Journal of Optimization Theory and Applications (2026) 208:56
https://doi.org/10.1007/s10957-025-02877-0

Equivalence of the Polyak-Łojasiewicz-Kurdyka Exponent Via Difference-of-Moreau-Envelope Smoothing

Shuqin Sun[1,2,3]

Received: 8 July 2024 / Accepted: 26 October 2025

Abstract
For unconstrained nonsmooth difference-of-convex (DC) optimization problems, the difference-of-Moreau-envelope (DME) smoothing serves as a significant smooth approximation for them. Maintaining DC structure, the resulted DME-based model has a one-to-one correspondence for the stationary points with the original DC problem. This has led to the development of DME-specific algorithms to indirectly solve the DC problems by solving their DME-based models. In this paper, we obtain the global convergence and the specific local convergence rate of various DME-specific algorithms to find the stationary points of the corresponding DC problems. These results are based on the Polyak-Łojasiewicz-Kurdyka (PLK) property and the specific PLK exponent assumed on the DME-based model or the potential function designed in the DME-specific algorithm. More importantly, we establish the equivalence of the PLK exponent between the DC problems and their DME-based models. Combined with our local convergence rate result, we are allowed to show the linear and sublinear convergence rates of these specific algorithms. Moreover, the equivalence result also provides a new tool to explore the PLK exponent of the DC problem from that of its DME-based model. An example is then provided to show that the PLK exponent of a nonconvex compressed sensing model that incorporates a logistic penalty term is 1/2, which was previously unknown.

Keywords Difference-of-Moreau-Envelope smoothing · Polyak-Łojasiewicz-Kurdyka exponent · Difference-of-Convex problems · Inexact gradient descent algorithm · Convergence rate

---

Communicated by Francisco J. Aragón Artacho.

✉ Shuqin Sun
sunshuqinsusan@163.com

1 School of Mathematical Sciences, China West Normal University, Nanchong, Sichuan, China

2 Visual Computing and Virtual Reality Key Laboratory of Sichuan Province, Sichuan Normal University, Chengdu, Sichuan, China

3 Key Laboratory of Optimization Theory and Applications at China West Normal University of Sichuan Province, China West Normal University, Nanchong, Sichuan, China

Published online: 17 November 2025

Mathematics Subject Classification 90C25 · 90C26 · 90C90 · 68Q25

1 Introduction

The unconstrained nonsmooth DC problem constitutes an important class of problems in structured nonconvex and nonsmooth optimization, which arises naturally in many contemporary applications, such as digital communication system [1], assignment and power allocation [37] and compressed sensing [38, 40, 42, 43]; more applications can be found in [19]. Interested readers may refer to [20, 21] and the references therein for an overview of the DC programming. The general DC problem admits the form as

$$\min_{x\in\mathbb{R}^n} F(x):=f(x)-g(x), \tag{1}$$

where $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ and $g:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ are proper closed convex functions (take $+\infty-(+\infty)=+\infty$ by convention as in [40]). The celebrated DC algorithm for solving (1) was first proposed in [31]. Since then, many variants of DC algorithms [16, 24, 25, 42] and their global convergence, which refers to the convergence of the entire sequence of iterates to a stationary point of (1), have been developed extensively. Additionally, the convergence rate has been established under the assumption that g has Lipschitz gradient and the objective or potential function has the PLK exponent ranging from $[0, 1)$; see, e.g., [42, Theorem 4.3] and [24, Theorem 4.2]. However, it is worth noting that reformulating a DC problem to incorporate a Lipschitz differentiable g is not always achievable, such as the commonly used models in compressed sensing: Capped ℓ_1 regularization model [15], ℓ_{1-2} regularization model [43] and Truncated ℓ_1 regularization model [41]. To circumvent the strict assumption on g, one way is to apply the difference-of-Moreau-envelope (DME) smoothing [17] to (1) and consider the resulted DME-based model:

$$\min_{x\in\mathbb{R}^n} F_\lambda(x):=e_\lambda f(x)-e_\lambda g(x); \tag{2}$$

see, e.g., [38, 40], where $e_\lambda f$ and $e_\lambda g$ are Moreau envelopes (defined as in (7)) of f and g with $\lambda>0$, respectively. Compared to the original DC problem (1), this DME-based model possesses several nice properties while preserving the DC structure: (i) $F_\lambda:\mathbb{R}^n\to\mathbb{R}$ is Lipschitz differentiable, with $1/\lambda$ as its Lipschitz modulus; (ii) the stationary point of (1) can be obtained by applying the proximal operator of f or g to that of (2), as shown in [40, Lemma 4]. These advantageous properties suggest that developing DME-specific algorithms is promising.[1] These algorithms aim to minimize F_λ instead of F by using (inexact) gradient of F_λ which has an explicit formula that involves the computation of proximal operators with a parameter $\lambda>0$.

