

Convergence of Descent Optimization Algorithms under Polyak-Łojasiewicz-Kurdyka Conditions

G. C. Bento^{*} B. S. Mordukhovich[†] T. S. Mota[‡] Yu. Nesterov[§]

February 12, 2025

Abstract

This paper develops a comprehensive convergence analysis for generic classes of descent algorithms in non-smooth and nonconvex optimization under several conditions of the Polyak-Łojasiewicz-Kurdyka (PLK) type. Along other results, we prove the finite termination of generic algorithms under the PLK conditions with lower exponents. Specifications are given to establish new convergence rates for inexact reduced gradient methods and some versions of the boosted algorithm in DC programming. It is revealed, e.g., that the lower exponent PLK conditions for a broad class of difference programs are incompatible with the gradient Lipschitz continuity for the plus function around a local minimizer. On the other hand, we show that the above inconsistency observation may fail if the Lipschitz continuity is replaced by merely the gradient continuity.

Keywords: nonsmooth optimization, descent methods, global convergence analysis, Polyak-Łojasiewicz-Kurdyka conditions.

Mathematics Subject Classification (2020) 65K05, 65K10, 90C26, 47N10

1 Introduction, Basic Definitions, and Discussions

The paper examines two generic classes of descent algorithms in (generally nonsmooth and nonconvex) constrained optimization. Our main goals are to establish global convergence and derive convergence rates for such algorithms under appropriate conditions. It has been well recognized that in optimization theory and many applications that conditions of the Polyak-Łojasiewicz-Kurdyka type play a crucial role in the study of global convergence and convergence rates of various numerical algorithms.

For functions ϕ of class $C^{1,1}$ (i.e., continuously differentiable functions with Lipschitzian gradients), Polyak introduced in 1962 (the English translation of his paper was published in [23]) the condition

$$\|\nabla\phi(x)\| \geq (1/2M)|\phi(x) - \phi(\bar{x})|^{1/2}, \quad c > 0,$$

and used it to prove a linear convergence of the gradient descent method in Hilbert spaces (Theorem 4 in [23]). Independently, Łojasiewicz [18] introduced the inequality

$$\|\nabla\phi(x)\| \geq b|\phi(x) - \phi(\bar{x})|^q, \quad b := 1/M(1 - q), \quad q \in [0, 1), \quad (1)$$

for real analytic functions in the finite-dimensional framework of semialgebraic geometry with no applications to optimization. The gradient inequality (1) is referred to (especially in the literature on machine learning and computer science) as the *Polyak-Łojasiewicz* (PL) condition; see, e.g., [13]. The subsequent algebraic-geometric extension of (1) was developed by Kurdyka [15] for the class of definable differentiable functions on \mathbb{R}^n . A nonsmooth extension of the latter condition was first proposed by Bolté et al. [7] in the form of Definition 1(ii) under the name of *Kurdyka-Łojasiewicz* (KL) inequality without any reference to the pioneering work by Polyak. We suggest using the name of *Polyak-Łojasiewicz-Kurdyka* (PLK) conditions for the properties of this type formulated below. More details on the employed subgradient constructions and their calculi can be found in [19, 20, 27]. We also refer the reader to [17] for important calculus rules for PLK exponents and their applications.

^{*}Federal University of Goiás, Goiânia, Goiás, BR (glaydston@ufg)

[†]Wayne State University, Detroit, MI 48202, USA (aa1086@wayne.edu)

[‡]Federal University of Goiás, Goiânia, Goiás, BR (tiago.mota@discente.ufg)

[§]Corvinus Centre for Operations Research, Corvinus Institute for Advanced Studies, Corvinus University of Budapest, and School of Data Sciences (Chinese University of Hong Kong (Shenzhen); part time). Professor emeritus at CORE/INMA, UCLouvain, Belgium.

Definition 1 (PLK conditions). Let $\phi : \mathbb{R}^n \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ be an extended-real-valued lower semicontinuous (l.s.c.) function with the domain $\text{dom } \phi := \{x \in \mathbb{R}^n \mid \phi(x) < \infty\}$. We say that the function ϕ satisfies:

(i) The *basic Polyak-Łojasiewicz-Kurdyka (PLK) condition* at $\bar{x} \in \text{dom } \phi$ if there exist a number $\eta \in (0, \infty)$, a neighborhood U of \bar{x} , and a concave continuous function $\varphi : [0, \eta] \rightarrow [0, \infty)$, called the *desingularizing function*, such that

$$\varphi(0) = 0, \quad \varphi \in C^1(0, \eta), \quad \varphi'(s) > 0 \text{ for all } s \in (0, \eta), \text{ and} \quad (2)$$

$$\varphi'(\phi(x) - \phi(\bar{x})) \text{dist}(0, \partial\phi(x)) \geq 1 \text{ for all } x \in U \cap [\phi(\bar{x}) < \phi(x) < \phi(\bar{x}) + \eta] \quad (3)$$

where $\partial\phi(\bar{x})$ stands for the *Mordukhovich/limiting subdifferential* of ϕ at \bar{x} defined by

$$\partial\phi(\bar{x}) := \left\{ v \in \mathbb{R}^n \mid \exists x_k \xrightarrow{\phi} \bar{x}, v_k \in \widehat{\partial}\phi(x_k), \quad v_k \rightarrow v \text{ as } k \rightarrow \infty \right\}, \quad (4)$$

with $x_k \xrightarrow{\phi} \bar{x}$ meaning that $x_k \rightarrow \bar{x}$, $\phi(x_k) \rightarrow \phi(\bar{x})$, and with $v_k \in \widehat{\partial}\phi(x_k)$ meaning that

$$\limsup_{u \rightarrow x_k} \frac{\phi(u) - \phi(x_k) - \langle v_k, u - x_k \rangle}{\|u - x_k\|} \geq 0, \quad k \in \mathbb{N} := \{1, 2, \dots\}.$$

(ii) The *symmetric PLK condition* at \bar{x} if ϕ is continuous around \bar{x} and $\partial\phi$ is replaced in (3) by the *symmetric subdifferential* of ϕ at \bar{x} defined by

$$\partial^0\phi(\bar{x}) := \partial\phi(\bar{x}) \cup (-\partial(-\phi)(\bar{x})). \quad (5)$$

(iii) The *strong PLK condition* if ϕ is Lipschitz continuous around \bar{x} and ∂ is replaced in (3) by the *Clarke/-convexified subdifferential* of ϕ at \bar{x} given by

$$\overline{\partial}\phi(\bar{x}) := \text{co } \partial\phi(\bar{x}), \quad (6)$$

where “co” stands for the convex hull of the set in question.

(iv) The *exponent versions* of the *PLK conditions* in (i)–(iii) if the desingularizing function in (2) and (3) is selected in the form $\varphi(t) = Mt^{1-q}$, where M is a positive constant, and where $q \in [0, 1]$. We refer to the case where $q \in (0, 1/2)$ as the *PLK conditions with lower exponents*.

In [6], Attouch, Bolté and Svaiter studied minimization of l.s.c. functions $\phi : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ by using a *generic* class of *descent methods* satisfying the following properties:

(H1) *Sufficient decrease*: for each $k \in \mathbb{N}$, we have

$$\phi(x^{k+1}) + a\|x^{k+1} - x^k\|^2 \leq \phi(x^k). \quad (7)$$

(H2) *Relative error*: for each $k \in \mathbb{N}$, there exists

$$w^{k+1} \in \partial\phi(x^{k+1}) \text{ with } \|w^{k+1}\| \leq b\|x^{k+1} - x^k\|, \quad (8)$$

with $a, b > 0$. It is shown in [6] that a great variety of important algorithms of optimization satisfy conditions (H1) and (H2). Employing the basic PLK property from Definition 1(i), a qualitative convergence analysis, with arriving at the *limiting/(M)ordukhovich-stationary point* as $0 \in \partial\phi(\bar{x})$, is developed in [6] for the general class of descent algorithms satisfying (H1) and (H2) while without establishing any convergence rate.¹ A detailed qualitative (with convergence rates) analysis of the generic class of descent algorithms under the exponent PLK conditions married to (H1) and (H2) was developed in [12], where the the obtained results were similar to [5] established for the proximal algorithm; namely, the finite step convergence if $q = 0$, linear convergence if $q \in (0, 1/2]$, and polynomial convergence if $q \in (1/2, 1)$.

A major motivation for our study came from the paper by Aragón-Artacho and Vuong [4] who observed that the relative error condition (H2) may fail for the *Boosted Difference of Convex Algorithm* (BDCA) proposed in [2] for DC (difference of convex) programs. As shown in [4], the convergence analysis of [6] can be given for BDCA *provided that* the relative error condition (8) is replaced by

$$\text{there exists } w^k \in \overline{\partial}\phi(x^k) \text{ with } \|w^k\| \leq b\|x^{k+1} - x^k\|, \quad k \in \mathbb{N}, \quad (9)$$

¹In fact, an additional technical and nonrestrictive assumption, labeled as (H4) in what follows, was imposed in [6] and the subsequent publications discussed below.

expressed in terms of the convexified subdifferential (6). To furnish this, it is assumed in [4] that the functions g and h in the DC decomposition $\phi = g - h$ are strongly convex with g being of class $\mathcal{C}^{1,1}$, and that the basic PLK condition in (3) is replaced by its strong version from Definition 1(iii). Under these requirements, [4] establishes the convergence of BDCA iterates to a *(C)larke-stationary point* $0 \in \partial\phi(\bar{x})$.

In the current paper, besides establishing new results on the convergence rate for generic algorithms satisfying conditions $(\mathcal{H}1)$ and $(\mathcal{H}2)$ under the PLK properties, we also investigate another generic class of algorithms where the relative error condition $(\mathcal{H}2)$ is replaced by the following one:

$(\mathcal{H}3)$ *Modified relative error*: for each $k \in \mathbb{N}$, there exists

$$w^k \in \partial\phi(x^k) \text{ with } \|w^k\| \leq b\|x^{k+1} - x^k\| \quad (10)$$

expressed in terms of the limiting subdifferential (4). The convergence analysis conducted below allows us to achieve, under the basic PLK property, the global convergence of any algorithm satisfying $(\mathcal{H}1)$ and $(\mathcal{H}3)$ to an M -stationary point $0 \in \partial\phi(\bar{x})$. This answers in the affirmative the question posted in [4, Remark 4.5], even with the improvement (10) of the stronger condition in (9) for general l.s.c. functions.

While the global convergence results under $(\mathcal{H}3)$ are parallel to those developed under $(\mathcal{H}2)$, this is not the case when it comes to convergence rate analysis. In what follows, we prove that the *lower exponent PLK* condition ($0 < q < 1/2$) always yields the *finite termination* of any descent algorithm satisfying $(\mathcal{H}1)$ and $(\mathcal{H}3)$. On the other hand, we present an example showing that the finite termination *fails* under PLK lower exponents for the proximal method, which surely satisfies $(\mathcal{H}1)$ and $(\mathcal{H}2)$.

