Local Convergence Analysis for Nonisolated Solutions to Derivative-Free Methods of Optimization

Dao Huy Cuong* Pham Duy Khanh† Boris S. Mordukhovich‡ Dat Ba Tran§

July 30, 2024

Dedicated to Professor Alexander Zaslavski in honor of his 65th Birthday

Abstract. This paper provides a local convergence analysis for newly developed derivative-free methods in problems of smooth nonconvex optimization. We focus here on local convergence to local minimizers, which might be nonisolated and hence more challenging for convergence analysis. The main results provide efficient conditions for local convergence to arbitrary local minimizers under the fulfillment of the Kurdyka-Łojasiewicz property.

Key words: derivative-free optimization, nonconvex smooth objective functions, nonisolated local minimizers, finite difference algorithms, local convergent analysis, Kurdyka-Łojasiewicz gradient inequality.

Mathematics Subject Classification (2020) 90C25, 90C26, 90C30, 90C56

1 Introduction

This paper is devoted to the local convergence analysis of *derivative-free methods* for solving unconstrained optimization problems given in the form

$$\text{minimize} \quad f(x) \quad \text{subject to} \;\; x \in \mathbb{R}^n, \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable ($\mathcal{C}^1$-smooth) function, not necessarily convex. In the context of derivative-free optimization (DFO), we assume that only information of the function values is available, but for gradient values we have access to some approximations under certain computation errors. The reader is referred to the excellent

*Group of Analysis and Applied Mathematics, Department of Mathematics, Ho Chi Minh City University of Education, in Chi Minh City, Vietnam. E-mail: cuongdh@hcmue.edu.vn. Research of this author is funded by the Ministry of Education and Training Research Funding under the project B2024-SPS-07.

†Group of Analysis and Applied Mathematics, Department of Mathematics, Ho Chi Minh City University of Education, Ho Chi Minh City, Vietnam. E-mail: khanhpd@hcmue.edu.vn. Research of this author is funded by the Ministry of Education and Training Research Funding under the project B2024-SPS-07.

‡Department of Mathematics, Wayne State University, Detroit, Michigan, USA. E-mail: aa1086@wayne.edu. Research of this author was partly supported by the US National Science Foundation under grant DMS-2204519, by the Australian Research Council under grant DP-190100555, and by Project 111 of China under grant D21024.

§Department of Mathematics, Wayne State University, Detroit, Michigan, USA. E-mail: tranbadat@wayne.edu. Research of this author was partly supported by the US National Science Foundation under grant DMS-2204519.

book by Zaslavski [33] for general aspects of numerical optimization with computational errors, while here we concentrate for specific DFO algorithms in noiseless settings.

It has been well recognized that DFO theory and algorithms have a lot of applications in data science [14, 24], biomedical imaging [30] as well as in infinite-dimensional optimization problems governed by partial differential equations [27], etc. Therefore, derivative-free optimization methods have received much attention, with major developments provided by the *Nelder-Mead simplex method* [28], *trust-region methods* [11], and *finite-difference-based methods* [5, 6, 32]. We refer the reader to the books [4, 10] with the vast bibliographies therein for various developments and historical remarks. Recently, empirical results conducted in [5, 32] show that derivative-free optimization methods based on *finite differences* are more preferable than other state-of-the-art derivative-free optimization methods developed in the literature.

In [23], some general derivative-free optimization methods are proposed to solve smooth problems with and without noise. Specifically, a derivative-free method with constant stepsize (DFC) is developed in [23] to deal with objective functions of $\mathcal{C}^{1,1}_L$ class, (i.e., C^1-smooth functions whose gradients are globally Lipschitz with constant L), and a derivative-free method with backtracking stepsize (DFB) is proposed therein to solve problems with $\mathcal{C}^{1,1}$ objectives, i.e., C^1-smooth functions having locally Lipschitzian gradients. The ideas for developing these methods come from the previous papers on inexact first-order methods, which started from [20, 21] for inexact gradient descent methods and then continued in [22] for inexact proximal methods with applications to deep learning provided in [19]. The generality in these algorithmic schemes allows us to encompass various implementations in many types of gradient approximations within DFC and DFB that include finite differences [29], the Gupal estimates [12, 15], and gradient estimations via linear interpolation [5]. Observe that the convergence analysis conducted in [23] focuses solely on global convergence properties with proving the stationarity of accumulation points and convergence of iterates under the Kurdyka-Łojasiewicz (KL) inequality.

In this paper, we enrich the convergence analysis for the algorithms in [23], proving the *local convergence* of the iterative sequences to nonisolated local minimizers under the presence of the KL property. First we propose a general framework for derivative-free methods that cover both DFC and DFB. Then we show that under the presence of the KL property, the algorithms satisfying the general framework with proper selections of initial points and parameters converge to a *local minimizer*, where the *isolation* of this local minimizer is *not required*. The latter feature also emphasizes the main difference between our work and the [8, Proposition 1.2.5], which assumes the isolation of the local minimizer. In [3, Theorem 2.10], a local convergence analysis of descent methods is conducted under the presence of the KL property as well. However, an additional global quadratic growth condition is assumed in that result, which is not required in our analysis. In [7], a local convergence analysis of the trust-region method is also considered, which is a derivative-free method different from ours.

The rest of the paper is organized as follows. Section 2 presents some basic definitions and preliminaries used throughout the entire paper. The main results of the paper are presented in Section 3. Concluding remarks on the main contributions of this paper together with some perspectives of our future research are presented in Section 4.

2 Preliminaries

First we recall some basic notions and notations frequently used in the paper. All our considerations are given in the space $\mathbb{R}^n$ with the Euclidean norm $\|\cdot\|$. For any $i = 1,\ldots,n$, let e_i denote the i^{th} basic vector in $\mathbb{R}^n$. As always, $\mathbb{N} := \{1, 2, \ldots\}$ signifies the collection of natural numbers. For any $x \in \mathbb{R}^n$ and $\varepsilon > 0$, let $\mathbb{B}(x, \varepsilon)$ and $\overline{\mathbb{B}}(x, \varepsilon)$ stand for the open and closed balls centered at x with radius ε, respectively. When $x = 0$, these balls are denoted simply by $\varepsilon\mathbb{B}$ and $\varepsilon\overline{\mathbb{B}}$.

