq 0} f(x_k) = -\infty$. Thus the proof is completed for part (a).

Now, we prove part (b).

Let ψ be the set of l given by
$$\psi = \Big\{ l | \delta_l = \frac{\delta_{l-1}}{2}, l \geq 1 \Big\}.$$
We obtain
$$\sigma_k = \sigma_{k-1} + \alpha_{k-1}\|v_{k-1}\| = \sum_{j=k_l}^{k-1} \alpha_j \|v_j\|$$
from Steps 10-16 and 20 . When the length of the path becomes greater than the upper value $\sum_{j=k_l}^{k-1} \alpha_j D > \sum_{j=k_l}^{k-1} \alpha_j \|v_j\| > R$ at Steps 14-16, k_{l+1} becomes equal to iteration number $k_{l+1} = k$ where $l+1 \in \psi$. Thus, the sum gives
$$\sum_{j=k_{l-1}}^{k-1} \alpha_j > \frac{R}{D} \quad \forall l \in \psi,$$
and, since the cardinality of ψ is infinite, we have the inequality,
$$\sum_{j=0}^{\infty} \alpha_j \geq \sum_{l\in\psi} \sum_{k_{l-1}}^{k-1} \alpha_j > \sum_{l\in\psi} \frac{R}{D} = \infty. \tag{10}$$

Now, assume to contrary that there exists some $\varepsilon > 0$
$$\inf_{k\geq 0} f(x_k) > f^* + \varepsilon,$$

$$\inf_{k\geq 0} f(x_k) - \varepsilon > f^*.$$

Since $\lim_{l\to\infty} \delta_l = 0$, let $\bar{l}$ be large enough so that there exits some ε such that $\delta_l \leq \varepsilon$ for all $l \leq \bar{l}$. Thereby for all $k \leq k_{\bar{l}}$ we obtain,

$$f_k^{lev} = f_{k_l}^{rec} - \delta_l \geq \inf_{k\geq 0} f(x_k) - \varepsilon > f^*. \tag{11}$$

By using the inequality obtained in (11) and by Lemma 1, and in addition, with assumption (5), the diameter of $\mathbb{S}$ given in (7), the dynamic stepsize (8), and finally using the fact that $\gamma_k < 2, \gamma_k^2 \leq \gamma_k$, $0 < \underline{\gamma} < \gamma_k \leq \bar{\gamma} < 2$, the following inequality is obtained

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \underline{\gamma}(2-\underline{\gamma})\frac{(f(x_k) - f_k^{lev} - c_k d_{\mathbb{S}})^2}{D^2}. \tag{12}$$

By summing these inequalities over $k \geq k_{\bar{l}}$, we have

$$\|x_{k+1} - x^*\|^2 \leq \|x_{k_{\bar{l}}} - x^*\| - \frac{\underline{\gamma}(2-\underline{\gamma})}{D^2}\sum_{k=k_{\bar{l}}}^{\infty}(f(x_k) - f_k^{lev} - c_k d_{\mathbb{S}})^2. \tag{13}$$

Due to (8) and (10), the last term $\sum_{k=k_{\bar{l}}}^{\infty}(f(x_k) - f_k^{lev} - c_k d_{\mathbb{S}})^2$ of the inequality (13) goes to infinity. Then, the relation cannot hold true. Thus, we obtain the contradiction. □

Now, we give a convergence rate analysis.

Proposition 2. *If the weak subgradient method with the PBTL algorithm terminates after a finite number of K iterations, then K is the largest positive integer such that*

$$\sum_{k=0}^{K-1}(\delta_k - Ld_{\mathbb{S}})^2 \leq \frac{D^2}{\underline{\gamma}(2-\bar{\gamma})}\|x_0 - x^*\|^2$$

and we have

$$\inf_{0\leq k\leq K} f(x_k) \leq f^* + \delta_0.$$

Proof. Assume to the contrary that

$$f(x_k) \geq f^* + \delta_0 \tag{14}$$

for all $k = 0, ..., K$.

Since $f_k^{lev} = \min_{0\geq j\geq k} f(x_j) - \delta_k$ and $\delta_k \leq \delta_0$ for all $k \geq 0$, with (14) we have

$$f_k^{lev} \geq f^* - \delta_0 \geq f^* - \delta_k \geq f^* \tag{15}$$

for all $k = 0, ..., K$.