Moreover, it comes as a surprise that in the case of difference programs to minimize $\phi = g - h$, where $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ is convex while g is of class $\mathcal{C}^{1,1}$ without any convexity assumption, there exists *no desingularizing function* of the type $\varphi(t) = t^{1-q}$ with $q \in (0, 1/2)$ for which the PLK condition holds at *local minimizers* of ϕ .

The rest of the paper is organized as follows. Section 2 establishes *global convergence* of descent algorithms satisfying the conditions in $(\mathcal{H}1)$ and $(\mathcal{H}3)$, and in parallel to those in $(\mathcal{H}1)$ and $(\mathcal{H}2)$, to the corresponding stationary points defined by the subdifferentials in the PLK properties from Definition 1(i–iii).

In Section 3, we develop a novel *rate convergence analysis* of the generic algorithms satisfying $(\mathcal{H}1)$ and $(\mathcal{H}2)$ under the exponent PLK conditions. A striking result of this section shows that the *lower exponent PLK* case exhibits *either the finite termination, or the superlinear* global convergence rate. It is a serious improvement of the linear convergence rate known in such settings. The failure of finite termination is demonstrated here by a simple example for the proximal algorithm.

Section 4 continues our rate convergence analysis under the exponent PLK conditions, while now with replacing $(\mathcal{H}2)$ but its *modified relative error* counterpart $(\mathcal{H}3)$. The main novelty here is to show that any algorithm satisfying $(\mathcal{H}1)$ and $(\mathcal{H}3)$ *always terminates in finite steps* under the lower exponent PLK conditions. Moreover, we present an additional assumption under which the latter conditions are *inconsistent* with the properties in $(\mathcal{H}1)$ and $(\mathcal{H}3)$.

Section 5 contains some specifications of applications to particular algorithms that lie within the generic class of algorithms satisfying the conditions in $(\mathcal{H}1)$ and $(\mathcal{H}3)$. These specific algorithms concern nonsmooth versions of the inexact reduced gradient methods recently proposed in [14] as well as the aforementioned BDCA. It will be demonstrated that the obtained results for the generic descent methods lead us to significant improvements and extensions of the previously known ones.

Section 6 addresses the fulfillment of the PLK conditions with lower exponents $q \in (0, 1/2)$ in the case of minimizing functions that admit the difference decomposition $\phi = g - h$. We show that the $\mathcal{C}^{1,1}$ structure of g and the convexity of h is *inconsistent* with the lower exponent PLK of ϕ at local minimizers. Section 7 summarizes the major achievements of the paper and discusses some directions of the future research.

The notation of this paper is standard in variational analysis and optimization. Recall that $B(x, r)$ stands for the open ball centered at x with radius $r > 0$.

2 Global Convergence of Generic Algorithms

This section studies in parallel the two classes of descent algorithms satisfying the generic conditions in **(a)**: $(\mathcal{H}1)$ and $(\mathcal{H}2)$, and in **(b)**: $(\mathcal{H}1)$ and $(\mathcal{H}3)$, respectively. Similarly to [6], we add to our convergence analysis the following technical assumption:

$(\mathcal{H}4)$ *Continuity condition*: There are a subsequence $\{x^{k_j}\}$ and $\bar{x} \in \text{dom } \phi$ such that

$$x^{k_j} \longrightarrow \bar{x} \text{ and } \phi(x^{k_j}) \longrightarrow \phi(\bar{x}) \text{ as } j \longrightarrow \infty.$$

From now on, unless otherwise stated, condition $(\mathcal{H}4)$ is assumed to be satisfied.

The major theorem below tells us that the basic PLK property from Definition 1 ensures the global convergence of generic algorithms in both cases (a) and (b). In Remark 1, we discuss the corresponding modifications of the convergence results under the symmetric and strong PLK conditions formulated in Definition 1(ii,iii).

Theorem 1 (global convergence of generic algorithms under the basic PLK). *Let $\phi : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be a proper l.s.c. function bounded from below, and let the sequence $\{x^k\}$ be constructed by the generic algorithms satisfying either $(\mathcal{H}1)$ and $(\mathcal{H}2)$, or $(\mathcal{H}1)$ and $(\mathcal{H}3)$ properties. If the basic PLK condition holds at some accumulation point $\bar{x} \in \text{dom } \phi$ of $\{x^k\}$, then we have that*

$$\sum_{k=0}^{\infty} \|x^k - x^{k+1}\| < \infty, \quad (11)$$

and that $\{x^k\}$ converges to \bar{x} as $k \rightarrow \infty$. Moreover, \bar{x} is an M -stationary point of ϕ .

Proof. The result under $(\mathcal{H}1)$ and $(\mathcal{H}2)$ is known from [6, Theorem 2.9], and thus we proceed with the proof in the case where $(\mathcal{H}1)$ and $(\mathcal{H}3)$ hold. Since \bar{x} is an accumulation point of $\{x^k\}$, there exists a subsequence $\{x^{k_j}\}$ of $\{x^k\}$ converging to \bar{x} as $j \rightarrow \infty$. By the boundedness from below of ϕ and the decreasing of $\{\phi(x^k)\}$, the continuity condition $(\mathcal{H}4)$ implies that $\{\phi(x^k)\}$ converges to $\phi(\bar{x})$ as $k \rightarrow \infty$ with $\phi(\bar{x}) < \phi(x^k)$, $k \in \mathbb{N}$. In particular, we get

$$\phi(\bar{x}) < \phi(x^k) < \phi(\bar{x}) + \eta \text{ for large } k \in \mathbb{N} \text{ and small } \eta > 0.$$

Taking the latter into account, define the sequence $\{b_k\}$ by

$$b_k := \|x^k - \bar{x}\| + (a^{-1}(\phi(x^k) - \phi(\bar{x})))^{\frac{1}{2}} + ba^{-1}\varphi(\phi(x^k) - \phi(\bar{x}))$$

and deduce from the continuity of φ that the origin $0 \in \mathbb{R}^n$ is an accumulation point of $\{b_k\}$. Thus there exists $k_0 := k_{j_0}$ such that the inequalities

$$\|x^{k_0} - \bar{x}\| + a^{-1}(\phi(x^{k_0}) - \phi(\bar{x}))^{\frac{1}{2}} + ba^{-1}\varphi(\phi(x^{k_0}) - \phi(\bar{x})) < \epsilon, \quad (12)$$

$$\phi(\bar{x}) < \phi(x^{k_0}) < \phi(\bar{x}) + \eta \quad (13)$$

are satisfied and yield in turn the inclusions

$$x^{k_0} \in B(\bar{x}, \epsilon) \cap [\phi(\bar{x}) < \phi < \phi(\bar{x}) + \eta] \subset U \cap [\phi(\bar{x}) < \phi < \phi(\bar{x}) + \eta].$$

Moreover, conditions (7) and (13) ensure that

$$0 \leq \phi(x^{k_0+1}) - \phi(\bar{x}) \leq \phi(x^{k_0}) - \phi(\bar{x}) < \eta, \quad (14)$$

which implies therefore the inequality

$$\varphi(\phi(x^{k_0}) - \phi(\bar{x})) - \varphi(\phi(x^{k_0+1}) - \phi(\bar{x})) \geq \varphi'(\phi(x^{k_0}) - \phi(\bar{x}))(\phi(x^{k_0}) - \phi(x^{k_0+1}))$$

by the concavity of φ . Combining (7) and (10) with $k := k_0$, we have

$$\phi(x^{k_0}) - \phi(x^{k_0+1}) \geq ba^{-1}\|x^{k_0+1} - x^{k_0}\| \cdot \|w^{k_0}\|.$$

The last two inequalities above lead us to

$$\begin{aligned} & \varphi(\phi(x^{k_0}) - \phi(\bar{x})) - \varphi(\phi(x^{k_0+1}) - \phi(\bar{x})) \\ & \geq \frac{b}{a}\varphi'(\phi(x^{k_0}) - \phi(\bar{x}))\|w^{k_0}\|\|x^{k_0} - x^{k_0+1}\| \end{aligned} \quad (15)$$

with $w^{k_0} \in \partial f(x^{k_0})$, and so the basic PLK property of f at \bar{x} yields $0 \notin \partial f(x^{k_0})$ while ensuring together with (3) that

$$\varphi'(\phi(x^{k_0}) - \phi(\bar{x}))\|w^{k_0}\| \geq \varphi'(\phi(x^{k_0}) - \phi(\bar{x}))\text{dist}(0, \partial\phi(x^{k_0})) \geq 1.$$

Combining the latter with (15) justifies the estimates

$$ba^{-1}(\varphi(\phi(x^k) - \phi(\bar{x})) - \varphi(\phi(x^{k+1}) - \phi(\bar{x}))) \geq \|x^k - x^{k+1}\|, \quad k = k_0. \quad (16)$$

Our next step is to verify the inclusion

$$x^k \in B(\bar{x}, \epsilon) \text{ for all } k \geq k_0. \quad (17)$$

To proceed by induction, observe that we have already proved that (17) holds for $k = k_0$ and now aim at showing that $x^{k_0+1} \in B(\bar{x}, \epsilon)$. It follows from (7) and (14) that

$$\|x^{k_0} - x^{k_0+1}\| \leq (a^{-1}(\phi(x^{k_0}) - \phi(x^{k_0+1})))^{\frac{1}{2}} < (a^{-1}(\phi(x^{k_0}) - \phi(\bar{x})))^{\frac{1}{2}}. \quad (18)$$

The last estimate along with the triangle inequality gives us

$$\|\bar{x} - x^{k_0+1}\| \leq \|\bar{x} - x^{k_0}\| + \|x^{k_0} - x^{k_0+1}\| \leq \|\bar{x} - x^{k_0}\| + (a^{-1}(\phi(x^{k_0}) - \phi(\bar{x})))^{\frac{1}{2}}.$$

Combining this with (12) implies that $x^{k_0+1} \in B(\bar{x}, \epsilon)$. Take $j > 1$ and assume that (17) holds for all $k = k_0 + 1, \dots, k_0 + j - 1$. In this case, (16) is satisfied for $k = k_0 + 1, \dots, k_0 + j - 1$. Therefore,

$$\sum_{i=1}^{j-1} \|x^{i+k_0} - x^{i+k_0+1}\| \leq ba^{-1}(\varphi(\phi(x^{k_0+1}) - \phi(\bar{x})) - \varphi(\phi(x^{k_0+j}) - \phi(\bar{x}))).$$

Using the triangle inequality again tells us that

$$\|\bar{x} - x^{k_0+j}\| \leq \sum_{i=1}^{j-1} \|x^{i+k_0} - x^{i+k_0+1}\| + \|x^{k_0} - x^{k_0+1}\| + \|\bar{x} - x^{k_0}\|.$$

Now we combine the last two inequalities with (18) to get the estimates

$$\begin{aligned} \|\bar{x} - x^{k_0+j}\| &\leq ba^{-1}\varphi(\phi(x^{k_0+1}) - \phi(\bar{x})) + (a^{-1}(\phi(x^{k_0+1}) - \phi(\bar{x})))^{\frac{1}{2}} + \|\bar{x} - x^{k_0}\|, \\ &\leq b^{-1}a\varphi(\phi(x^{k_0}) - \phi(\bar{x})) + (a^{-1}(\phi(x^{k_0+1}) - \phi(\bar{x})))^{\frac{1}{2}} + \|\bar{x} - x^{k_0}\|, \end{aligned}$$

where the second one follows from (14) and the fact that φ is increasing. The latter inequality along with (12) ensures that $x^{k_0+j} \in B(\bar{x}, \epsilon)$, which verifies (17) and shows that (16) holds for all $k \geq k_0$. Thus we arrive at

$$\sum_{k=k_0}^{\infty} \|x^k - x^{k+1}\| \leq ba^{-1}\varphi(\phi(x^{k_0}) - \phi(\bar{x})) \text{ whenever } k > k_0,$$

which clearly yields (11). This tells us that $\{x^k\}$ is a Cauchy sequence and hence converges to \bar{x} since it is an accumulation point of $\{x^k\}$. It follows from (10) and (11) that the sequence $\{w^k\}$ converges to zero. By definition (4) of the limiting subdifferential $\partial\phi$ and the convergence $\phi(x^k) \rightarrow \phi(\bar{x})$ as $k \rightarrow \infty$ shown above, we get that $0 \in \partial\phi(\bar{x})$, i.e., \bar{x} is an M -stationary point of ϕ , and the proof is complete. \square

The following remark reveals global convergence properties of generic descent methods under the *symmetric* and *strong* PLK conditions.