Recall that a mapping $G\colon \mathbb{R}^n \to \mathbb{R}^m$ is *Lipschitz continuous on a subset D* of $\mathbb{R}^n$ if there exists a constant $L > 0$ such that we have

$$\|G(x) - G(y)\| \le L\,\|x - y\| \text{ for all } x, y \in D.$$

If $D = \mathbb{R}^n$, the mapping G is said to be *globally Lipschitz continuous*. The *local Lipschitz continuity* of G on $\mathbb{R}^n$ is understood as the Lipschitz continuity of this mapping on every compact subset of $\mathbb{R}^n$. The latter is equivalent to saying that for any $x \in \mathbb{R}^n$ there exists a neighborhood U of x such that G is Lipschitz continuous on U. In what follows, we denote by $\mathcal{C}^{1,1}$ the class of $\mathcal{C}^1$-smooth mappings that have a *locally Lipschitz continuous gradient* on $\mathbb{R}^n$, while $\mathcal{C}^{1,1}_L$ signifies the class of $\mathcal{C}^1$-smooth mappings that have a *globally Lipschitz continuous gradient with constant $L > 0$* on the entire space $\mathbb{R}^n$.

Our convergence analysis of the numerical algorithms developed in the subsequent sections largely exploits the following important results and notions. The first result, which is taken from [17, Lemma A.11], presents a simple albeit very useful property of real-valued functions with Lipschitz continuous gradients.

Lemma 2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$, let $x, y \in \mathbb{R}^n$, and let $L > 0$. If f is differentiable on the line segment $[x, y]$ with its derivative being L-Lipschitz continuous on this segment, then*

$$|f(y) - f(x) - \langle \nabla f(x), y - x\rangle| \le \frac{L}{2}\,\|y - x\|^2. \tag{2.1}$$

The version of the fundamental *Kurdyka-Łojasiewicz (KL) property* formulated below is taken from Absil et al. [1, Theorem 3.4].

Definition 2.2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function. We say that f satisfies the *KL property* at $\bar{x} \in \mathbb{R}^n$ if there exist a number $\eta > 0$, a neighborhood U of $\bar{x}$, and a nondecreasing function $\psi : (0, \eta) \to (0, \infty)$ such that the function $1/\psi$ is integrable over $(0, \eta)$ and we have

$$\|\nabla f(x)\| \ge \psi\big(f(x) - f(\bar{x})\big) \text{ for all } x \in U \text{ with } f(\bar{x}) < f(x) < f(\bar{x}) + \eta. \tag{2.2}$$

Remark 2.3. If f satisfies the KL property at $\bar{x}$ with a neighborhood U, it is clear that the same property holds for any $x \in U$ where $f(x) = f(\bar{x})$. It has been realized that the KL property is satisfied in broad settings. In particular, it holds at every *nonstationary point* of f; see [2, Lemma 2.1 and Remark 3.2(b)]. Furthermore, it is proved in the seminal paper by Łojasiewicz [26] that any analytic function $f : \mathbb{R}^n \to \mathbb{R}$ satisfies the KL property at every point $\bar{x}$ with $\psi(t) = Mt^q$ for some $q \in [0, 1)$; this property with $q = 1/2$ was independently i=introduced and applied to gradient methods by Polyak [31]. As demonstrated in [20, Section 2], the KL property formulated in Attouch et al. [2] is stronger than the one in Definition 2.2. Typical smooth functions that satisfy the KL property from [2], and hence the one from Definition 2.2, are smooth *semialgebraic* functions and also those from the more general class of functions known as *definable in o-minimal structures*; see [2, 3, 25]. The latter property is fulfilled, e.g., in important models arising in deep neural networks, low-rank matrix recovery, principal component analysis, and matrix completion as discussed in [9, Section 6.2].

3 Main results

We begin this section by recalling the global and local approximation considered in [23].

Definition 3.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^1$-smooth function. A mapping $\mathcal{G} : \mathbb{R}^n \times (0, \infty) \to \mathbb{R}^n$ is said to be :

(i) A *global approximation* of ∇f if there exists a constant $C > 0$ such that

$$\|\mathcal{G}(x, \delta) - \nabla f(x)\| \le C\delta \;\text{ for any }\; (x, \delta) \in \mathbb{R}^n \times (0, \infty). \tag{3.1}$$

(ii) A *local approximation* of ∇f if for any bounded set $\Omega \subset \mathbb{R}^n$ and any $\Delta > 0$, there exists a constant $C > 0$ with

$$\|\mathcal{G}(x, \delta) - \nabla f(x)\| \le C\delta \;\text{ for any }\; (x, \delta) \in \Omega \times (0, \Delta]. \tag{3.2}$$

Remark 3.2. It is obvious that any global approximation of ∇f is also a local approximation of this mapping. Examples of both global and local approximations can be given via finite differences [29] as follows:

- *Forward finite difference*:

$$\mathcal{G}(x, \delta) := \frac{1}{\delta} \sum_{i=1}^{n} \left(f(x + \delta e_i) - f(x)\right) e_i \text{ for any } (x, \delta) \in \mathbb{R}^n \times (0, \infty), \tag{3.3}$$

  where e_i is the i^{th} basis of $\mathbb{R}^n$, i.e., the vector with a one in the i^{th} position and zeros everywhere else.

- *Central finite difference*:

$$\mathcal{G}(x, \delta) := \frac{1}{2\delta} \sum_{i=1}^{n} \left(f(x + \delta e_i) - f(x - \delta e_i)\right) e_i \text{ for any } (x, \delta) \in \mathbb{R}^n \times (0, \infty). \tag{3.4}$$

  It is shown in [23, Proposition 3.4] that if f belongs to the classes $\mathcal{C}_L^{1,1}$ and $C^{1,1}$, then the forward and central differences given by (3.3) and (3.4) are global and local approximations of ∇f, respectively. More examples of these types of approximations can be found via Gupal estimates [12, 15] and linear interpolation [5].

Now we are ready to propose a general framework for derivative-free methods using the local approximation mentioned above.

Algorithm 1.

Step 0. Choose a local approximation $\mathcal{G}$ of ∇f under condition (3.2). Select an initial point $x^1 \in \mathbb{R}^n$, an initial sampling radius $\delta_1 > 0$, a sequence $\{C_k\} \subset \mathbb{R}_+$, a reduction factor $\theta \in (0, 1)$, and a scaling factor $\mu > 2$. Choose a sequence of manually controlled sampling radii $\{\nu_k\} \subset [0, \infty)$. Set $k := 1$.