Hereby, by using the inequality $f_k^{lev} \geq f^*$ obtained in (15) and by Lemma 1, and in addition, with the diameter of $\mathbb{S}$ given in (7), the definition of the dynamic stepsize (8),$0 < \underline{\gamma} < \gamma_k \leq \bar{\gamma} < 2$ (similar to the Proposition 1), we get the inequality (12).

When we combine the inequality (12) using the fact $f(x_k) - f_k^{lev} \geq \delta_k \forall k$ and $c_k \leq L$ given in (6), for all $k \leq K$ the following inequality

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \underline{\gamma}(2 - \underline{\gamma})\frac{(\delta_k - Ld_{\mathbb{S}})^2}{D^2}$$

is obtained.

By summing these inequalities over $k = 0, ...K$, we have

$$\|x_{K+1} - x^*\|^2 \leq \|x_0 - x^*\|^2 - \frac{\underline{\gamma}(2 - \underline{\gamma})}{D^2}\sum_{k=0}^{K}(\delta_k - Ld_{\mathbb{S}})^2.$$

The last relation cannot hold for sufficiently large K because of the compactness of the set $\mathbb{S}$. Thus, it implies

$$\sum_{k=0}^{K-1}(\delta_k - Ld_{\mathbb{S}})^2 \leq \frac{D^2}{\underline{\gamma}(2 - \bar{\gamma})}\|x_0 - x^*\|^2.$$

$\square$

Remark 2. *Let the Assumption 1 hold true. Then, there exists a weak subgradient* $(v_k, c_k) \in W(e, \alpha) \subset \partial^w_{\mathbb{S}_L} f(x_k)$ *and thus, we have*

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

for all $k \geq 0$*, which plays an important role in proving Lemma* 1*. Since Lemma* 1 *is essential to prove the results on the propositions of convergence analysis and convergence rate, all the results of this section are valid if the weak subgradient is computed via Algorithm 1.*

4. Computational Results

In this section, we verify the performance and analyze the efficiency of the weak subgradient method with the PBTL algorithm by solving completely 49 nonsmooth, nonconvex test problems, of which 19 are small scale, P-SS, (P1-P19) with 2 to 10 decision variables and 15+15 are large scale (P20-P34), with 50, P-LS-50, (shown as P20-50 to P34-50) and 200, P-LS-200 (shown as P20-200 to P34-200) decision variables, respectively. Table 1 shows the properties of the test problems, including the names given in the literature and references to where they were taken from, the variable numbers, n, and the optimal values of the problems, f^*. Note that the optimal values of some problems are approximate, P12 and P13 are the L_1 version of the Rosenbrock and Wood functions, respectively, and P21 is the nonsmooth version of the Brown function.

Table 1: Nonsmooth nonconvex test problems

Small Scale Problem	**n**	f^*	**Large Scale Problem**	**n**	f^*
P1 Crescent [35]	2	0	P20 Active faces [28]	any	0
P2 Mifflin 2 [44]	2	-1	P21 Brown function [29]	any	0
P3 WF [42]	2	0	P22 Chained crescent I [29]	any	0
P4 SPIRAL [42]	2	0	P23 Chained crescent II [29]	any	0
P5 EVD52 [42]	3	3.5991193	P24 Problem 6 in [43]	any	0
P6 PBC3 [42]	3	0.0042021427	P25 Problem 17 in [43]	any	0
P7 Bard [42]	3	0.050816327	P26 Problem 19 in [43]	any	0
P8 Polak 6 [49]	4	-44	P27 Problem 20 in [43]	any	0
P9 El-Attar [42]	6	0.5598131	P28 Problem 22 in [43]	any	0
P10 Gill [42]	10	9.7857721	P29 Problem 24 in [43]	any	0
P11 Problem 1 [5]	2	2	P30 DC Maxl [5]	any	0
P12 Rosenbrock [5]	2	0	P31 DC Maxlq [7]	any	0
P13 Wood [5]	4	0	P32 Problem 6 in [7]	any	0
P14 EXP [42]	5	0.00012237125	P33 Problem 7 in [7]	any	0
P15 Kow.-Osb. [42]	4	0.0080843684	P34 Chained Mifflin 2 [29]	50	-34.795
P16 OET5 [42]	4	0.0026359735		200	-140.86
P17 OET6 [42]	4	0.0020160753			
P18 PBC1 [42]	5	0.022340496			
P19 EVD61 [42]	6	0.034904926			

The constraint set is $\mathbb{S} = \{x | x_i \in [-5, 5] \quad i = 1, ..., n\}$ in the problems, however if any component of the optimal solution is not in this interval, then the constraint set is updated as $[x_i^* - 5, x_i^* + 5] \quad i \in \{1, ..., n\}$. In addition, the starting points of the problems needed in the algorithm are the same in reference to the corresponding problems.