Remark 1 (convergence under symmetric and strong PLK conditions).

- (i) Theorem 1 resolves in the affirmative the question posed in [4, Remark 4.5] about the possibility to provide a counterpart of the convergence analysis in [6] under the modified error bound condition (10) in $(\mathcal{H}3)$.
- (ii) It follows from the proof of Theorem 1 that our analysis works with replacing (10) by (9) under the strong PLK condition from Definition 1(iii) and also by using the symmetric subdifferential (5) in (10) under the symmetric PLK property from Definition 1(ii). However, such replacements lead us to the less informative C -stationary $0 \in \bar{\partial}\phi(\bar{x})$ and *symmetric stationary* points $0 \in \partial^0\phi(\bar{x})$, respectively. Similar conclusions concerning the strong and symmetric PLK conditions can be made for generic algorithms in the case of $(\mathcal{H}2)$. Moreover, the same observation applies to the *convergence rate results* presented in Section 3 and Section 4 below.

3 Convergence Rates under Relative Error Condition

In this section, we provide explicit convergence rates in terms of the PLK exponents of the desingularizing function $\varphi(t) = t^{1-q}$, $q \in [0, 1)$, for generic algorithms that satisfy the conditions in (H1) and (H2). While the convergence rate results obtained under the exponent PLK property for $q = 0$ and $q \in [1/2, 1)$ are generic extensions of those given for particular algorithms (see, e.g., [2, 3, 4, 5, 16, 17, 14, 21] and the references therein), it is worth highlighting that the main challenge lies in the case where $q \in (0, 1/2)$ for which we provide significant convergence rate improvements.

To proceed, we first present two technical lemmas that will be instrumental for our subsequent results. The idea behind the proof of the first lemma can be found within the proof of [1, Theorem 3.2]. As for the second lemma, to the best of our knowledge, it has not been previously documented in the literature and will play a crucial role in our generic convergence rate analysis.

Lemma 1 (estimates for monotone sequences). *Let $\{a_k\} \subset [0, \infty)$ be a monotonically decreasing sequence, and let $q \in (0, 1) \cup (1, 2)$. Then we have the estimate*

$$\sum_{j=k}^{k+l} \frac{a_j - a_{j+1}}{a_j^q} \leq \frac{1}{1-q} \left(a_k^{1-q} - a_{k+l+1}^{1-q} \right) \quad \text{whenever } k, l \in \mathbb{N}.$$

Proof. Since $\{a_k\}$ is monotonically decreasing, it follows that $a_j \geq t$ for all $t \in [a_{j+1}, a_j]$ and $j \in \mathbb{N}$. This leads us to

$$\frac{1}{(a_j)^q} \leq \frac{1}{t^q}, \quad t \in [a_{j+1}, a_j], \quad q \in (0, 1),$$

which implies in turn that

$$\frac{a_j - a_{j+1}}{a_j^q} = \int_{a_{j+1}}^{a_j} \frac{1}{a_j^q} dt \leq \int_{a_{j+1}}^{a_j} \frac{1}{t^q} dt.$$

The latter readily ensures that

$$\frac{a_j - a_{j+1}}{a_j^q} \leq \frac{1}{1-q} \left(a_j^{1-q} - a_{j+1}^{1-q} \right), \quad j \in \mathbb{N},$$

and thus we arrive at the claimed estimate. \square

The next lemma is essential to establish the main results of this section on either finite or superlinear convergence of generic algorithms satisfying the conditions in (H1) and (H2) under the PLK lower exponents.

Lemma 2 (convergence rates for monotone sequences). *Let $\sigma > 0$ and $p \in (0, 1)$, and let $\{a_k\} \subset [0, \infty)$ be a monotonically decreasing sequence. Assume that*

$$a_{k+1} \leq a_k - \sigma a_{k+1}^p \quad k \in \mathbb{N}. \quad (19)$$

If $\{a_k\}$ converges to \bar{a} , then $\bar{a} = 0$ and one of the following two conditions holds:

- (i) *The sequence $\{a_k\}$ converges superlinearly to zero.*
- (ii) *There exist $k_0 \in \mathbb{N}$ such that $a_k = 0$ for all $k \geq k_0$.*

Proof. Since $a_k \geq 0$ for all $k \in \mathbb{N}$, we have $\bar{a} \geq 0$ be the limit of this sequence. On the other hand, the passage to the limit in (19) with taking into account that $\sigma > 0$ yields

$$\bar{a} \leq \bar{a} - \sigma \bar{a}^p \implies \sigma \bar{a}^p \leq 0,$$

which implies that $\bar{a} = 0$. Fix $p \in (0, 1)$ and observe that if there exists $k_0 \in \mathbb{N}$ such that $a_{k_0} = 0$, then (ii) holds. Suppose now that there exists no $k_0 \in \mathbb{N}$ with $a_{k_0} = 0$, i.e., $a_k > 0$ for all $k \in \mathbb{N}$. Then it follows from (19) that

$$a_{k+1} + \sigma a_{k+1}^p \leq a_k, \quad k \in \mathbb{N}. \quad (20)$$

Remembering that $a_k > 0$ for all $k \in \mathbb{N}$, we divide both sides of (20) by a_{k+1} and take the limit as $k \rightarrow \infty$ therein. This gives us

$$\lim_{k \rightarrow \infty} \left(1 + \sigma a_{k+1}^{p-1} \right) \leq \lim_{k \rightarrow \infty} \frac{a_k}{a_{k+1}}.$$

By $a_k \downarrow 0$ and $p - 1 < 0$, the obtained inequality yields $\lim_{k \rightarrow \infty} \frac{a_k}{a_{k+1}} = \infty$, which indicates that $\{a_k\}$ converges superlinearly to zero and thus completes the proof. \square

Now we are ready to establish the first result on convergence rates of any generic algorithm satisfying $(\mathcal{H}1)$ and $(\mathcal{H}2)$ under the exponent PLK conditions. The theorem below concerns the *value convergence* with the most significant improvement of the known results in the case of lower exponents.

Theorem 2 (value convergence rates under exponent PLK and relative error conditions). *In the setting of Theorem 1, let $\{x_k\}$ be a generic sequence of iterates satisfying $(\mathcal{H}1)$ and $(\mathcal{H}2)$, and let the exponent PLK property of ϕ hold at \bar{x} with $\varphi(t) = Mt^{1-q}$ for some $M > 0$ and $q \in [0, 1)$. The following convergence rates are guaranteed for the value sequence $\{\phi(x^k)\}$:*

- (i) *If $q = 0$, then the sequence $\{\phi(x^k)\}$ terminates at $\phi(\bar{x})$ in finite steps.*
- (ii) *If $q \in (0, \frac{1}{2})$, then either $\{\phi(x^k)\}$ terminates at $\phi(\bar{x})$ in a finite number of steps, or $\{\phi(x^k)\}$ converges superlinearly to $\phi(\bar{x})$ as $k \rightarrow \infty$.*
- (iii) *If $q = 1/2$, then $\{\phi(x^k)\}$ linearly converges to $\phi(\bar{x})$ as $k \rightarrow \infty$.*
- (iv) *If $q \in (1/2, 1)$, then there exists a positive constant η such that*

$$\phi(x^k) - \phi(\bar{x}) \leq \eta k^{-\frac{1}{2q-1}} \quad \text{for all large } k \in \mathbb{N}. \quad (21)$$

Proof. It follows from Theorem 1 that $x^k \rightarrow \bar{x}$ as $k \rightarrow \infty$ and that \bar{x} is an M -stationary point of ϕ . Defining $\psi(x) := \phi(x) - \phi(\bar{x})$, we see that the sequence $\{\psi(x^k)\} \subset [0, \infty)$ is decreasing. Using (3) with $\varphi(t) = Mt^{1-q}$ as $q \in [0, 1)$ and $M > 0$, it follows that

$$\begin{aligned} \|w^{k+1}\| &\geq \text{dist}(0, \partial\phi(x^{k+1})) \geq \frac{1}{M(1-q)} (\phi(x^{k+1}) - \phi(\bar{x}))^q \\ &\text{with } w^{k+1} \in \partial\phi(x^{k+1}), \quad k \in \mathbb{N}. \end{aligned} \quad (22)$$

To verify (i), suppose on the contrary that the sequence $\{x^k\}$ is infinitely generated and deduce from (22) with $q = 0$ that $\|w^{k+1}\| \geq M^{-1}$ for all large k . This tells us together with (8) in $(\mathcal{H}2)$ that $(Mb)^{-1} \leq \|x^{k+1} - x^k\|$, which contradicts the fact that $\{x^k\}$ converges to \bar{x} and thus justifies the statement in (i).

For the proof of (ii), subtract $\phi(\bar{x})$ from both sides of (7) and get

$$\psi(x^{k+1}) \leq \psi(x^k) - a\|x^{k+1} - x^k\|^2,$$

which being combined with (8) yields

$$\psi(x^{k+1}) \leq \psi(x^k) - \frac{a}{b^2}\|w^{k+1}\|^2.$$

Using the obtained inequality together with (22) and the definition of ψ gives us

$$\psi(x^{k+1}) \leq \psi(x^k) - \frac{a}{b^2 M^2 (1-q)^2} (\psi(x^{k+1}))^{2q}. \quad (23)$$

It follows from Lemma 2 with $\sigma = \frac{a}{b^2 M^2 (1-q)^2}$ and $p = 2q$ that if $q \in (0, 1/2)$, then the sequence $\{\psi(x^k)\}$ is either finite or converges superlinearly, and thus (ii) is verified.

To prove (iii), let $q = 1/2$ and $\alpha = \frac{4a}{b^2 M^2}$. By (23) we get

$$\psi(x^{k+1}) \leq \frac{1}{1+\alpha} \psi(x^k). \quad (24)$$

Since $1/(1+\alpha) \in (0, 1)$, assertion (iii) is a consequence of (24).