Step 1 (approximate gradient)**.** Find g^k and the smallest integer $i_k \ge 0$ such that

$$g^k = \mathcal{G}\big(x^k, \min\left\{\theta^{i_k}\delta_k, \nu_k\right\}\big) \;\text{ and }\; \left\|g^k\right\| > \mu C_k \theta^{i_k}\delta_k.$$

Then set $\delta_{k+1} := \theta^{i_k}\delta_k$.

Step 2 (update)**.** Choose a stepsize $\tau_k \ge 0$, set $x^{k+1} := x^k - \tau_k g^k$. and return to Step 1.

The result addressing the local convergence of Algorithm 1 to local minimizers of problem (1.1) is presented below.

Theorem 3.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^1$-smooth function, let $\bar{x} \in \mathbb{R}^n$, and let $\Delta > 0$. Assume that $\bar{x}$ is a local minimizer of f satisfying the KL property at $\bar{x}$, and that ∇f is locally Lipschitz continuous around $\bar{x}$. Then we have the assertions:*

(i) *There exist positive numbers ξ, T, and C such that for any initial point $x^1 \in \mathbb{B}(\bar{x}, \xi)$, an initial radius $\delta_1 \in (0, \Delta]$ as well as sequences $\{\tau_k\} \subset [0, T]$, $\{C_k\} \subset [C, \infty)$, and $\{\nu_k\} \subset [0, \infty)$, it holds that the iterative sequence $\{x^k\}$ of Algorithm 1 converges provided that $\nabla f(x^k) \neq 0$ for all $k \in \mathbb{N}$.*

(ii) *If in addition $\sum_{k=1}^{\infty} \tau_k = \infty$, then the sequence $\{x^k\}$ converges to a local minimizer $\widetilde{x}$ of f with $f(\widetilde{x}) = f(\bar{x})$.*

Proof. Since f satisfies the KL property at $\bar{x}$, there exist a bounded neighborhood U of $\bar{x}$, a number $\eta > 0$, and a nonincreasing function $\psi : (0, \eta) \to (0, \infty)$ such that $1/\psi$ is integrable over $(0, \eta)$ and that

$$\|\nabla f(x)\| \geq \psi(f(x) - f(\bar{x})) \text{ for all } x \in U \text{ with } f(\bar{x}) < f(x) < f(\bar{x}) + \eta. \tag{3.5}$$

Remark 2.3 tells us that (3.5) also holds if $\bar{x}$ is replaced by any $\widetilde{x} \in U$ with $f(\widetilde{x}) = f(\bar{x})$. Since $\bar{x}$ is a local minimizer of f and since f is continuous, we can assume by shrinking U if necessary that $f(\bar{x}) \leq f(x) < f(\bar{x}) + \eta$ for all $x \in U$. Combining this with (3.5) yields

$$\|\nabla f(x)\| \geq \psi(f(x) - f(\bar{x})) > 0 \text{ for all } x \in U \text{ with } f(x) \neq f(\bar{x}), \tag{3.6}$$

which implies that $f(\bar{x})$ is the only critical value of f within U. The local Lipschitz continuity of ∇f around $\bar{x}$ gives us numbers $\rho, L \geq 0$ such that $\mathbb{B}(\bar{x}, 2\rho) \subset U$ and ∇f is Lipschitz continuous with constant L on $\mathbb{B}(\bar{x}, 2\rho)$. Choose further $T := \min\left\{\frac{1}{3L}, \frac{\mu-2}{\mu L}\right\}$ and define $\varphi : [0, \eta) \to [0, \infty)$ by $\varphi(x) := \int_0^x \frac{1}{\psi(t)} dt$ for $x \in (0, \eta)$ with $\varphi(0) := 0$. By the right continuity of φ at 0 and the continuity of f at $\bar{x}$, we find such $\xi \in (0, \rho)$ that

$$\|x - \bar{x}\| + 4\varphi(f(x) - f(\bar{x})) < \rho \text{ for all } x \in \mathbb{B}(\bar{x}, \xi). \tag{3.7}$$

Since $\mathcal{G}$ is a local approximation of ∇f under condition (3.2), it gives us $C > 0$ for which

$$\|\mathcal{G}(x, \delta) - \nabla f(x)\| \leq C\delta \text{ whenever } (x, \delta) \in U \times (0, \Delta]. \tag{3.8}$$

Assuming that $x^1 \in \mathbb{B}(\bar{x}, \xi)$, $\delta_1 \in (0, \Delta]$, $\{\tau_k\} \subset [0, T]$, and $\{C_k\} \subset [C, \infty)$, we now aim at verifying (i). To proceed, let us first prove the following claim.

Claim 1. *Algorithm 1 generates the well-defined iterative sequence $\{x^k\}$, which stays inside $\mathbb{B}(\bar{x}, \rho)$ whenever $k \in \mathbb{N}$.*

Indeed, by $\nabla f(x^k) \neq 0$ for all $k \in \mathbb{N}$, the existence of g^k in Step 2 of Algorithm 1 is guaranteed in each iteration, which ensures that the iterative sequence $\{x^k\}$ is well-defined. To verify that $\{x^k\} \subset \mathbb{B}(\bar{x}, \rho)$, we proceed by induction. Fix $K \in \mathbb{N}$ and assume that $x^k \in \mathbb{B}(\bar{x}, \rho)$ for all $k = 1, \ldots, K$. To show that $x^{K+1} \in \mathbb{B}(\bar{x}, \rho)$, observe that $\delta_k \leq \delta_1 \leq \Delta$ for all $k \in \mathbb{N}$ by the selection of δ_1 and the construction of $\{\delta_k\}$ in Algorithm 1. Since $\{C_k\} \subset [C, \infty)$, we deduce from (3.8) and $g^k = \mathcal{G}(x^k, \min\{\delta_{k+1}, \nu_k\})$ in Step 1 of Algorithm 1 that

$$\left\|g^k - \nabla f(x^k)\right\| \leq C \min\{\delta_{k+1}, \nu_k\} \leq C_k \delta_{k+1} \leq \mu^{-1} \left\|g^k\right\| \text{ for all } k = 1, \ldots, K. \tag{3.9}$$

It follows from (3.9) with $k := K$, the triangle inequality, and the choice of $\mu > 2$ that

$$\left\|\nabla f(x^K)\right\| \geq \left\|g^K\right\| - \left\|g^K - \nabla f(x^K)\right\| > (1 - \mu^{-1}) \left\|g^K\right\| > \tfrac{1}{2} \left\|g^K\right\|. \tag{3.10}$$