We code the weak subgradient method with the PBTL algorithm in the Python programming language and carry out numerical experiments on MacBook Pro with 2.5GHz Intel Core i7 processor and with 16GB 1600 MHz DDR3 RAM. The algorithm is terminated if δ_k becomes less than $tol = 0.001$ or the CPU (s) time reaches $3600s$ for all test problems. δ_0 is defined as $|f(x_0)|$. The prescribed upper bound R is defined as $100, 5000$ and 50 for P-SS, P-LS-50 and P-LS-200, respectively. For P7, R is defined as 10000. The parameters α and λ is set as 1 and 0.001 for the approximate computing of the weak subgradient via Algorithm 1, respectively. The upper bound $\bar{c}_k$ of the scaler parameter c_k of the weak subgradient is $\bar{c}_k = \frac{f(x_k) - f_k^{lev}}{d_{\mathbb{S}}}$ to ensure the positiveness of the stepsize. The scaler parameter c_k is defined $c_k = \bar{c}_k * 0.5$ to compute the vector part v_k of the subgradient in Algorithm 1.

The computational results, the CPU times, and iteration numbers of nonsmooth, nonconvex test problems obtained via weak subgradient method with the PBTL algorithm are given in Table 2, where the following notations are used:

- $WSM-Path$: The weak subgradient method with the PBTL algorithm.
- $WSM-Dyn$: The weak subgradient method with dynamic stepsize with dynamic f_k^{lev} from [22].
- f_{wsa}^{path}: The best value of the objective function, computed using $WSM-Path$.
- f_{wsa}^{dyn}: The best value of the objective function, computed using $WSM-Dyn$.
- $iter$: The number of iterations at which the weak subgradient method with the PBTL algorithm is terminated.

Table 2 compares the results with the (approximate) optimal solutions obtained so far and the results obtained by $WSM-Dyn$. The better results are shown in bold. The results show that $WSM-Path$ outperforms $WSM-Dyn$ in 29 out of 49 test problems and two algorithms find the same value in 7 out of 49 test problems.

$$\frac{f-f^*}{1+|f^*|} \leq \varepsilon. \tag{16}$$

We evaluate the results with the evaluation criteria (16) given above, where f is the results obtained by the relevant method (f_{wsm}^{dyn} or f_{wsm}^{path} in this paper). When the evaluation criteria of each result is less than ε, the results is accepted as successful. The successful percentage is computed by the total number of successful results over the total number of the problems. We take ε as $10^{-2}, 10^{-3}$, and 5×10^{-5}. We summarize the results in Table 3.

If we take the $\varepsilon = 10^{-2}$, then $WSA-Path$ reaches the optimal value with $\%95, \%60$ and $\%40$ percentages for P-SS, P-LS-50, and P-LS-200, respectively. Similar, If we take the $\varepsilon = 10^{-3}$, then $\%95, \%46$ and $\%33$ percentages are obtained. Last, if we take the $\varepsilon = 5\times10^{-5}$, then $\%68, \%40$ and $\%27$ percentages are observed. Additional, $WSA-Path$ finds better solution for P14 (EXP). Moreover, $WSA-Path$ outperforms the successful percentage of $WSM-Dyn$.

5. Conclusion

In this paper, we propose a new version of the weak subgradient method with the PBTL algorithm ($WSA-Path$). A weak subgradient of the current point with a version of dynamic stepsize is used to produce a new solution at each iteration, where the weak subgradient is computed with Algorithm 1 using the theorem about the directional derivative and weak subdifferential. The target level in the dynamic stepsize is computed with the PBTL algorithm. Then, the difference with the PBTL algorithm compared to the other dynamic stepsizes is the method of defining the target level to ensure $f_k^{lev} \to f^*$. We give the convergence analysis and converge rate of the method. Furthermore, we show the tests performed using the method