It remains to verify (iv). For $q \in (1/2, 1)$, it follows from (23) with $\beta := \frac{a}{b^2 M^2 (1-q)^2}$ that we have

$$\beta \leq \frac{\psi(x^k) - \psi(x^{k+1})}{\psi(x^{k+1})^{2q}} = [\psi(x^k) - \psi(x^{k+1})] \Phi(\psi(x^{k+1})), \quad (25)$$

where $\Phi(t) := \frac{1}{t^{2q}}$. Suppose first that there exists $Q \in (1, \infty)$ such that $\Phi(\psi(x^{k+1})) \leq Q\Phi(\psi(x^k))$. Combining this fact with (25) and Lemma 1 brings us to

$$\begin{aligned}\beta &\leq \frac{\psi(x^k) - \psi(x^{k+1})}{\psi(x^{k+1})^{2q}} = \int_{\psi(x^{k+1})}^{\psi(x^k)} \Phi(\psi(x^{k+1})) dt \leq Q \int_{\psi(x^{k+1})}^{\psi(x^k)} \Phi(\psi(x^k)) dt \\ &\leq \frac{Q}{1-2q} [(\psi(x^k))^{1-2q} - (\psi(x^{k+1}))^{1-2q}] = \frac{Q}{2q-1} [(\psi(x^{k+1}))^{1-2q} - (\psi(x^k))^{1-2q}],\end{aligned}$$

which implies by $2q - 1 > 0$ that

$$\frac{\beta(2q-1)}{Q} \leq (\psi(x^{k+1}))^{1-2q} - (\psi(x^k))^{1-2q}, \quad k \geq k_0. \quad (26)$$

Letting $j - 1 > k_0$, we deduce from (26) that

$$\sum_{k=k_0}^j \frac{\beta(2q-1)}{Q} \leq \sum_{k=k_0}^j (\psi(x^{k+1}))^{1-2q} - (\psi(x^k))^{1-2q}$$

giving us in turn that

$$\frac{(j-1-k_0)\beta(2q-1)}{Q} \leq (\psi(x^j))^{1-2q} - (\psi(x^{k_0}))^{1-2q}.$$

Therefore, we arrive at the estimate

$$(\psi(x^j))^{1-2q} \geq \frac{(j-1-k_0)\beta(2q-1)}{Q} + (\psi(x^{k_0}))^{1-2q}.$$

Since the function $t \mapsto t^{-\frac{1}{2q-1}}$ is decreasing for $q \in (1/2, 1)$, it follows that

$$\psi(x^j) \leq \left(\frac{(j-1-k_0)\beta(2q-1)}{Q} + (\psi(x^{k_0}))^{1-2q} \right)^{-\frac{1}{2q-1}},$$

which ensures the existence of some $\eta > 0$ with

$$\psi(x^j) \leq \eta j^{-\frac{1}{2q-1}}.$$

and thus verifies the convergence rate estimate in (21) in this case.

Consider now the alternative situation, where the number Q used above does not exist, i.e., whenever $Q \in (1, \infty)$ we have $\Phi(\psi(x^{k+1})) > Q\Phi(\psi(x^k))$ for all k sufficiently. Fix some $Q \in (1, \infty)$ and define $Q_1 := \frac{1}{Q^{2q}}$. Then it follows from the definitions that $\psi(x^{k+1}) \leq Q_1\psi(x^k)$ and hence $\psi(x^{k+1})^{1-2q} \geq Q_1^{1-2q}\psi(x^k)^{1-2q}$ by $1-2q < 0$. Subtracting $\psi(x^k)^{1-2q}$ from both sides of the latter inequality gives us

$$\psi(x^{k+1})^{1-2q} - \psi(x^k)^{1-2q} \geq (Q_1^{1-2q} - 1)\psi(x^k)^{1-2q}.$$

Note that $Q_1 \in (0, 1)$ and therefore $Q_1^{1-2q} > 1$ for all $q \in (1/2, 1)$. The convergence $\psi(x^k) \rightarrow 0$ yields $\psi(x^k)^{1-2q} \rightarrow \infty$ as $k \rightarrow \infty$, and hence

$$\psi(x^{k+1})^{1-2q} - \psi(x^k)^{1-2q} \geq \frac{\beta(2q-1)}{Q}$$

whenever k is sufficiently large. The rest of the proof is similar to the above arguments based on (26), and thus we are done with verifying (iv) and the whole theorem. \square

The next theorem establishes the convergence rates for the sequence of iterates $\{x^k\}$ of any algorithm satisfying $(\mathcal{H}1)$ and $(\mathcal{H}2)$ under the exponent PLK conditions.

Theorem 3 (convergence rates of iterates under exponent PLK and relative error conditions). *In the setting of Theorem 2, the following assertions hold:*

- (i) *If $q = 0$, then the sequence $\{x^k\}$ terminates at \bar{x} in finite steps.*

- (ii) If $q \in (0, \frac{1}{2})$, then either $\{x^k\}$ terminates at \bar{x} in finite steps, or $\{x^k\}$ converges superlinearly to \bar{x} as $k \rightarrow \infty$.
- (iii) If $q = 1/2$, then $\{x^k\}$ converges linearly to \bar{x} as $k \rightarrow \infty$.
- (iv) If $q \in (1/2, 1)$, then there exists a positive constant σ such that

$$\|x^k - \bar{x}\| \leq \sigma k^{-\frac{1-q}{2q-1}} \quad \text{for all } k \text{ sufficiently large.}$$

Proof. By taking into account that

$$0 \leq a\|x^{k+1} - x^k\|^2 \leq \phi(x^k) - \phi(x^{k+1}), \quad k \in \mathbb{N},$$

assertion (i) of this theorem is an immediate consequence of Theorem 2(i). To verify (ii), deduce first from the concavity of φ in (3) together with $(\mathcal{H}1)$ and $(\mathcal{H}2)$ that

$$\frac{\|x^{k+1} - x^k\|^2}{\|x^k - x^{k-1}\|} \leq \frac{b}{a} [\varphi(\phi(x^k) - \phi(\bar{x})) - \varphi(\phi(x^{k+1}) - \phi(\bar{x}))]. \quad (27)$$

for all large $k \in \mathbb{N}$. Given $r \in (0, 1)$, we have the two possibilities for such k :

- (a) $\|x^{k+1} - x^k\| \geq r\|x^k - x^{k-1}\|$.
- (b) $\|x^{k+1} - x^k\| < r\|x^k - x^{k-1}\|$.

In case (a), the usage of (27) with $\varphi(t) = Mt^{1-q}$ and $\psi(x) = \phi(x) - \phi(\bar{x})$ yields

$$\|x^{k+1} - x^k\| \leq \frac{bM}{ar} (\psi(x^k)^{1-q} - \psi(x^{k+1})^{1-q}) \leq r\|x^k - x^{k-1}\| + \frac{bM}{ar} (\psi(x^k)^{1-q} - \psi(x^{k+1})^{1-q}).$$

On the other hand, in case (b) we readily get

$$\|x^{k+1} - x^k\| \leq r\|x^k - x^{k-1}\| \leq r\|x^k - x^{k-1}\| + \frac{bM}{ar} (\psi(x^k)^{1-q} - \psi(x^{k+1})^{1-q}),$$

and therefore arrive at the inequality

$$\|x^{k+1} - x^k\| \leq r\|x^k - x^{k-1}\| + \frac{bM}{ar} (\psi(x^k)^{1-q} - \psi(x^{k+1})^{1-q}).$$

The latter provides the following estimates valid for all the natural numbers $l \geq 1$:

$$\sum_{j=k}^{k+l} \|x^{j+1} - x^j\| \leq \frac{r}{(1-r)} \|x^k - x^{k-1}\| + \frac{bM}{ar(1-r)} (\psi(x^k)^{1-q} - \psi(x^{k+l+1})^{1-q}). \quad (28)$$

Define $s_k := \sum_{j=k}^{\infty} \|x^{j+1} - x^j\|$ and let $l \rightarrow \infty$ in (28). Remembering that $\{\psi(x^k)\}$ decreases to zero, we get from (28) that

$$s_k \leq \frac{r}{(1-r)} \|x^k - x^{k-1}\| + \frac{bM}{ar(1-r)} \psi(x^k)^{1-q}.$$

Combining the last inequality with $(\mathcal{H}1)$ and using $\phi(x^{k-1}) - \phi(x^k) \leq \phi(x^{k-1}) - \phi(\bar{x}) = \psi(x^{k-1})$ bring us to the estimate

$$s_k \leq \frac{r}{a^{1/2}(1-r)} (\psi(x^{k-1}))^{1/2} + \frac{bM}{ar(1-r)} \psi(x^k)^{1-q}. \quad (29)$$

When $q \in (0, 1/2)$, it follows from the above that

$$s_k \leq \frac{r}{a^{1/2}(1-r)} (\psi(x^{k-1}))^{1/2} + \frac{bM}{ar(1-r)} \psi(x^k)^{1/2}. \quad (30)$$

Since the sequence $\{\psi(x^k)\}$ is monotonically decreasing, we deduce from (30) that

$$s_k \leq C(\psi(x^{k-1}))^{1/2} \quad \text{with } C := \frac{a^{1/2}r^2 + bM}{ar(1-r)},$$

which allows us to verify (ii) by using Theorem 2 and the estimate $\|x^k - \bar{x}\| \leq s_k$. Assertions (iii) and (iv) of this theorem follow directly from Theorem 2(iii,iv)), respectively. Indeed, in case (iii) we set $q = 1/2$ in (29), while (iv) with $q \in (1/2, 1)$ is deduced from the inequality $(\psi(x^k))^{1/2} \leq (\psi(x^k))^{1-q}$ held for large k . \square

The following example presents a family of *univariate smooth functions* for which the classical proximal algorithm, satisfying (\mathcal{H}_1) and (\mathcal{H}_2) , does *not terminate in finite steps* under the PLK conditions with lower exponents.

Example 1 (failure of finite termination for the proximal algorithm under PLK lower exponents). Consider the parametric class of functions $\phi_\alpha(x) := |x|^{1+\alpha}$ with $0 < \alpha < 1$; cf. [24, page 22], where such functions were considered from different perspectives. Each function ϕ_α is continuously differentiable, achieves its global minimum at $\bar{x} = 0$, is C^1 , and satisfies the exponent PLK condition with $q = \frac{\alpha}{1+\alpha} \in (0, 1/2)$. The classical proximal algorithm reads for ϕ_α as follows:

$$x_{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ |x|^{1+\alpha} + \frac{1}{2\lambda_k} |x - x_k|^2 \right\}, \quad \lambda_k > 0, \quad k \in \mathbb{N}. \quad (31)$$

If we suppose that $\{x_k\}$ is finite, then there exists $k_0 \in \mathbb{N}$ such that $(1 + \alpha)\operatorname{sign}(x_{k_0})|x_{k_0}|^\alpha = 0$ and consequently $x_{k_0} = 0$. Applying the stationary rule to the function in (31) gives us the equation

$$0 = (1 + \alpha)\operatorname{sign}(x_{k_0})|x_{k_0}|^\alpha + x_{k_0} - x_{k_0-1},$$

which yields $x_{k_0-1} = x_{k_0} = 0$. Recursively, we arrive at $x_0 = 0$ and thus conclude that for any starting point $x_0 \neq 0$ the sequence of iterates in (31) must be infinite. Theorem 3 tells us that $x_k \rightarrow 0$ as $k \rightarrow \infty$ superlinearly.