Since $x^K \in \mathbb{B}(\bar{x}, \rho)$, we deduce from the Lipschitz continuity of f on $\mathbb{B}(\bar{x}, 2\rho)$ that

$$\left\|\nabla f(x^K)\right\| = \left\|\nabla f(x^K) - \nabla f(\bar{x})\right\| \leq L \left\|x^K - \bar{x}\right\| \leq L\rho.$$

Combining this with the update $x^{k+1} = x^k - \tau_k g^k$, $\tau_k \leq T \leq (3L)^{-1}$, and (3.10) gives us

$$\left\|x^{K+1} - x^K\right\| = \tau_k \left\|g^K\right\| \leq 2T \left\|\nabla f(x^K)\right\| \leq 2TL\rho < \rho,$$

which means that $x^{K+1} \in \mathbb{B}(\bar{x}, 2\rho)$. Since $\nabla f(x^{K+1}) \neq 0$ and $f(\bar{x})$ is the minimum value of f within $\mathbb{B}(\bar{x}, 2\rho)$, we get that $f(x^{K+1}) > f(\bar{x})$. It follows from the triangle inequality and estimate (3.9) that

$$\left\|g^k\right\| \geq \left\|\nabla f(x^k)\right\| - \left\|g^k - \nabla f(x^k)\right\| \geq \left\|\nabla f(x^k)\right\| - \mu^{-1} \left\|g^k\right\|,$$

which implies in turn that

$$\left\|\nabla f(x^k)\right\| \leq (1 + \mu^{-1}) \left\|g^k\right\| \leq 2 \left\|g^k\right\| \quad \text{for all} \quad k = 1, \ldots, K. \tag{3.11}$$

The Lipschitz continuity of ∇f on $\mathbb{B}(\bar{x}, 2\rho) \supset \left\{x^k \mid k = 1, \ldots, K+1\right\}$ with the Lipschitz constant $L > 0$ ensures that

$$\begin{aligned}
f(x^{k+1}) - f(x^k) &\leq \left\langle \nabla f(x^k), x^{k+1} - x^k \right\rangle + \frac{L}{2} \left\|x^{k+1} - x^k\right\|^2 && \text{(3.12a)}\\
&= -\tau_k \left\langle \nabla f(x^k), g^k \right\rangle + \frac{L\tau_k^2}{2} \left\|g^k\right\|^2 && \text{(3.12b)}\\
&= -\tau_k \left\langle \nabla f(x^k) - g^k, g^k \right\rangle - \tau_k \left\|g^k\right\|^2 + \frac{L\tau_k^2}{2} \left\|g^k\right\|^2 &&\\
&\leq \tau_k \left\|\nabla f(x^k) - g^k\right\| \left\|g^k\right\| - \tau_k \left\|g^k\right\|^2 + \frac{L\tau_k^2}{2} \left\|g^k\right\|^2 && \text{(3.12c)}\\
&\leq \tau_k \frac{1}{\mu} \left\|g^k\right\|^2 - \tau_k \left\|g^k\right\|^2 + \frac{L\tau_k^2}{2} \left\|g^k\right\|^2 && \text{(3.12d)}\\
&\leq \left\|\tau_k g^k\right\| \left\|g^k\right\| \left(\frac{1}{\mu} - 1 + \frac{L\tau_k}{2}\right) &&\\
&\leq -\frac{1}{2} \left\|x^{k+1} - x^k\right\| \left\|g^k\right\| && \text{(3.12e)}\\
&\leq -\frac{1}{4} \left\|x^{k+1} - x^k\right\| \left\|\nabla f(x^k)\right\| \quad \text{for all} \quad k = 1, \ldots, K, && \text{(3.12f)}
\end{aligned}$$

where (3.12a) follows from Lemma 2.1, (3.12b) follows from the iterative update $x^{k+1} = x^k - \tau_k g^k$, (3.12c) follows from the Cauchy-Schwarz inequality, (3.12d) is deduced by $\left\|g^k - \nabla f(x^k)\right\| \leq \frac{1}{\mu} \left\|g^k\right\|$ from (3.9), (3.12e) follows from $\tau_k \leq T \leq \dfrac{\mu - 2}{\mu L}$, and (3.12f) is deduced by $\left\|\nabla f(x^k)\right\| \leq 2 \left\|g^k\right\|$ from (3.11). Therefore,

$$\frac{1}{4} \left\|x^{k+1} - x^k\right\| \left\|\nabla f(x^k)\right\| \leq f(x^k) - f(x^{k+1}) \text{ whenever } k = 1, \ldots, K. \tag{3.13}$$

As a consequence of the above, we have the inequalities

$$f(\bar{x}) + \eta > f(x^k) \geq f(x^{K+1}) > f(\bar{x}) \quad \text{for all} \quad k = 1, \ldots, K+1,$$

which ensure together with $x^k \in \mathbb{B}(\bar{x}, 2\rho)$ and (3.6) that $\left\|\nabla f(x^k)\right\| \geq \psi(f(x^k) - f(\bar{x}))$ for all $k = 1, \ldots, K$. Combining the latter with (3.13) leads us to the conditions

$$\begin{aligned}
\frac{1}{4} \left\|x^{k+1} - x^k\right\| \leq \frac{f(x^k) - f(x^{k+1})}{\psi(f(x^k) - f(\bar{x}))} &= \int_{f(x^{k+1})}^{f(x^k)} \frac{1}{\psi(f(x^k) - f(\bar{x}))} dt\\
&\leq \int_{f(x^{k+1})}^{f(x^k)} \frac{1}{\psi(t - f(\bar{x}))} dt\\
&= \varphi(f(x^k) - f(\bar{x})) - \varphi(f(x^{k+1}) - f(\bar{x})) \quad \text{for all } k = 1, \ldots, K,
\end{aligned}$$

where the second inequality follows from the nondecreasing property of ψ. Therefore, the triangle inequality gives us the relationships

$$\begin{aligned}
\left\|x^{K+1}-\bar{x}\right\| &\leq \|x_1-\bar{x}\| + \sum_{k=1}^{K}\left\|x^{k+1}-x^k\right\| \\
&\leq \|x_1-\bar{x}\| + 4\sum_{k=1}^{K}[\varphi(f(x^k)-f(\bar{x}))-\varphi(f(x^{k+1})-f(\bar{x}))] \\
&= \|x_1-\bar{x}\| + 4\left[\varphi(f(x^1)-f(\bar{x}))-\varphi(f(x^{K+1})-f(\bar{x}))\right] \\
&\leq \|x_1-\bar{x}\| + 4\varphi(f(x^1)-f(\bar{x})) < \rho,
\end{aligned}$$

where the latter inequality is a consequence of the selection $x^1 \in \mathbb{B}(\bar{x},\xi)$ and ξ in (3.7). This means that $x^{K+1} \in \mathbb{B}(\bar{x},\rho)$. By induction we arrive at $x^k \in \mathbb{B}(\bar{x},\rho)$ for all $k \in \mathbb{N}$, which verifies the claimed assertion.