Table 2: Computational results for nonsmooth test problems for test problems

Prob.	f^*	f_{wsm}^{dyn}	$WSM - Path$	
			f_{wsm}^{path}	CPU (s)
P1	0	0	0	86.19
P2	-1	-1	-1	80.41
P3	0	**0**	0.00000169	81.30
P4	0	0	0	1.14
P5	3.5991193	3.59984305	**3.59973074**	123.51
P6	0.0042021427	0.00421077	**0.00420479**	429.55
P7	0.050816327	0.0508552	**0.050829**	232.20
P8	-44	-43.99	-43.99	215.58
P9	0.5598131	0.56171104	**0.55993735**	1859.47
P10	9.7857721	9.813723	**9.79246244**	2739.11
P11	2	2	2	118.39
P12	0	0.00015433	**0**	16.50
P13	0	0.0090316	**0**	0.04
P14	0.00012237125	-0.0024076	$\mathbf{-6}$	552.60
P15	0.0080843684	0.00815057	**0.00810742**	114.80
P16	0.0026359735	0.00325996	**0.0026544**	367.68
P17	0.0020160753	0.00317971	**0.00209686**	353.22
P18	0.022340496	0.11826176	**0.02251701**	155.31
P19	0.034904926	**0.03578041**	0.07816864	12.48
P20-50	0	0.004235249	**0**	90.75
P21-50	0	0.01909278	**0**	1656.90
P22-50	0	0.045976722	**0**	3606.05
P23-50	0	0.0048727756	**0**	3636.80
P24-50	0	0.004071199	**0.00300322**	3544.41
P25-50	0	**0**	0.87361276	200, 47
P26-50	0	**0.002417618**	0.18014279	4106, .27
P27-50	0	**0.0073205**	0.103582595	552.50
P28-50	0	**0.000680983**	0.00068109	0.004
P29-50	0	0.012916485	**0.00928192**	3019.05
P30-50	0	2.575630571	**0**	1054.63
P31-50	0	1	1	906.04
P32-50	0	0.028024848	**0.02395819**	3548.23
P33-50	0	**0**	0.07654164	3684.00
P34-50	-34.795	-34.70324	$\mathbf{-34.774069}$	2517.02
P20-200	0	**0.01170317**	0.76033843	27.00
P21-200	0	0.096967	**0**	3600.04

Prob.	f^*	f_{wsm}^{dyn}	$WSM - Path$	
			f_{wsm}^{path}	CPU (s)
P22-200	0	0.662850133	**0**	1211.46
P23-200	0	0.08312041	**0**	3600.07
P24-200	0	0.0093181	**0.00489712**	1698.22
P25-200	0	**0**	0.83829569	544.62
P26-200	0	**0.00852129**	0.0300995	398.30
P27-200	0	0.505147266	**0.2388156**	1018.13
P28-200	0	0	0	0.04
P29-200	0	**0.01205072**	0.02550528	1037.98
P30-200	0	9.54184555	**0.0787**	3600.05
P31-200	0	1	1	152.23
P32-200	0	**0.0061375149**	0.09293225	3600.28
P33-200	0	1.170935921	**0.32948383**	3600.28
P34-200	-140.86	-139.8939	**−140.75363**	3196.39

Table 3: Success percentage of $WSM - Path$ for nonsmooth test problems versus the optimal value and $WSM - Dyn$

Type of Prob.	**Criteria** $\frac{f-f*}{1+\|f*\|}$	f_{wsm}^{dyn}	f_{wsm}^{path}
P-SS	$< 5 \times 10^{-5}$	63%	68%
	$< 10^{-3}$	74%	95%
	$< 10^{-2}$	95%	95%
P-LS-50	$< 5 \times 10^{-5}$	14%	40%
	$< 10^{-3}$	20%	46%
	$< 10^{-2}$	60%	60%
P-LS-200	$< 5 \times 10^{-5}$	14%	27%
	$< 10^{-3}$	14%	33%
	$< 10^{-2}$	34%	40%

on nonsmooth, nonconvex optimization problems. The performance of $WSA - Path$ over the (approximate) optimal values and $WSM - Dyn$ is shown by the computational experiments. Besides $WSM - Path$ shows good performance in reaching the optimal values, it also outperforms $WSM - Dyn$ in 29 out of 49 test problems and the two algorithms find the same value from 7 out of 49 test problems. We intend to investigate the ways of weakening Assumption 1 as a part of our future work. Additionally, we would like to solve other nonsmooth optimization problems, such as those found in machine learning problems.