The next example illustrates the alternative situation in Theorem 3, where the proximal algorithm *terminates in finite steps* in contrast to Example 1.

Example 2 (finite step termination of the proximal algorithm). Let

$$\phi(x) = \begin{cases} -x, & x < 0, \\ x^{3/2}, & x \geq 0. \end{cases} \quad (32)$$

We see that the function ϕ in (32) achieves its global minimum at $\bar{x} = 0$, where it is nondifferentiable, and it satisfies the exponent PLK property with $q = 1/3 \in (0, 1/2)$ at \bar{x} . The proximal algorithm for (32) is written as

$$x_{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ \phi(x) + \frac{1}{2\lambda_k} |x - x_k|^2 \right\}, \quad \lambda_k > 0, \quad k \in \mathbb{N}, \quad (33)$$

First we claim that if $x_0 < 0$, then $x_k \leq 0$ for all $k \in \mathbb{N}$. On the contrary, suppose that there exists $k_0 \in \mathbb{N}$ for which $x_{k_0} < 0$ while $x_{k_0+1} > 0$. By the smoothness of (32) at the points under consideration, we deduce from the stationary rule applied to (33) that

$$x_{k_0} = \frac{3\lambda_{k_0}}{2} x_{k_0+1}^{\frac{1}{2}} + x_{k_0+1},$$

which is a contradiction that verifies the claim. Letting $\lambda_k := \bar{\lambda} > 0$ for all $k \in \mathbb{N}$ and assuming that the sequence is infinitely generated, we get $x_k < 0$. Applying again the stationary rule to (33) in this setting yields $\bar{\lambda} = x_{k+1} - x_k$, which leads us to a contradiction. Therefore, for any fixed $\bar{\lambda} > 0$, the sequence $\{x_k\}$ converges in finite steps if the starting point is selected as $x_0 < 0$.

We finalize this section with the following comments.

Remark 2 (comparison with known results).

- (i) The main novelty of Theorem 2 and Theorem 3 is for the case of lower exponents $q \in (0, 1/2)$, where the obtained results significantly improve the known ones with *linear convergence* for various algorithms satisfying (\mathcal{H}_1) and (\mathcal{H}_2) ; see [12, Theorem 3.4] and many other publications in this direction.²
- (ii) The convergence rate results obtained in Theorems 2, 3 for the cases where $q = 0$ and $q \in (1/2, 1)$ under the exponent PLK conditions are closely related to those established in [12, Theorem 3.4] for an implicit abstract model, which encompasses our explicit model defined by (\mathcal{H}_1) and (\mathcal{H}_2) . The proofs of Theorems 2, 3 are different from those given in [12].

²When the results of our paper have been completed and prepared for publication, we got familiar with preprint [25] containing an alternative proof of superlinear convergence of descent algorithms satisfying (\mathcal{H}_1) and (\mathcal{H}_2) under the PLK conditions with lower exponents.

4 Convergence Rates under Modified Error Property

In this section, we provide a comprehensive rate convergence analysis of the generic algorithms satisfying (\mathcal{H}_1) and (\mathcal{H}_3) under the exponent PLK conditions. It is striking to observe that, in contrast to the results of Section 3, the replacement of (\mathcal{H}_2) by (\mathcal{H}_3) fully *excludes infinite sequences of iterates* in the case of PLK with lower exponents. Moreover, we show below that, under some additional assumption, the lower exponent PLK property is *inconsistent* with the conditions in (\mathcal{H}_1) and (\mathcal{H}_3) .

Theorem 4 (convergence rates under exponent PLK and modified relative error conditions). *In the setting of Theorem 1, assume that the exponent PLK property of ϕ holds at \bar{x} with $\varphi(t) = Mt^{1-q}$ for some $M > 0$ and $q \in [0, 1)$. The following convergence rates are guaranteed for the generic iterative sequences satisfying the conditions in (\mathcal{H}_1) and (\mathcal{H}_3) :*

- (i) *If $q \in [0, \frac{1}{2})$, then $\{x^k\}$ and $\{\phi(x^k)\}$ are terminated in a finite number of steps to \bar{x} and $\bar{\phi} = \phi(\bar{x})$, respectively.*
- (ii) *If $q = \frac{1}{2}$, then the sequences $\{x^k\}$ and $\{\phi(x^k)\}$ converge linearly at \bar{x} and $\bar{\phi}$, respectively.*
- (iii) *If $q \in (1/2, 1)$, then there exists a positive constant σ such that*

$$\|x^k - \bar{x}\| \leq \sigma k^{-\frac{1-q}{2q-1}} \quad \text{for all large } k \in \mathbb{N}.$$

Proof. It is shown in the proof of Theorem 1 that $\phi(x^k) \rightarrow \phi(\bar{x})$ as $k \rightarrow \infty$. Defining $\psi(x) := \phi(x) - \phi(\bar{x})$, we see that the sequence $\{\psi(x^k)\} \subset [0, \infty)$ is decreasing, and hence $\psi(x^j) \geq t$ for all $t \in [\psi(x^{j+1}), \psi(x^j)]$, $j \in \mathbb{N}$. Thus

$$\frac{1}{\psi(x^j)^q} \leq \frac{1}{t^q} \quad \text{whenever } t \in [\psi(x^{j+1}), \psi(x^j)] \quad \text{and } q \in (0, 1),$$

which implies in turn that

$$\frac{\psi(x^j) - \psi(x^{j+1})}{\psi(x^j)^q} = \int_{\psi(x^{j+1})}^{\psi(x^j)} \frac{1}{\psi(x^j)^q} dt \leq \int_{\psi(x^{j+1})}^{\psi(x^j)} \frac{1}{t^q} dt.$$

Therefore, we arrive at the inequality

$$\frac{\psi(x^j) - \psi(x^{j+1})}{\psi(x^j)^q} \leq \frac{1}{1-q} (\psi(x^j)^{1-q} - \psi(x^{j+1})^{1-q}),$$

which gives us the estimate

$$\sum_{j=k}^{k+l} \frac{\psi(x^j) - \psi(x^{j+1})}{\psi(x^j)^q} \leq \frac{1}{1-q} (\psi(x^k)^{1-q} - \psi(x^{k+l+1})^{1-q}) \quad \text{for all } k, l \in \mathbb{N}. \quad (34)$$

On the other hand, combining (7), the construction of ψ , and (10) in (\mathcal{H}_3) yields

$$\psi(x^{k+1}) \leq \psi(x^k) - a\|x^{k+1} - x^k\|^2 \leq \psi(x^k) - \frac{a}{b^2}\|w^k\|^2.$$

By the PLK property (3) with $\varphi(t) = Mt^{1-q}$, we find $M > 0$ such that

$$\|w^k\| \geq \text{dist}(0, \partial\phi(x^k)) \geq \frac{1}{M(1-q)} (\phi(x^k) - \phi(\bar{x}))^q. \quad (35)$$

Using the latter together with (35) tells us that

$$\psi(x^{k+1}) \leq \psi(x^k) - \frac{a}{b^2 M^2 (1-q)^2} (\psi(x^k))^{2q}. \quad (36)$$

To furnish now the proof of assertion (i), consider first the case where $q \in (0, 1/2)$ and suppose, arguing by contradiction, that the sequence $\{x^k\}$ is infinitely generated, which tells us that $\psi(x^k) > 0$ for all $k \in \mathbb{N}$. Recall the convergence $\psi(x^k) \rightarrow 0$ as $k \rightarrow \infty$ and deduce from $2q - 1 < 0$, (36), and the equality

$$\psi(x^k) - \frac{a}{b^2 M^2 (1-q)^2} \psi(x^k)^{2q} = \psi(x^k) \left(1 - \frac{a}{b^2 M^2 (1-q)^2} \right) (\psi(x^k))^{2q-1}$$

that there exists a number $k_0 \in \mathbb{N}$ such that

$$\psi(x^{k_0+1}) \leq \psi(x^{k_0}) \left(1 - \frac{a}{b^2 M^2 (1-q)^2} (\psi(x^{k_0}))^{2q-1}\right) < 0,$$

which contradicts the fact that $\psi(x^k) > 0$ for all $k \in \mathbb{N}$. Thus for $q \in (0, 1/2)$, the sequence $\{x^k\}$ reaches the M -stationary point \bar{x} in a finite number of steps.

To verify (i), it remains to consider the case where $q = 0$. Arguing again by contradiction, suppose that the sequence $\{x^k\}$ is infinitely generated and deduce from (35) with $q = 0$ that $\|w^k\| \geq M^{-1}$ for all large k , which tells us together with (10) in (\mathcal{H}_3) that $(Mb)^{-1} \leq \|x^{k+1} - x^k\|$. This contradicts the fact that $\{x^k\}$ converges to \bar{x} and thus concludes the proof of assertion (i).

To verify assertion (ii), we get from (36) with $q = 1/2$ that

$$\psi(x^{k+1}) \leq \left(1 - \frac{4a}{b^2 M}\right) \psi(x^k). \quad (37)$$

Due to $\psi(x^k) \geq 0$ as $k \in \mathbb{N}$, it follows from (37) that $4a(b^2 M)^{-1} \in (0, 1)$, and hence $\{\phi(x^k)\}$ converges linearly to $\phi(\bar{x})$. On the other hand, we have by (7) and (10) that

$$\phi(x^{k+1}) \leq \phi(x^k) - ab^{-1} \|x^{k+1} - x^k\| \cdot \|w^k\| \quad \text{for all } k \in \mathbb{N},$$

which being combined with (35) and the construction of ψ yields

$$\psi(x^{k+1}) \leq \psi(x^k) - \frac{2a}{Mb} \|x^{k+1} - x^k\| (\psi(x^k))^{\frac{1}{2}}.$$

The obtained inequality readily ensures the estimate

$$\|x^{k+1} - x^k\| \leq \frac{Mb}{2a} \left(\frac{\psi(x^k) - \psi(x^{k+1})}{\psi(x^k)^{\frac{1}{2}}} \right),$$

which tells us together with (34) that

$$\begin{aligned} \sum_{j=k}^{k+l} \|x^{j+1} - x^j\| &\leq \frac{Mb}{2a} \sum_{j=k}^{k+l} \frac{\psi(x^j) - \psi(x^{j+1})}{\psi(x^j)^{\frac{1}{2}}} \\ &\leq \frac{Mb}{a} (\psi(x^k)^{\frac{1}{2}} - \psi(x^{k+l+1})^{\frac{1}{2}}). \end{aligned}$$

By taking the limit in the last inequality as $l \rightarrow \infty$, we arrive at

$$s_k := \sum_{j=k}^{\infty} \|x^{j+1} - x^j\| \leq \frac{Mb}{a} (\psi(x^k))^{\frac{1}{2}},$$

which being combined with (35) and (10) brings us to the estimates

$$s_{k+1} \leq \frac{Mb}{a} (\psi(x^k))^{\frac{1}{2}} \leq \frac{M^2 b}{2a} \|w^k\| \leq \frac{M^2 b^2}{2a} \|x^{k+1} - x^k\|.$$

Since $\|x^{k+1} - x^k\| = s_k - s_{k+1}$, the last inequality implies that

$$s_{k+1} \leq (2a)^{-1} M^2 b^2 (s_k - s_{k+1})$$

from which it clearly follows that

$$s_{k+1} \leq \frac{\tau}{1+\tau} s_k \quad \text{with } \tau := \frac{M^2 b^2}{2a}.$$

This tells us that $\{s_k\}$ converges linearly to 0 as $k \rightarrow \infty$. On the other hand, using the triangle inequality yields $\|x^k - x^{k+l}\| \leq \sum_{j=k}^{k+l} \|x^{j+1} - x^j\|$ from which we get that $\|x^k - \bar{x}\| \leq s_k$. Therefore, the linear convergence of $x^k \rightarrow \bar{x}$ follows from the linear convergence of $s_k \rightarrow 0$, and thus (ii) is fully justified. Assertion (iii) can be verified similarly to the proof of [5, Theorem 2] given in the case of the proximal algorithm. \square

The next significant result shows that the lower exponent PLK property of the cost function ϕ is *inconsistent* *fails* with $(\mathcal{H}1)$ and $(\mathcal{H}3)$ under the fulfillment of some additional assumption on the generic algorithm iterates. Note that another inconsistency result is established in Section 6 for a class of difference programming problems under the assumptions independent of $(\mathcal{H}1)$ and $(\mathcal{H}3)$.