Claim 2. *The sequence of iterates $\left\{x^k\right\}$ converges to some $\widetilde{x} \in \overline{\mathbb{B}}(\bar{x},\rho)$. If in addition we have $\sum_{k=1}^{\infty}\tau_k = \infty$, then $\widetilde{x}$ is a local minimizer of f.*

Picking any $K \in \mathbb{N}$ and arguing similarly to the proof of Claim 1 with taking into account that $x^k \in \mathbb{B}(\bar{x},\rho)$ for all $k \in \mathbb{N}$, we get the estimates

$$\sum_{k=1}^{K}\left\|x^{k+1}-x^k\right\| \leq 4\sum_{k=1}^{K}[\varphi(f(x^{k+1})-f(\bar{x}))-\varphi(f(x^k)-f(\bar{x}))] \leq 4\varphi(f(x^1)-f(\bar{x})).$$

Passing there to the limit as $K \to \infty$ yields $\sum_{k=1}^{\infty}\left\|x^{k+1}-x^k\right\| < \infty$, which tells us that $\left\{x^k\right\}$ converges to some point $\widetilde{x} \in \overline{\mathbb{B}}(\bar{x},\rho)$. Let now $\sum_{k=1}^{\infty}\tau_k = \infty$ be satisfied. Since $\left\{x^k\right\} \subset \mathbb{B}(\bar{x},\rho)$, we proceed similarly to the proof of (3.12e) in Claim 1 to show that

$$f(x^{k+1})-f(x^k) \leq -\frac{1}{2}\left\|x^{k+1}-x^k\right\|\left\|g^k\right\| \quad \text{for all} \quad k \in \mathbb{N}. \tag{3.14}$$

Combining this with the fact that $f(x^k) \geq f(\bar{x})$ for all $k \in \mathbb{N}$ yields

$$\begin{aligned}
\sum_{k=1}^{\infty}\tau_k\left\|g^k\right\|^2 = \sum_{k=1}^{\infty}\left\|x^{k+1}-x^k\right\|\left\|g^k\right\| &\leq 2\sum_{k=1}^{\infty}(f(x^k)-f(x^{k+1})) \\
&\leq 2(f(x^1)-f(\bar{x})) < \infty.
\end{aligned}$$

Supposing that there is $r>0$ with $\left\|g^k\right\| \geq r$ for all k sufficiently large, the above inequality gives us $\sum_{k=1}^{\infty}\tau_k < \infty$, which is a contradiction. Therefore, $0 \in \mathbb{R}^n$ is an accumulation point of $\left\{g^k\right\}$, i.e., there exists an infinite set $J \subset \mathbb{N}$ such that $g^k \xrightarrow{J} 0$. As in the proof of (3.9) in Claim 1 with taking now $x^k \in \mathbb{B}(\bar{x},\rho)$ into account, we get the estimate

$$\left\|g^k-\nabla f(x^k)\right\| \leq \mu^{-1}\left\|g^k\right\| \quad \text{for all} \quad k \in \mathbb{N}.$$

Unifying the latter with $g^k \xrightarrow{J} 0$ ensures that $\nabla f(x^k) \xrightarrow{J} 0$. Remembering that $\left\{x^k\right\}$ converges to $\widetilde{x}$, we have that $\nabla f(x^k) \to \nabla f(\widetilde{x})$ as $k \to \infty$. Therefore, $\nabla f(\widetilde{x}) = 0$, i.e., $\widetilde{x}$ is a stationary point of f on $\mathbb{B}(\bar{x},2\rho)$. Since $f(\bar{x})$ is the only critical value of f within $\mathbb{B}(\bar{x},2\rho)$, we get that $f(\widetilde{x}) = f(\bar{x})$, which tells us that $\widetilde{x}$ is a local minimizer of f. □

Remark 3.4. Note that the condition $\nabla f(x^k) \neq 0$ for all $k \in \mathbb{N}$ in Theorem 3.3 is a standard assumption in the convergence analysis of derivative-free optimization methods since there is no tool available to determine whether $\nabla f(x^k)$ equals zero or not. Similar assumptions can also be found at [10, Section 4] and [15, Corollary 3.3]. Quite recently, the paper by Josz et al. [18] has developed a local convergence analysis for *exact momentum methods* with *constant stepsizes* while mainly focusing on *semi-algebraic* functions with gradients being *locally Lipschitzian everywhere*. It is necessary to emphasize that this analysis does not encompass the obtained convergence properties in Theorem 3.3. Specifically, our work focuses on *derivative-free methods* with *variable stepsizes* when applied to $\mathcal{C}^1$-smooth functions satisfying the KL property. These functions have gradients that are *locally Lipschitzian* but only in the vicinity of *local minimizers*. Observe also that the local convergence result above does not follow from the one in [3, Theorem 2.10]. The latter relies on the global conditions (H1) and (H2) therein, along with the local growth condition (H4), which are not assumed in our derivative-free context and may not be satisfied within the scope of our given assumptions.

If the condition $\sum_{k=1}^{\infty} \tau_k = \infty$ is removed in Theorem 3.3, a trivial example with $\tau_k = 0$ for all $k \in \mathbb{N}$ can be used to show that the iterative sequence $\{x^k\}$ generated by Algorithm 1 may remain at the given initial point, which is not a stationary one, and thus it does not converge to any local minimizer of f. Even when $\tau_k > 0$ for all $k \in \mathbb{N}$ but $\sum_{k=1}^{\infty} \tau_k < \infty$, the following example shows that $\{x^k\}$ may still converge to a nonstationary point, which confirms the necessity of the assumption $\sum_{k=1}^{\infty} \tau_k = \infty$ in deriving the local optimality of $\widetilde{x}$ in Theorem 3.3.