Declaration of Competing Interests There are no relevant financial or non-financial competing interests to report.

Acknowledgment This study is supported by The Scientific and Technological Research Council of Turkey (TUBITAK) under Grant No. 217M487.

References

[1] Akbari, Z., Yousefpour, R., Peyghami, M. R., A new nonsmooth trust region algorithm for locally Lipschitz unconstrained optimization problems, *Journal of Optimization Theory and Applications, 164* (3) (2015), 733–754, https://dx.doi.org/10.1007/s10957-014-0534-6.

[2] Allen, E., Helgason, R., Kennington, J., Shetty, B., A generalization of Polyak's convergence result for subgradient optimization, *Mathematical Programming, 37* (3) (1987), 309–317.

[3] Azimov, A., Gasimov, R., On weak conjugacy, weak subdifferentials and duality with zero gap in nonconvex optimization, *International Journal of Applied Mathematics, 1* (2) (1999), 171–192.

[4] Azimov, A., Gasimov, R., Stability and duality of nonconvex problems via augmented Lagrangian, *Cybernetics and Systems Analysis, 38* (3) (2002), 412–421, https://dx.doi.org/10.1023/A:1020316811823.

[5] Bagirov, A., A method for minimization of quasidifferentiable functions, *Optimization Methods and Software, 17* (1) (2002), 31–60, https://dx.doi.org/10.1080/10556780290027837.

[6] Bagirov, A. M., A method for minimizing convex functions based on continuous approximations to the subdifferential, *Optimization Methods and Software, 9* (1998), 1–17, https://dx.doi.org/10.1080/10556789808805683.

[7] Bagirov, A. M., Ugon, J., Codifferential method for minimizing nonsmooth DC functions, *Journal of Global Optimization, 50* (1) (2011), 3–22, https://dx.doi.org/10.1007/s10898-010-9569-x.

[8] Beck, A., Teboulle, M., Smoothing and first order methods: A unified framework, *SIAM Journal on Optimization, 22* (2) (2012), 557–580, https://dx.doi.org/10.1137/100818327.

[9] Bolte, J., Sabach, S., Teboulle, M., Proximal alternating linearized minimization for nonconvex and nonsmooth problems, *Mathematical Programming, 146* (1-2) (2014), 459–494, https://dx.doi.org/10.1007/s10107-013-0701-9.

[10] Borges, P., Sagastizábal, C., Solodov, M., A regularized smoothing method for fully parameterized convex problems with applications to convex and nonconvex two-stage stochastic programming, *Mathematical Programming* (2020), 1–33, https://dx.doi.org/10.1007/s10107-020-01582-2.

[11] Boţ, R. I., Csetnek, E. R., An inertial Tsengs type proximal algorithm for nonsmooth and nonconvex optimization problems, *Journal of Optimization Theory and Applications, 171* (2) (2016), 600–616, https://dx.doi.org/10.1007/s10957-015-0730-z.

[12] Bot, R. I., Csetnek, E. R., Nguyen, D. K., A proximal minimization algorithm for structured nonconvex and nonsmooth problems, *SIAM Journal on Optimization, 29* (2) (2019), 1300–1328, https://dx.doi.org/10.1137/18M1190689.

[13] Boţ, R. I., Hendrich, C., A variable smoothing algorithm for solving convex optimization problems, *TOP, 23* (1) (2015), 124–150, https://dx.doi.org/10.1007/s11750-014-0326-z.

[14] Brannlund, U., On Relaxation Methods for Nonsmooth Convex Optimization, PhD thesis, Stockholm: Department of Mathematics, Royal Institute of Technology, 1993.

[15] Burke, J. V., Hoheisel, T., Epi-convergent smoothing with applications to convex composite functions, *SIAM Journal on Optimization, 23* (3) (2013), 1457–1479, https://dx.doi.org/10.1137/120889812.

[16] Burke, J. V., Lewis, A. S., Overton, M. L., A robust gradient sampling algorithm for nonsmooth, nonconvex optimization, *SIAM Journal on Optimization, 15* (3) (2005), 751–779, https://dx.doi.org/10.1137/030601296.

[17] Chen, X., Smoothing methods for nonsmooth, nonconvex minimization, *Mathematical Programming, 134* (1) (2012), 71–99, https://dx.doi.org/10.1007/s10107-012-0569-0.