Theorem 5 (inconsistency of lower exponent PLK with $(\mathcal{H}1)$ and $(\mathcal{H}3)$). *In addition to the general assumption of Theorem 4, suppose that $x^{k+1} \neq x^k$ for all $k \in \mathbb{N}$. Then the exponent PLK property of f at \bar{x} fails whenever $q \in (0, 1/2)$.*

Proof. Observe that if $x^k \neq x^{k+1}$ whenever $k \in \mathbb{N}$, then it follows from

$$a\|x^{k+1} - x^k\|^2 \leq \phi(x^k) - \phi(x^{k+1})$$

that $\phi(x^k) \neq \phi(x^{k+1})$ for all $k \in \mathbb{N}$. Arguing by contradiction, suppose that ϕ satisfies the PLK property at \bar{x} with exponent $q \in (0, 1/2)$. It follows from $(\mathcal{H}1)$ and $(\mathcal{H}3)$ that

$$\phi(x^k) - \phi(x^{k+1}) \geq \frac{a}{b^2} \|w^k\|^2.$$

Combining the last inequality with the the exponent PLK condition yields

$$\phi(x^k) - \phi(x^{k+1}) \geq \frac{a}{(1-q)^2 M^2 b^2} (\phi(x^k) - \phi(\bar{x}))^{2q}.$$

Using the latter together with $\phi(x^k) - \phi(\bar{x}) \geq \phi(x^k) - \phi(x^{k+1})$ brings us to

$$(\phi(x^k) - \phi(\bar{x}))^{1-2q} \geq \beta, \quad \text{where } \beta := \frac{a}{(1-q)^2 M^2 b^2}. \quad (38)$$

Since $\phi(x^k) - \phi(\bar{x})$ as $k \rightarrow \infty$, (38) fails for large k when $q \in (0, 1/2)$. The obtained contradiction verifies the claimed assertion and completes the proof of theorem. \square

Remark 3 (novelty under the modified elative error condition).

- (i) Similarly to the case of the relative error condition (\mathcal{H}_2) discussed in Remark 2, the convergence rate results obtained in Theorem 4 under the modified version $(\mathcal{H}3)$ and the exponent PLK properties with $q = 0$ and $q \in [1/2, 1)$ continue to be generic extensions of those given for particular algorithms (see, e.g., [2, 3, 4, 5, 12, 16, 14, 21]).
- (ii) In the case of *continuous-time* gradient/subgradient systems, a *finite-time* convergence estimate has already appeared in the literature when the exponent PLK varies within the interval $(0, 1/2)$; see [8, Theorem 2.7]. Inspired by the latter result, it is shown in [12, Theorem 3.5(i)] that for certain algorithms including gradient-related methods, this better estimate remains true at least when dealing with the sequence of functional values. Our approach to this issue is novel being also in accordance with the estimates presented in the continuous case.
- (iii) To the best of our knowledge, the *inconsistency* result in the setting of Theorem 5 has never been observed before. Its analogs in the different setting are discussed in Section 6; see Theorem 7 and the discussion after its proof.

5 Specifications of Generic Descent Algorithms

This section discusses some specifications of the general class of descent methods and convergence analysis for them developed in Sections 2 and Section 4.

We consider here two particular methods and start with the following model.

5.1 Inexact Reduced Graduate Methods

A class of *inexact reduced graduate (IRG) methods* with various stepsize selections have been recently proposed and developed by Khanh et al. [14] for problems of smooth nonconvex optimization. The model unifying the IRG methods in [14] considers the objective function $\phi: \mathbb{R}^n \rightarrow \mathbb{R}$ and the iterative sequence $\{x^k\}$ as $k \in \mathbb{N}$ satisfying

$$\phi(x^k) - \phi(x^{k+1}) \geq \frac{\beta}{t_k} \|x^{k+1} - x^k\|^2 \quad \text{and} \quad \|\nabla \phi(x^k)\| \leq \frac{c}{t_k} \|x^{k+1} - x^k\|, \quad (39)$$

where $t_k \subset \mathbb{R}_+$ and $\beta, c > 0$. This can be viewed as a smooth version of the general nonsmooth scheme investigated above. The convergence analysis of IRG algorithms conducted in [14] is based on the observation that the sequences of iterates generated by (39) satisfy the following properties:

- *Primary descent condition*: There exists $\sigma > 0$ such that

$$\phi(x^k) - \phi(x^{k+1}) \geq \sigma \|\nabla \phi(x^k)\| \cdot \|x^{k+1} - x^k\| \quad \text{for large } k \in \mathbb{N}.$$

- *Complementary descent condition*: we have the implication

$$[\phi(x^{k+1}) = \phi(x^k)] \implies [x^{k+1} = x^k] \quad \text{for large } k \in \mathbb{N}.$$

In fact, the latter conditions, while not (39), are used by Absil et al. in [1, page 536] to prove the convergence of the corresponding iterative sequences when the cost function is real analytical. Let us emphasize that the general scheme (39) comprises *new linesearch methods* with an *inexact gradient information* for finding stationary points of \mathcal{C}^1 -smooth functions by using different choices of stepsize rules. To the best of our knowledge, [14] is a first paper dealing with convergence rates for methods satisfying condition (39). In particular, it extends some special settings considered, e.g., in [23] and [13], where the authors analyze convergence properties of the exact gradient descent method for functions satisfying the PLK conditions. More precisely, [14, Theorem 2.5] extends the scope of applicability of the results in the aforementioned papers. The results established above allow us to essentially improve those obtained in [14] for IRG methods and their various linesearch specifications. In particular, we now have the *finite termination* of such algorithms under (39) and the exponent PLK with $q \in [0, 1/2)$ at any *accumulation point* of the iterative sequences *provided* that the corresponding stepsize sequence $\{t_k\}$ is *bounded away* from zero. The latter assumption is required to accommodate (H3) with a fixed $b > 0$ in (10).

5.2 Boosted DC Algorithm

Let us recall the following Boosted DC Algorithm with Backtracking (BDCA), which was proposed and investigated by Aragón-Artacho and Vuong [4] to solve the problem:

$$(\mathcal{P}) \quad \min_{x \in \mathbb{R}^n} \phi(x) := g(x) - h(x),$$

where both $g, h: \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ are assumed to be strongly convex with modulus $\rho > 0$, with g being strictly differentiable at the points in question:

Algorithm 1 BDCA

1. Fix $\alpha > 0, \bar{\lambda} > 0$ and $\beta \in (0, 1)$. Let x_0 be any initial point and set $k := 0$.
2. Select $u_k \in \partial h(x^k)$ and solve the strongly convex optimization problem

$$(\mathcal{P}'_k) \quad \min_{x \in \mathbb{R}^n} g(x) - \langle u_k, x \rangle$$

to obtain its unique solution y^k .

3. Let $d^k := y^k - x^k$. If $d^k = 0$, **stop** and **return** x^k . Otherwise, go to Step 4.
 4. Choose any $\bar{\lambda}_k \geq 0$. Set $\lambda_k := \bar{\lambda}_k$. **While** $\phi(y^k + \lambda_k d^k) > \phi(y^k) - \alpha \lambda_k^2 \|d^k\|^2$, **do** $\lambda_k := \beta \lambda_k$.
 5. Let $x^{k+1} := y^k + \lambda_k d^k$. If $x^{k+1} = x^k$, **stop** and **return** x^k . Otherwise, set $k := k + 1$ and go to Step 2.
-

When h is also smooth, BDCA was introduced and analyzed in Aragón-Artacho et al. [2]. In both papers, BDCA accelerates the convergence of the classical Difference of Convex Functions Algorithm (DCA). It is shown in [4] that the *strong PLK condition* imposed on ϕ at an accumulation point yields the global convergence of iterates with deriving convergence rates under the exponent strong PLK condition with $\varphi(t) = Mt^{1-q}$ for $q \in [0, 1)$ and $M > 0$. Moreover, it is observed in [4, Remark 4.4], with the reference to the second author of the current paper, that for Lipschitz continuous functions ϕ , the strong PLK property may be replaced by the symmetric one.

Next we show that BDCA can be viewed as a specification, for non-Lipschitzian *continuous* functions ϕ , of the generic algorithm developed in Section 2 with using the *symmetric subdifferential* (5) and the *symmetric PLK property* of ϕ . This approach allows us to improve and extend the results of [4] about the convergence and convergence rates of BDCA to the case of continuous functions. First we recall some properties of the symmetric subdifferential of continuous functions used below; all of them and much more on (5) (including full calculus) can be found in [19, 20].

(\mathcal{A}_1) $\partial^0(-\phi(x)) = -\partial^0\phi(x)$ (plus-minus symmetry).

(\mathcal{A}_2) $\partial^0\phi(x) = \{\nabla\phi(x)\}$ when ϕ is strictly differentiable.

(\mathcal{A}_3) We have the representation $\partial^0\phi(\bar{x}) = \partial\phi(\bar{x}) = \{x^* \in \mathbb{R}^n \mid \langle x^*, x - \bar{x} \rangle \leq \phi(x) - \phi(\bar{x}) \text{ for all } x \in \mathbb{R}^n\}$ when ϕ is convex.

(\mathcal{A}_4) If $\phi = f_1 + f_2$ with f_1 being strictly differentiable, then we have the sum rule $\partial^0\phi(x) = \nabla f_1(x) + \partial^0 f_2(x)$.

Observe that $\partial^0\psi(x)$ is often nonconvex being significantly smaller than Clarke's subdifferential $\bar{\partial}\psi(x)$, with the relationship $\bar{\partial}\psi(x) = \text{co } \partial^0\phi(x)$ when ϕ is locally Lipschitzian around x . Note also that Clarke's counterpart of the plus-minus symmetry property as in (\mathcal{A}_1) requires the local Lipschitz continuity of ϕ around \bar{x} .