Example 1. Considering the function $f(x) := x^2$, it is clear that its derivative f' is globally Lipschitz continuous, and that f satisfies the KL property at the local minimizer 0 with $\psi(t) := 2t^{\frac{1}{2}}$. Therefore, all the assumptions in Theorem 3.3 are satisfied except for $\sum_{k=1}^{\infty} \tau_k = \infty$. Now we consider Algorithm 1 with $\mathcal{G}(x, \delta)$ being chosen via the central finite difference. Then for any $x \in \mathbb{R}$, we have

$$\mathcal{G}(x,\delta) = \frac{f(x+\delta) - f(x-\delta)}{2\delta} = \frac{(x+\delta)^2 - (x-\delta)^2}{2\delta} = \frac{4x\delta}{2\delta} = 2x = f'(x) \text{ whenever } \delta > 0. \tag{3.15}$$

Let us show that for any positive numbers ξ, T and C, there exists an iterative sequence $\{x^k\}$ generated by Algorithm 1 with the initial point $x^1 \in \mathbb{B}(0, \xi)$, with the sequence of stepsizes $\{\tau_k\} \subset (0, T]$, and with $C_k = C$ for all $k \in \mathbb{N}$ such that $\{x^k\}$ converges to a *nonstationary point* of f. To proceed, take any $\xi, T, C > 0$ and choose $x^1 := \xi/2 \in \mathbb{B}(0, \xi)$. The sequence $\{\tau_k\}$ is constructed inductively as follows. For any $k \in \mathbb{N}$ with $x^k > x^1/2 > 0$, we deduce from Step 1 of Algorithm 1 and (3.15) that

$$g^k = \mathcal{G}(x^k, \delta_{k+1}) = f'(x^k) = 2x^k \quad \text{for all} \quad k \in \mathbb{N}.$$

Choosing $\tau_k := \min\left\{T, \frac{1}{4} - \frac{x^1}{8x^k}\right\} > 0$ tells us that

$$\begin{aligned} x^{k+1} = x^k - \tau_k g^k = (1 - 2\tau_k)x^k &\geq \left(1 - \frac{1}{2} + \frac{x^1}{4x^k}\right)x^k \\ &= \frac{x^k}{2} + \frac{x^1}{4} > \frac{x^1}{2}. \end{aligned}$$

Arguing by induction with the usage of $x^1 > \frac{x^1}{2}$ ensures that the sequence of positive stepsizes $\{\tau_k\}$ constructed as above provides $x^k > \frac{x^1}{2}$ for all $k \in \mathbb{N}$. This implies that $\{x^k\}$ does not converge to 0, which is the only stationary point of f. We can actually find the exact limit for $\lim\limits_{k\to\infty} x^k$ to see that it is clearly not 0. Since $x^k > 0$ for all $k \in \mathbb{N}$, the iterative procedure $x^{k+1} = x^k - 2\tau_k x^k$ shows that $\{x^k\}$ is decreasing, and thus it has a limit $\bar{x}$ by taking into account its boundedness from below. Therefore,

$$\tau_k = \frac{x^k - x^{k+1}}{2x^k} \leq \frac{x^k - x^{k+1}}{4x^1} \to 0 \quad \text{as} \quad k \to \infty$$

which implies by the selection of τ_k that $\lim_{k\to\infty} \frac{x^1}{8x^k} = \frac{1}{4}$, i.e., $\lim_{k\to\infty} x^k = \frac{x^1}{2}$. It can also be observed that the assumption $\sum_{k=1}^{\infty} \tau_k = \infty$ is not satisfied in this example since $x^1\tau_k \leq 2\tau_k x^k = x^k - x^{k+1}$ yields

$$\sum_{k=1}^{\infty} \tau_k \leq \frac{1}{x^1} \sum_{k=1}^{\infty} (x^k - x^{k+1}) = \frac{1}{x^1}(x^1 - \lim_{k\to\infty} x^k) < \infty.$$

Next we recall the algorithms designed in [23] and derive their corresponding *local convergence* results by employing Theorem 3.3.

Algorithm 2 (DFC).

Step 0. Choose a global approximation $\mathcal{G}$ of ∇f under condition (3.1). Select an initial point $x^1 \in \mathbb{R}^n$, an initial sampling radius $\delta_1 > 0$, a constant $C_1 > 0$, a reduction factor $\theta \in (0, 1)$, and scaling factors $\mu > 2, \eta > 1, \kappa > 0$. Set $k := 1$.

Step 1 (approximate gradient)**.** Find g^k and the smallest integer $i_k \geq 0$ such that

$$g^k = \mathcal{G}(x^k, \theta^{i_k}\delta_k) \text{ and } \|g^k\| > \mu C_k \theta^{i_k}\delta_k.$$

Then set $\delta_{k+1} := \theta^{i_k}\delta_k$.

Step 2 (update)**.** If $f\left(x^k - \frac{\kappa}{C_k} g^k\right) \leq f(x^k) - \frac{\kappa(\mu - 2)}{2C_k\mu} \|g^k\|^2$, then $x^{k+1} := x^k - \frac{\kappa}{C_k} g^k$ and $C_{k+1} := C_k$. Otherwise, $x^{k+1} := x^k$ and $C_{k+1} := \eta C_k$.

Corollary 3.5. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^1$-smooth function with a globally Lipschitz continuous gradient, let $\bar{x} \in \mathbb{R}^n$, and let $\Delta > 0$. Assume that $\bar{x}$ is a local minimizer of f satisfying the KL property at $\bar{x}$, and that $\nabla f(x^k) \neq 0$ for all $k \in \mathbb{N}$. Then there exist constants $\xi, C > 0$ such that for any initial point $x^1 \in \mathbb{B}(\bar{x}, \xi)$, any initial sampling radius $\delta_1 \in (0, \Delta]$, any $C_1 \geq C$ and other parameters listed in Algorithm 2, we have that $\{x^k\}$ converges to a local minimizer $\widetilde{x}$ of f with $f(\widetilde{x}) = f(\bar{x})$.*

Proof. By Theorem 3.3, there exist $\xi, T, C > 0$ such that for any initial point $x^1 \in \mathbb{B}(\bar{x}, \xi)$, initial radius $\delta_1 \in (0, \Delta]$, and sequences $\tau_k \subset [0, T]$ and $C_k \subset [C, \infty)$, it holds that any sequence of iterates generated by Algorithm 1 exhibits the convergence properties presented in Theorem 3.3. By choosing a larger C if necessary, we can assume that $\kappa/C \leq T$, where $\kappa > 0$ is the parameter taken from Algorithm 2.