[18] Clarke, F. H., Optimization and Nonsmooth Analysis, SIAM, 1990.

[19] Curtis, F. E., Que, X., An adaptive gradient sampling algorithm for nonsmooth optimization, *Optimization Methods and Software, 28* (6) (2013), 1302–1324, https://dx.doi.org/10.1080/10556788.2012.714781.

[20] Curtis, F. E., Que, X., A quasi-Newton algorithm for nonconvex, nonsmooth optimization with global convergence guarantees, *Mathematical Programming Computation, 7* (4) (2015), 399–428, https://dx.doi.org/10.1007/s12532-015-0086-2.

[21] Dennis, J. E., Li, S. B. B., Tapia, R. A., A unified approach to global convergence of trust region methods for nonsmooth optimization, *Mathematical Programming, 68* (1-3) (1995), 319–346, https://dx.doi.org/10.1007/BF01585770.

[22] Dinc Yalcin, G., Kasimbeyli, R., Weak subgradient method for solving nonsmooth nonconvex optimization problems, *Optimization, 70* (7) (2021), 1513–1553, https://dx.doi.org/10.1080/02331934.2020.1745205.

[23] Ermoliev, Y. M., Methods of solution of nonlinear extremal problems, *Cybernetics, 2* (4) (1966), 1–14, https://dx.doi.org/10.1007/BF01071403.

[24] 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) (2004), 89–102, https://dx.doi.org/10.1080/10556780410001648112.

[25] Fuduli, A., Gaudioso, M., Giallombardo, G., Minimizing nonconvex nonsmooth functions via cutting planes and proximity control, *SIAM Journal on Optimization, 14* (3) (2004), 743–756, https://dx.doi.org/10.1137/S1052623402411459.

[26] Gasimov, R. N., Augmented Lagrangian duality and nondifferentiable optimization methods in nonconvex programming, *Journal of Global Optimization, 24* (2) (2002), 187–203, https://dx.doi.org/10.1023/A:1020261001771.

[27] Goffin, J. L., Kiwiel, K. C., Convergence of a simple subgradient level method, *Mathematical Programming, 85* (1) (1999), 207–211, https://dx.doi.org/10.1007/s101070050053.

[28] Grothey, A., Decomposition Methods for Nonlinear Nonconvex Optimization Problems, PhD thesis, Citeseer, 2001.

[29] Haarala, M., Miettinen, K., Mäkelä, M. M., New limited memory bundle method for large-scale nonsmooth optimization, *Optimization Methods and Software, 19* (6) (2004), 673–692, https://dx.doi.org/10.1080/10556780410001689225.

[30] Haarala, N., Miettinen, K., Mäkelä, M. M., Globally convergent limited memory bundle method for large-scale nonsmooth optimization, *Mathematical Programming, 109* (1) (2007), 181–205, https://dx.doi.org/10.1007/s10107-006-0728-2.

[31] Hoseini, N., Nobakhtian, S., A new trust region method for nonsmooth nonconvex optimization, *Optimization, 67* (8) (2018), 1265–1286, https://dx.doi.org/10.1080/02331934.2018.1470175.

[32] Hu, Y., Yang, X., Sim, C.-K., Inexact subgradient methods for quasi-convex optimization problems, *European Journal of Operational Research, 240* (2) (2015), 315–327, https://dx.doi.org/10.1016/j.ejor.2014.05.017.

[33] Kasimbeyli, R., Mammadov, M., On weak subdifferentials, directional derivatives, and radial epiderivatives for nonconvex functions, *SIAM Journal on Optimization, 20* (2) (2009), 841–855, https://dx.doi.org/10.1137/080738106.

[34] Kasimbeyli, R., Mammadov, M., Optimality conditions in nonconvex optimization via weak subdifferentials, *Nonlinear Analysis: Theory, Methods & Applications, 74* (7) (2011), 2534–2547, https://dx.doi.org/10.1016/j.na.2010.12.008.

[35] Kiwiel, K. C., Methods of Decent for Nondifferentiable Optimization, Lecture Notes in Mathematics, No. 1133, Amer Chemical Soc 1155 16TH St, NW, Washington, DC 20036, 1985.

[36] Kiwiel, K. C., Restricted step and Levenberg–Marquardt techniques in proximal bundle methods for nonconvex nondifferentiable optimization, *SIAM Journal on Optimization, 6* (1) (1996), 227–249, https://dx.doi.org/10.1137/0806013.