The theorem below incorporates BDCA into the generic class of algorithms satisfying conditions ($\mathcal{H}1$) and ($\mathcal{H}3$) via the symmetric subdifferential and provides its convergence analysis under the symmetric exponent PLK properties.

Theorem 6 (BDCA for continuous functions). *Consider problem (P), where ϕ is continuous on its domain. Let the sequence $\{x^k\}$ be generated by BDCA, and let ∇g be Lipschitz continuous with modulus L around an accumulation point \bar{x} of $\{x^k\}$. Then for all large $k \in \mathbb{N}$ we have the conditions:*

$$\phi(x^{k+1}) \leq \phi(x^k) - \frac{\alpha\lambda_k^2 + \rho}{(1 + \lambda_k)^2} \|x^{k+1} - x^k\|^2, \quad (40)$$

$$\text{there exists } w^k \in \partial^0\phi(x^k) \text{ such that } \|w^k\| \leq L\|x^{k+1} - x^k\|, \quad (41)$$

and $\{x^k\}$ converges to \bar{x} as $k \rightarrow \infty$, which is a symmetric stationary point of (P) satisfying $0 \in \partial^0\phi(\bar{x}) = \nabla g(\bar{x}) - \partial h(\bar{x})$. If furthermore ϕ enjoys the exponent PLK property with $\varphi(t) = Mt^{1-q}$ for some $M > 0$ and $q \in [0, 1)$, then the convergence rates of $x^k \rightarrow \bar{x}$ are as given in Theorem 4.

Proof. Using the definitions of x^{k+1} and d^k in BDCA gives us the equalities

$$\|x^{k+1} - x^k\| = \|y^k + \lambda_k d^k - x^k\| = \|y^k + \lambda_k(y^k - x^k) - x^k\| = (1 + \lambda_k)\|d^k\|$$

from which we deduce the estimate

$$\|d^k\| = \frac{1}{1 + \lambda_k} \|x^{k+1} - x^k\| \leq \|x^{k+1} - x^k\|. \quad (42)$$

Hence the estimate in (40) follows from the conditions

$$\phi(y^k) \leq \phi(x^k) - \rho\|d^k\|^2 \text{ and } \phi(y^k + \lambda_k d^k) \leq \phi(y^k) - \alpha\lambda_k^2\|d^k\|^2 \text{ for some } \delta_k > 0$$

obtained for BDCA in [4, Proposition 3.1]. To verify further (41), we employ properties (\mathcal{A}_1)–(\mathcal{A}_4) of the symmetric subdifferential and get the relationships

$$\nabla g(y_k) - \nabla g(x_k) \in \partial h(x_k) - \nabla g(x_k) = \partial^0(-\phi(x_k)) = -\partial^0\phi(x_k).$$

Recalling that y^k is the unique solution to the subproblem (\mathcal{P}'_k) in BDCA, i.e., $u^k = \nabla g(y^k)$ ensures that $w^k := \nabla g(x^k) - \nabla g(y^k) \in \partial^0\phi(x^k)$. Take now a neighborhood U of \bar{x} on which ∇g is Lipschitz continuous with modulus L . It follows from (42) that $x^k, y^k \in U$ for all large $k \in \mathbb{N}$, and hence

$$\|\nabla g(x^k) - \nabla g(y^k)\| \leq L\|x^k - y^k\|,$$

which yields (41). Since \bar{x} is an accumulation point of $\{x^k\}$ and ϕ is continuous on its domain, we deduce from (40) and (41) by the proof of Theorem 1 that $x_k \rightarrow \bar{x}$ as $k \rightarrow \infty$ with $0 \in \partial^0\phi(\bar{x})$. The latter implies by (\mathcal{A}_1)–(\mathcal{A}_4) that $0 \in \nabla g(\bar{x}) - \partial h(\bar{x})$. The convergence rates under the exponent PLK conditions follows from Theorem 4. \square

Note that the finite termination of BDCA at any accumulation point under the exponent PLK condition with $q \in (0, 1/2)$ is a significant improvement of [4] even for Lipschitz continuous objectives, where merely linear convergence of $\{x^k\}$ is justified. We'll see in Section 6 the further clarification of this fact at local minimizers of ϕ .

6 Lower Exponent PLK for Difference Programs

In this section, we take a close look at the exponent PLK property from Definition 1(iv) with *lower exponents* $q \in (0, 1/2)$ for a class of *difference programs*.

As shown in [9, Theorem 1]³, the classical PL gradient inequality always holds for any *analytic* function $\phi: \mathbb{R}^n \rightarrow \mathbb{R}$ with some exponent $q \in [1/2, 1)$. Although it does not exclude the fulfillment of the PLK conditions with lower exponents, we show below that this property *fails* for *local minimizers* of smooth functions with *Lipschitzian gradients*, i.e., of class $\mathcal{C}^{1,1}$, without any connections with decent algorithms satisfying (\mathcal{H}_1) and (\mathcal{H}_3) as in Theorem 5. In fact, our analysis provides such a result for problems of *difference programming* of type (\mathcal{P}) (see Subsection 5.2) with $\phi = g - h$, where g is of class $\mathcal{C}^{1,1}$ (may not be convex), while $h: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ is l.s.c. and convex. Recall that for any $x \in \text{int}(\text{dom } h)$, the directional derivative of ϕ is represented by

$$\phi'(x, d) = \langle \nabla g(x), d \rangle - \max_{v \in \partial h(x)} \langle v, d \rangle \quad \text{whenever } d \in \mathbb{R}^n. \quad (43)$$

Let us introduce the two important *measures of stationarity* at the point $x \in \text{int}(\text{dom } h)$ as follows. Observe first from (43) that

$$\begin{aligned} \min_{\|d\|=1} \phi'(x, d) &= \min_{\|d\|=1} \left\{ \langle \nabla g(x), d \rangle - \max_{v \in \partial h(x)} \langle v, d \rangle \right\} = \min_{\|d\|=1} \min_{v \in \partial h(x)} \langle \nabla g(x) - v, d \rangle \\ &= \min_{v \in \partial h(x)} \left(\{-\|\nabla g(x) - v\|\} \right) = - \max_{v \in \partial h(x)} \|\nabla g(x) - v\| := -\sigma_1(x). \end{aligned}$$

Thus $\sigma_1(x)$ measures the *fastest local decrease* of the function ϕ . For each $x \in \mathbb{R}^n$, fix

$$v_1(x) \in \arg\max_{v \in \partial h(x)} \|\nabla g(x) - v\|$$

and note that this point satisfies the necessary condition for local maximum:

$$\langle v_1(x) - \nabla g(x), v - v_1(x) \rangle \leq 0 \quad \text{whenever } v \in \partial h(x).$$

In other words, we have the equality

$$\langle \nabla g(x) - v_1(x), v_1(x) \rangle = \min_{v \in \partial h(x)} \langle \nabla g(x) - v_1(x), v \rangle. \quad (44)$$

The second measure of stationarity at x is defined from the conditions

$$\begin{aligned} \max_{\|d\|=1} \phi'(x, d) &= \max_{\|d\|=1} \left\{ \langle \nabla g(x), d \rangle - \max_{v \in \partial h(x)} \langle v, d \rangle \right\} = \max_{\|d\| \leq 1} \min_{v \in \partial h(x)} \langle \nabla g(x) - v, d \rangle \\ &= \min_{v \in \partial h(x)} \max_{\|d\| \leq 1} \langle \nabla g(x) - v, d \rangle = \min_{v \in \partial h(x)} \|\nabla g(x) - v\| := \sigma_0(x). \end{aligned}$$

Thus $\sigma_0(x)$ measures the *fastest local increase* of ϕ . We obviously get

$$\sigma_0(x) \leq \sigma_1(x) \quad \text{whenever } x \in \mathbb{R}^n.$$

Given an arbitrary point $x_0 \in \mathbb{R}^n$, consider its *own level set* $\mathcal{L}(x_0)$, which is the maximal connected component in $\{x \mid \phi(x) \leq \phi(x_0)\}$ containing x_0 . Let us define the *exact Lipschitzian bound* of ∇g with respect to $\mathcal{L}(x_0)$ by

$$L(x_0) = \inf \{L \mid \|\nabla g(x) - \nabla g(y)\| \leq L\|x - y\| \text{ over } x, y \in \mathcal{L}(x_0)\}. \quad (45)$$

The next lemma measures the difference between $\phi(x_0)$ and the global minimum of ϕ with respect to the level set $\mathcal{L}(x_0)$ in terms of the exact Lipschitzian bound (45) and the stationarity measure $\sigma_1(x_0)$.

Lemma 3 (distance estimate from global minimum). *Let $\bar{x} \in \text{int}(\text{dom } h)$ be a global minimizer of the function ϕ in (\mathcal{P}) with respect to the level set $\mathcal{L}(x_0)$, i.e.,*

$$\phi(x) \geq \phi(\bar{x}) \quad \text{for all } x \in \mathcal{L}(x_0).$$

If $L(x_0) < \infty$, then for $d(x_0) := \nabla g(x_0) - v_1(x_0)$ and any $\alpha \in [0, 1/L(x_0)]$, we have

$$\phi(x_0) - \phi(\bar{x}) \geq \phi(x_0) - \phi(x_0 - \alpha d(x_0)) \geq \frac{1}{2} \alpha (2 - \alpha L(x_0)) \sigma_1^2(x_0). \quad (46)$$

³The authors are grateful to Guoyin Li for drawing our attention to paper [9].

Proof. Since the case where $\sigma_1(x_0) = 0$ is obvious, suppose that $\sigma_1(x_0) > 0$. Denote $x_\alpha := x_0 - \alpha d(x_0)$ for $\alpha \geq 0$ and define the function $\xi(\alpha) := \phi(x_\alpha)$. Note that

$$\begin{aligned}\xi'(0, 1) &= \phi'(x_0, -d(x_0)) = \langle \nabla g(x_0), -d(x_0) \rangle - \max_{v \in \partial h(x_0)} \langle v, -d(x_0) \rangle \\ &= \langle \nabla g(x_0), -d(x_0) \rangle + \min_{v \in \partial h(x_0)} \langle v, d(x_0) \rangle \stackrel{(44)}{=} -\sigma_1(x_0) < 0.\end{aligned}$$

Let $\bar{\alpha}$ be the smallest element of the set $\Sigma = \{\alpha \geq 0 \mid \xi'(\alpha, 1) \geq 0\}$. If $\Sigma = \emptyset$, then $\bar{\alpha} = \infty$. Since $\xi'(0, 1) < 0$, we have $\xi'(\alpha, 1) < 0$ for all $\alpha \in [0, \bar{\alpha})$. Thus any x_α with $\alpha \in [0, \bar{\alpha}]$ belongs to the set $\mathcal{L}(x_0)$. Adding the fact that ∂h is monotone, this yields

$$\begin{aligned}\xi'(\alpha, 1) &= \langle \nabla g(x_\alpha), -d(x_0) \rangle - \max_{v \in \partial h(x_\alpha)} \langle v, -d(x_0) \rangle \\ &\leq \langle \nabla g(x_\alpha), -d(x_0) \rangle - \max_{v \in \partial h(x_0)} \langle v, -d(x_0) \rangle \\ &\stackrel{(44)}{=} \langle \nabla g(x_\alpha), -d(x_0) \rangle + \langle v_1(x_0), d(x_0) \rangle \\ &\leq (\alpha L(x_0) - 1) \sigma_1^2(x_0).\end{aligned}$$

Hence $\bar{\alpha} \geq \hat{\alpha} := 1/L(x_0)$, and therefore for all $\alpha \in [0, \hat{\alpha}]$ we get

$$\xi(\alpha) = \xi(0) + \int_0^\alpha \xi'(t, 1) dt \leq \xi(0) + \int_0^\alpha (tL(x_0) - 1) \sigma_1^2(x_0) dt = \xi(0) - \frac{\alpha(2 - \alpha L(x_0))}{2} \sigma_1^2(x_0).$$

By $x_{\hat{\alpha}} \in \mathcal{L}(x_0)$, the latter means that

$$\phi(x_0) - \phi(\bar{x}) \geq \phi(x_0) - \phi(x_{\hat{\alpha}}) \geq \frac{\sigma_1^2(x_0)}{2} 2L(x_0),$$

which verifies (46) and thus completes the proof of the lemma. \square

Now we are ready to establish our major observation about the lower exponent PLK property for problems of difference programming (\mathcal{P}) , with $\mathcal{C}^{1,1}$ functions g and convex functions h , at local minimizers of ϕ .