It suffices to show that Algorithm 2 with initial point $x^1 \in \mathbb{B}(\bar{x}, \xi)$, $C_1 \geq C$, $\delta_1 \in (0, \Delta]$, and the other parameters therein is a special case of Algorithm 1 with the parameters listed above, and thus Algorithm 2 enjoys the desired convergence properties. It is clear from the construction of Algorithm 2 that $C_{k+1} \geq C_k$ for all $k \in \mathbb{N}$, which implies that $C_k \geq C_1 \geq C$ whenever $k \in \mathbb{N}$. The selection of $\{g^k\}$ in Step 1 of Algorithm 1 reduces to that of Algorithm 2 by choosing $\nu_k \geq \delta_1$ whenever $k \in \mathbb{N}$. The iterative procedure of Algorithm 2 can be rewritten as

$$x^{k+1} = x^k - \tau_k g^k \quad \text{with either} \quad \tau_k = 0, \quad \text{or} \quad \tau_k = \kappa/C_k \leq \kappa/C \leq T \quad \text{for all} \quad k \in \mathbb{N}$$

telling us that $\{\tau_k\} \subset [0, T]$. The proof of [23, Theorem 4.3] guarantees that C_k are constant for large $k \in \mathbb{N}$, which ensures by Step 2 of Algorithm 2 that the stepsizes τ_k are equal to a positive constant for large $k \in \mathbb{N}$. This guarantees that the sequence $\{\tau_k\}$ is bounded away from 0, and furthermore $\sum_{k=1}^{\infty} \tau_k = \infty$. All the assumptions in Theorem 3.3 are satisfied, so $\{x^k\}$ converges to some local minimizer $\widetilde{x}$ of f with $f(\widetilde{x}) = f(\bar{x})$. $\square$

Algorithm 3 (DFB)**.**

Step 0 (initialization)**.** Choose a local approximation $\mathcal{G}$ of ∇f under condition (3.2). Select an initial point $x^1 \in \mathbb{R}^n$ and initial radius $\delta_1 > 0$, a constant $C_1 > 0$, factors $\theta \in (0, 1)$, $\mu > 2, \eta > 1$, linesearch constants $\beta \in (0, 1/2), \gamma \in (0, 1)$, $\bar{\tau} > 0$, and an initial bound $t_1^{\min} \in (0, \bar{\tau})$. Choose a sequence of manually controlled errors $\{\nu_k\} \subset [0, \infty)$ such that $\nu_k \downarrow 0$ as $k \to \infty$. Set $k := 1$.

Step 1 (approximate gradient)**.** Select g^k and the smallest integer $i_k \geq 0$ so that

$$g^k = \mathcal{G}(x^k, \min\{\theta^{i_k}\delta_k, \nu_k\}) \text{ and } \left\|g^k\right\| > \mu C_k \theta^{i_k} \delta_k. \tag{3.16}$$

Then set $\delta_{k+1} := \theta^{i_k}\theta_k$.

Step 2 (linesearch)**.** Set $t_k := \bar{\tau}$. While

$$f(x^k - t_k g^k) > f(x^k) - \beta t_k \left\|g^k\right\|^2 \text{ and } t_k \geq t_k^{\min}, \tag{3.17}$$

set $t_k := \gamma t_k$.

Step 3 (stepsize and parameters update)**.** If $t_k \geq t_k^{\min}$, then set $\tau_k := t_k$, $C_{k+1} := C_k$, and $t_{k+1}^{\min} := t_k^{\min}$. Otherwise, set $\tau_k := 0$, $C_{k+1} := \eta C_k$ and $t_{k+1}^{\min} := \gamma t_k^{\min}$.

Step 4 (iteration update)**.** Set $x^{k+1} := x^k - \tau_k g^k$. Increase k by 1 and return to Step 1.

Corollary 3.6. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^1-smooth function with a locally Lipschitzian gradient, let $\bar{x} \in \mathbb{R}^n$ and $\Delta > 0$. Assume that $\bar{x}$ is a local minimizer of f, which satisfies the KL property at $\bar{x}$, and that $\nabla f(x_k) \neq 0$ for all $k \in \mathbb{N}$. Then there are constants $\xi, T, C > 0$ such that for any initial point $x^1 \in \mathbb{B}(\bar{x}, \xi)$, any initial sampling radius $\delta_1 \in (0, \Delta]$, $\bar{\tau} \in (0, T], C_1 \geq C$, and the other parameters of Algorithm 3, the sequence of iterates $\{x_k\}$ in this algorithm converges to some local minimizer $\widetilde{x}$ with $f(\widetilde{x}) = f(\bar{x})$.*

Proof. By Theorem 3.3, there exist positive numbers ξ, T, C such that for any initial point $x^1 \in \mathbb{B}(\bar{x}, \xi)$, any initial radius $\delta_1 \in (0, \Delta]$, and any sequences $\{\tau_k\} \subset [0, T]$, and $\{C_k\} \subset [C, \infty)$, the sequence of iterates generated by Algorithm 1 exhibits the properties listed in Theorem 3.3(i,ii).

It is sufficient to verify that Algorithm 3 with the initial point $x^1 \in \mathbb{B}(\bar{x}, \xi)$, $C_1 \geq C, \bar{\tau} \in (0, T], \delta_1 \in (0, \Delta]$, and with other parameters taken from Algorithm 3 is a special case of the general Algorithm 1, and hence it enjoys the claimed convergence properties. We see from the structure of Algorithm 3 that $C_{k+1} \geq C_k$ for all $k \in \mathbb{N}$, which tells us that $C_k \geq C$ for all $k \in \mathbb{N}$.

It follows from Step 2 of Algorithm 3 that $\tau_k \leq \bar{\tau}$ for all $k \in \mathbb{N}$. Combining this with $\bar{\tau} \leq T$, ensures that $\tau_k \subset [0, T]$. Moreover, the selection of $\{g^k\}$ in Step 1 of Algorithm 1 also reduces to that of Algorithm 3 since $\{\nu_k\} \downarrow 0$ as $k \to \infty$. Theorem 3.3(i) tells us that the sequence of iterates $\{x^k\}$ generated by Algorithm 3 converges, and thus it is bounded. By using equation (5.27) in the proof of [23, Theorem 5.5], we get that the sequence of stepsizes $\{\tau_k\}$ is bounded away from 0. This guarantees the condition $\sum_{k=1}^{\infty} \tau_k = \infty$ in

Theorem 3.3(ii), which verifies therefore that $\{x^k\}$ converges to some local minimizer $\widetilde{x}$ of f with the convergence rates as in Theorem 3.3(ii). Thus the proof is complete. □

Remark 3.7. Let us now summarize the main differences between the three algorithms mentioned above. Algorithm 1 is a general one that does not specify the selection of the stepsize, while the stepsize of Algorithm 2 is determined via an inequality and will reduce to a constant stepsize after a finite number of iterations [23, Theorem 4.3], and the stepsize of Algorithm 3 is determined via the standard backtracking line search.