[37] Kiwiel, K. C., Convergence and efficiency of subgradient methods for quasiconvex minimization, *Mathematical Programming, 90* (1) (2001), 1–25, https://dx.doi.org/10.1007/s101070100198.

[38] Kiwiel, K. C., Convergence of approximate and incremental subgradient methods for convex optimization, *SIAM Journal on Optimization, 14* (3) (2004), 807–840, https://dx.doi.org/10.1137/S1052623400376366.

[39] Kiwiel, K. C., Convergence of the gradient sampling algorithm for nonsmooth nonconvex optimization, *SIAM Journal on Optimization, 18* (2) (2007), 379–388, https://dx.doi.org/10.1137/050639673.

[40] Lewis, A. S., Overton, M. L., Nonsmooth optimization via quasi-Newton methods, *Mathematical Programming, 141* (1-2) (2013), 135–163, https://dx.doi.org/10.1007/s10107-012-0514-2.

[41] Lukšan, L., Vlček, J., A bundle-Newton method for nonsmooth unconstrained minimization, *Mathematical Programming, 83* (1-3) (1998), 373–391, https://dx.doi.org/10.1007/BF02680566.

[42] 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.

[43] Lukšan, L., Vlček, J., Ramešová, N., UFO 2002, Interactive System for Universal Functional Optimization, Technická Zpráva 883, 2002.

[44] Mäkelä, M. M., N. P., Nonsmooth Optimization: Analysis and Algorithms with Applications to Optimal Control, World Scientific, 1992.

[45] Nedic, A., Bertsekas, D. P., Incremental subgradient methods for nondifferentiable optimization, *SIAM Journal on Optimization, 12* (1) (2001), 109–138, https://dx.doi.org/10.1137/S1052623499362111.

[46] Nedić, A., Bertsekas, D. P., The effect of deterministic noise in subgradient methods, *Mathematical Programming, 125* (1) (2010), 75–99, https://dx.doi.org/10.1007/s10107-008-0262-5.

[47] Nesterov, Y., Smooth minimization of non-smooth functions, *Mathematical Programming, 103* (1) (2005), 127–152, https://dx.doi.org/10.1007/s10107-004-0552-5.

[48] Pock, T., Sabach, S., Inertial proximal alternating linearized minimization (ipalm) for nonconvex and nonsmooth problems, *SIAM Journal on Imaging Sciences, 9* (4) (2016), 1756–1787, https://dx.doi.org/10.1137/16M1064064.

[49] Polak, E., Mayne, D. Q., Higgins, J. E., Superlinearly convergent algorithm for minmax problems, *Journal of Optimization Theory and Applications, 69* (3) (1991), 407–439, https://dx.doi.org/10.1007/BF00940683.

[50] Polyak, B. T., A general method for solving extremal problems, *Doklady Akademii Nauk, Russian Academy of Sciences, 174* (1) (1967), 33–36.

[51] Polyak, B. T., Minimization of unsmooth functionals, *USSR Computational Mathematics and Mathematical Physics, 9* (3) (1969), 14–29, https://dx.doi.org/10.1016/0041-5553(69)90061-5.

[52] Qi, L., Sun, J., A trust region algorithm for minimization of locally Lipschitzian functions, *Mathematical Programming, 66* (1-3) (1994), 25–43, https://dx.doi.org/10.1007/BF01581136.

[53] Shor, N. Z., Minimization Methods for Non-differentiable Functions (K. Kiwiel and A. Ruszcynski, translate), Heidelberg:Springer-Verlag Berlin, 1985.

[54] Tran-Dinh, Q., Adaptive smoothing algorithms for nonsmooth composite convex minimization, *Computational Optimization and Applications, 66* (3) (2017), 425–451, https://dx.doi.org/10.1007/s10589-016-9873-6.

[55] Vlček, J., Lukšan, L., Globally convergent variable metric method for nonconvex nondifferentiable unconstrained minimization, *Journal of Optimization Theory and Applications, 111* (2) (2001), 407–430, https://dx.doi.org/10.1023/A:1011990503369.

[56] Zhang, P., Bao, G., Path-based incremental target level algorithm on riemannian manifolds, *Optimization, 69* (4) (2020), 799–819, https://dx.doi.org/10.1080/02331934.2019.1671840.