Theorem 7 (inconsistency of the lower exponent PLK property with Lipschitz continuity of gradients). *Let $\bar{x} \in \text{int}(\text{dom } h)$ be a local minimizer of problem (\mathcal{P}) , where g is of class $\mathcal{C}^{1,1}$ around \bar{x} , and where h is convex. Then the exponent PLK property of ϕ at \bar{x} fails whenever $q \in (0, 1/2)$.*

Proof. Observe that the exponent PLK property of ϕ at \bar{x} can be represented, provided that ϕ is locally continuous around \bar{x} as follows from the assumptions of the theorem, in the following form: there exist a constant $M > 0$ and a neighborhood U of \bar{x} such that for all $x \in U$ taken inside of $\mathcal{L}(x_0)$, we have the inequality

$$\sigma(x) \geq M[\phi(x) - \phi(\bar{x})]^q, \tag{47}$$

where $\sigma(x)$ is some estimate for the norm of the corresponding subgradient from (3) of ϕ at x . The standard class $\sigma(x)$ is $\sigma(x) = \sigma_0(x)$, but we consider even a broader class where $\sigma(x) = \sigma_1(x)$ in (47). Then Lemma 3 tells us that for any $x_0 \in U$, the estimates

$$M \left[\frac{1}{2L(x_0)} \sigma_1^2(x_0) \right]^q \leq M [\phi(x_0) - \phi(\bar{x})]^q \stackrel{(47)}{\leq} \sigma_1(x_0)$$

are satisfied by definition (45). This implies that

$$M \sigma_1^{2q-1}(x_0) \leq [2L(x_0)]^q. \tag{48}$$

It follows from (48) for $q \in (0, \frac{1}{2})$ that $L(x_0) \rightarrow \infty$ as $x_0 \rightarrow \bar{x}$, which is inconsistent with the Lipschitz continuity of ∇g around \bar{x} and thus completes the proof. \square

As we see, the inconsistency result of Theorem 7 affects also problems of *unconstrained minimization of $\mathcal{C}^{1,1}$ functions* at their local minimizers. After a preliminary version of this paper was uploaded to arXiv, we received a message from Nicolas Boumal informing us that about the results and discussions on the failure of the lower exponent PLK conditions for problems of $\mathcal{C}^{1,1}$ -optimization given in [26, Remark 2.21].

The following example shows that the inconsistency observation of Theorem 7 may *fail* if g is *not of class $\mathcal{C}^{1,1}$* . The function below is taken from [16, Example 4.6], where it is used to demonstrate that the tight *quadratic* convergence rate can be achieved for the proximal methods applied to this function.

Example 3 ($\mathcal{C}^{1,1}$ property is essential for inconsistency with lower exponent PLK). Consider the univariate function $\phi(x) := |x|^{3/2}$, which has the global minimizer $\bar{x} = 0$. It is straightforward to compute that

$$\phi'(x) = \frac{3}{2} \text{sign}(x)|x|^{\frac{1}{2}} = \begin{cases} \frac{3}{2}x^{\frac{1}{2}} & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -\frac{3}{2}(-x)^{\frac{1}{2}} & \text{if } x < 0. \end{cases}$$

Therefore, the derivative of ϕ is not locally Lipschitzian around \bar{x} . On the other hand, it is not difficult to check that the lower exponent PLK property holds for this function with $\varphi(t) = t^{1-q}$ and $q = 1/3$.

7 Conclusions

This paper conducts a comprehensive convergence analysis of generic descent algorithms under several versions of the PLK conditions. Convergence rates are established by using the exponent PLK conditions, where some really surprising results are obtained for the case of lower exponents. It is shown, in particular, that the lower exponent PLK conditions under consideration are inconsistent with the Lipschitz continuity of the gradient ∇g for problems of minimizing the difference functions $g - h$, where h is extended-real-valued and convex.

There are several challenging open questions for the future research. Among such questions, the following should be mentioned. First, we intend to clarify for which particular algorithms the additional assumption ensuring the striking result of Theorem 5 holds. Another issue is to investigate difference programs discussed in Section 5.2 and Section 7 in the case where h is nonconvex with some specific structures, e.g., being *prox-regular* [21, 27]. Among other algorithms, we plan to study from this viewpoint are the semi-Newton regularized method of [3] in difference programming as well as the cubic Newton method by Nesterov and Polyak [22] and its far-going extension in [16]. Let us finally mention our intention to consider DC programs where the function g in the DC decomposition $\phi = g - h$ is *nondifferentiable*. Such problems exhibit significant differences from the differentiable case; see the recent works [10, 11] for new phenomena in *nonexact* versions of Boosted DCA. Our goal is to establish appropriate counterparts of the results obtained in this paper for difference problems without any differentiability assumptions.

Acknowledgements

Research of Boris S. Mordukhovich was supported by the US National Science Foundation under grant DMS-2204519 and by the Australian Research Council under Discovery Project DP250101112. Research of Glaydston de C. Bento was supported by CNPq grants 314106/2020-0. Research of Yurii Nesterov was supported by the National Research, Development and Innovation Office (NKFIH) under grant number 2024-1.2.3-HU-RIZONT-2024-00030.

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] F. J. Aragón-Artacho, R. M. Fleming and P. T. Vuong, Accelerating the DC algorithm for smooth functions, *Math. Program.* **169** (2018), 95–118.
- [3] F. J. Aragón-Artacho, B. S. Mordukhovich and P. Pérez-Aros, Coderivative-based semi-Newton method in nonsmooth difference programming, *Math. Program.*, DOI:10.1007/s10107-024-02142-8 (2024).
- [4] F. J. Aragón-Artacho and P. T. Vuong, The boosted difference of convex functions algorithm for nonsmooth functions, *SIAM J. Optim.* **30** (2020), 980–1006.
- [5] H. Attouch and J. Bolté, On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, *Math. Program.* **116** (2009), 5–16.
- [6] H. Attouch, J. Bolté and B. F. Svaiter, Convergence of descent methods for semialgebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, *Math. Program.* **137** (2013), 91–129.

- [7] J. Bolté, A. Daniilidis, A. Lewis and M. Shiota, Clarke subgradients of stratifiable functions, *SIAM J. Optim.* **18** (2007), 556–572.
- [8] R. Chill and A. Fiorenza, Convergence and decay rate to equilibrium of bounded solutions of quasilinear parabolic equations, *J. Diff. Eqs.* **228** (2006), 611–632.
- [9] P. M. M. Feehan, Resolution of singularities and geometric proofs of the Łojasiewicz inequalities, *Geom. Topol.* **23** (2019), 3273–3313.
- [10] O. P. Ferreira, B. S. Mordukhovich, W. M. S. Santos and J. C. O. Souza, An inexact boosted difference of convex algorithm for nondifferentiable functions, arXiv:2412.05697 (2024).
- [11] O. P. Ferreira, E. M Santos and J. C. O. Souza, A boosted DC algorithm for nondifferentiable DC components with nonmonotone linesearch, *Comput. Optim. Appl.* **88** (2024), 783–818.
- [12] P. Frankel, G. Garrigos and J. Peypouquet, Splitting methods with variable metric for Kurdyka–Łojasiewicz functions and general convergence rates, *J. Optim. Theory Appl.* **165** (2015), 874–900.
- [13] H. Karimi, J. Nutini and M. Schmidt, Linear convergence of gradient and proximal-gradient methods under the Polyak–Łojasiewicz condition, in *Machine Learning and Knowledge Discovery in Databases, Part 1*, pp. 795–811, Springer, Cham, Switzerland, 2016,
- [14] P. D. Khanh, B. S. Mordukhovich and D. B. Tran, Inexact reduced gradient methods in nonconvex optimization, *J. Optim. Theory Appl.* **203** (2024), 2138–2178.
- [15] K. Kurdyka, On gradients of functions definable in o-minimal structures, *Ann. Inst. Fourier* **48** (1998), 769–783.
- [16] G. Li, B. S. Mordukhovich and J. Zhu, Generalized metric subregularity with applications to high-order regularized Newton methods, arXiv:2406.13207 (2024).
- [17] G. Li and T. K. Pong, Calculus of the exponent of Kurdyka–Łojasiewicz inequality and its applications to linear convergence of first-order methods, *Found. Comput. Math.* **18** (2018), 1199–1232.
- [18] S. Łojasiewicz, Une propriété topologique des sous-ensembles analytiques réels. Coll. du CNRS, Les équations aux dérivées partielles, pp. 87–89, 1963.
- [19] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation I: Basic Theory*, Springer, Berlin, 2006.
- [20] B. S. Mordukhovich, *Variational Analysis and Applications*, Springer, Cham, Switzerland, 2018.
- [21] B. S. Mordukhovich, *Second-Order Variational Analysis in Optimization, Variational Stability, and Control: Theory, Algorithms, Applications*, Springer, Cham, Switzerland, 2024.
- [22] Yu. Nesterov and B. T. Polyak, Cubic regularization of Newton method and its global performance, *Math. Program.* **108** (2006), 177–205.
- [23] B. T. Polyak, Gradient methods for the minimization of functionals, *USSR Comput. Math. Math. Phys.* **3** (1963), 864–878.
- [24] B. T. Polyak, *Introduction to Optimization*, Optimization Software, New York, 1987.
- [25] Y. Qian and S. Pan, A superlinear convergence framework for Kurdyka–Łojasiewicz optimization, arXiv:2210.12449 (2023).
- [26] Q. Rebjock and N. Boumal, Fast convergence to non-isolated minima: four equivalent conditions for C^2 functions, *Math. Program.*, DOI: 10.1007/s10107-024-02136-6 (2024).
- [27] R. T. Rockafellar and R. J-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.