4 Conclusion

This paper conducts a local convergence analysis for derivative-free optimization methods introduced in [23], which are derivative-free methods with constant stepsize (DFC) and derivative-free methods with backtracking stepsize (DFB). The analysis shows that whenever the KL property holds at a nonisolated local minimizer of a smooth objective function with a proper initialization, the convergence of DFC and DFB to a local minimizer near the nonisolated local minimizer in question is guaranteed. An example is presented to illustrate the necessity of the imposed condition on the nonsummability of the stepsize sequence in the main theorem. In the future, we intend to develop this local analysis, as well as the global analysis in [23], to model-based derivative-free methods.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

[1] P.-A. Absil, R. Mahony and B. Andrews, Convergence of the iterates of descent methods for analytic cost functions, SIAM J. Optim. 16 (2005), 531–547.

[2] H. Attouch, J. Bolte, P. Redont and A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka-Łojasiewicz property, Math. Oper. Res. 35 (2010), 438–457.

[3] H. Attouch, J. Bolte and B. F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, Math. Program. 137 (2013), 91–129.

[4] C. Audet and W. Hare, Derivative-Free and Blackbox Optimization, Springer, Cham, Switzerland, 2017.

[5] A. S. Berahas, L. Cao, K. Choromanski and K. Scheinberg, A theoretical and empirical comparison of gradient approximations in derivative-free optimization, Found. Comput. Math. 22 (2022), 507–560.

[6] A. S. Berahas, L. Cao and K. Scheinberg, Global convergence rate analysis of a generic linesearch algorithm with noise, SIAM J. Optim. 31 (2021), 1489–1518.

[7] R. J. Baraldi, D. P. Kouri, Local convergence analysis of an inexact trust-region method for nonsmooth optimization, Optim. Lett. 18 (2024), 663–680.

[8] D. P. Bertsekas, Nonlinear Programming, 3rd edition, Athena Scientific, Belmont, MA, 2016.

[9] J. Bolte and E. Pauwels, Conservative set-valued fields, automatic differentiation, stochastic gradient methods and deep learning, Math. Program. 188 (2021), 19–51.

[10] A. R. Conn, K. Scheinberg and L. N. Vicente, Introduction to Derivative-Free Optimization, SIAM, Philadelphia, PA, 2009.

[11] A. R. Conn, K. Scheinberg and L. N. Vicente, Global convergence of general derivative-free trust-region algorithms to first- and second-order critical points, SIAM J. Optim. 20 (2009), 387–415.

[12] A. M. Gupal, On a minimization method for almost-differentiable functions, Cybernetics, 13 (1977), 115—117.

[13] G. A. Gray and T. G. Kolda, Algorithm 856: Appspack 4.0: Asynchronous parallel pattern search for derivative-free optimization, ACM Trans. Math. Softw. 32 (2006), 485–507.

[14] G. Gray, T. Kolda, K. Sale, and M. Young, Optimizing an empirical scoring function for transmembrane protein structure determination, INFORMS J. Comput. 16 (2004), 406–418.

[15] W. Hare and J. Nutini, A derivative-free approximate gradient sampling algorithm for finite minimax problems, Comput. Optim. Appl. 56 (2013), 1–38.

[16] R. Hooke and T. A. Jeeves, Direct search solution of numerical and statistical problems, J. ACM. 8 (1961), 212–229.

[17] A. F. Izmailov and M. V. Solodov, Newton-Type Methods for Optimization and Variational Problems, Springer, New York, 2014.

[18] C. Josz, L. Lai and X. Li, Convergence of the momentum method for semi-algebraic functions with locally Lipschitz gradients, SIAM J. Optim. 33 (2023), 2988–3011.

[19] P. D. Khanh, H. C. Luong, B. S. Mordukhovich and D. B. Tran, Fundamental convergence analysis of sharpness-aware minimization, arXiv:2401.08060 (2024).

[20] P. D. Khanh, B. S. Mordukhovich and D. B. Tran, Inexact reduced gradient methods in smooth nonconvex optimization, J. Optim. Theory Appl. (2023), https://doi.org/10.1007/s10957-023-02319-9.

[21] P. D. Khanh, B. S. Mordukhovich and D. B. Tran, A new inexact gradient descent method with applications to nonsmooth convex optimization, Optim. Methods Softw. (2024), https://doi.org/10.1080/10556788.2024.2322700.

[22] P. D. Khanh, B. S. Mordukhovich, V. T. Phat and D. B. Tran, Inexact proximal methods for weakly convex functions, to appear in J. Global Optim., arXiv:2307.15596 (2023).

[23] P. D. Khanh, B. S. Mordukhovich and D. B. Tran, Globally convergent derivative-free methods in nonconvex optimization with and without noise, https://optimization-online.org/?p=26889 (2024).

[24] T. Lin, Z. Zheng and M. I. Jordan, Gradient-free methods for deterministic and stochastic nonsmooth nonconvex optimization, Adv. Neural Inf. Process Syst. (2022), 26160–26175.

[25] K. Kurdyka, On gradients of functions definable in o-minimal structures, Ann. Inst. Fourier 48 (1998), 769–783.

[26] S. Łojasiewicz, Ensembles Semi-Analytiques, Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, France, 1965.

[27] R. Muthukumar, D. P. Kouri and M. Udell, Randomized sketching algorithms for low-memory dynamic optimization, SIAM J. Optim. 31 (2021), 1242–1275

[28] J. A. Nelder and R. Mead, A simplex method for function minimization, Comput. J. 7 (1965), 308–313.

[29] J. Nocedal and S. J. Wright, Numerical Optimization, 2nd edition. New York, 2006.

[30] R. Oeuvray, Trust-Region Methods Based on Radial Basis Functions with Application to Biomedical Imaging, PhD thesis, Institute of Mathematics, Swiss Federal Institute of Technology, Lausanne (2005).

[31] B. T. Polyak, Gradient methods for the minimization of functionals, USSR Comput. Math. Math. Phys. 3 (1965), 864–878.

[32] H. M. Shi, M. Q. Xuan, F. Oztoprak and J. Nocedal, On the numerical performance of finite-difference-based methods for derivative-free optimization, Optim. Methods Softw. 38 (2023), 289–311.

[33] A. J. Zaslavski, Numerical Optimization with Computational Errors, Springer, New York, 2016.