Several DME-specific algorithms have been proposed and achieved promising numerical performance, such as the inexact gradient descent (**IGD**) algorithm in [38, Algorithm 1] and the classical gradient descent (**GD**) algorithm in [40, Algorithm 1]. The convergence analysis for the sequence of gradients or subgradients at the iterates

[1] Although the computation of some proximal operators may be costly in general, the proximal operators of many applications from compressed sensing admit a closed-form; like the ℓ_{1-2} regularization model and the Capped ℓ_l regularization model mentioned above.

generated by these two DME-specific algorithms has been well-established in [38, 40]. However, the convergence of the whole iterates generated by some algorithms of this kind and the corresponding local convergence rate have been less studied. To do these, one typically resorts to the Polyak-Łojasiewicz-Kurdyka (PLK) properties. Indeed, the PLK property is an important tool for analyzing the convergence properties of many first-order optimization algorithms. In particular, the PLK exponent is commonly used to analyze the corresponding local convergence rate; for more details, refer to [4–6, 22, 23]. All of these PLK dependencies are usually assumed on the objective function or the potential function; see, e.g., [4, 24, 42]. In this paper, we demonstrate the global convergence and the specific local convergence rate of **GD** under the PLK assumption of F_λ and those of **IGD** under the PLK assumption of potential function $H_\lambda : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ (designed for the convergence analysis of **IGD** in [38]):

$$H_\lambda(x, y) := f(x) + \frac{1}{2\lambda}\|x - y\|^2 - e_\lambda g(y). \tag{3}$$

Although the PLK exponents of many functions are known (refer to [6, 23, 24] for more details), deducing the PLK exponent of a function from the known ones is not always straightforward, especially for nonconvex functions; see, e.g., [23, 26]. Many pieces of recent literature have focused on this issue, such as [23, 24, 30, 45]. In light of this, we proceed to explore the sufficient condition for the PLK exponent of the functions assumed to be known. This investigation further enables us to establish a closed circular relationship that connects the PLK exponent of (1) with (2), through the exponent of H_λ and a lifted function (designed for the convergence analysis of a DC algorithm [24, Algorithm 1]) $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$:

$$H(x, y) := f(x) - \langle x, y\rangle + g^*(y), \tag{4}$$

where g^* is the conjugate function of g.

Specifically, for **GD**, in order to have the PLK exponent of F_λ, we show that if H is a PLK function with an exponent within the range $[1/2, 1)$, then this exponent is transferred to F_λ with no further assumption. It has been proven in [24, Theorem 4.1] that if F is a PLK function with an exponent of $1/2$ and g is Lipschitz differentiable, then H is also a PLK function with the same exponent. In this study, we extend this conclusion to cover exponents ranging from $1/2$ to 1. This enables the determination of the PLK exponent of DME functions based on the known exponent of their original DC functions, as illustrated in examples from [24, Corollary 4.1]. As for **IGD**, remember that our local convergence rate result of **IGD** relies on the PLK exponent of the function H_λ. We show that the PLK exponent of H_λ with the range of $[1/2, 1)$ is inherited from that of F_λ when f is continuous. Furthermore, we show that the PLK exponent of H_λ implies that F in (1) possesses the same exponent when g is differentiable. At this point, we actually provide an equivalent relationship for the PLK exponent between F in (1) and F_λ in (2) through H in (4) and H_λ in (3) under suitable assumptions.

The aforementioned PLK equivalence result also provides a new approach to explore the previously unknown PLK exponent of certain DC functions from that of their DME functions. As an application, we aim to address an open question presented in [23], which involves exploring the PLK exponent of a least squares models with logistic penalty [29, 33] in compressed sensing, i.e.,

$$
\begin{aligned}
F(x) &= \frac{1}{2}\|Ax-b\|^2 + \sum_{i=1}^{n} \mu \log\left(1+\frac{|x_i|}{\sigma}\right) \\
&= \underbrace{\frac{1}{2}\|Ax-b\|^2 + \frac{\mu}{\sigma}\|x\|_1}_{f(x)} - \underbrace{\left(\frac{\mu}{\sigma}\|x\|_1 - \sum_{i=1}^{n} \mu \log\left(1+\frac{|x_i|}{\sigma}\right)\right)}_{g(x)},
\end{aligned}
\tag{5}
$$

where $A = I$, $b \in \mathbb{R}^n$, μ and $\sigma > 0$ with $\mu \gg \sigma$. Here, the condition $\mu \gg \sigma$ ensures that the logistic penalty approximates the behavior of the ℓ_0-norm near the origin, thereby inducing sparsity in the solution. We show that the PLK exponent of (5) is $1/2$, which was previously unknown. Prior to that, we establish that the PLK exponent of Lipschitz differentiable functions is $1/2$ under the suitable assumption. This finding extends the known result that twice differentiable functions have the same exponent if their Jacobian matrices at the critical point are nonsingular. Our contributions can be summarized as follows, which can also be visually observed through the following "Tai Chi"-like figure; see Fig. 1:

- We establish the global convergence of the iterates generated by two DME-specific algorithms: **GD** in [40, Algorithm 1] and **IGD** in [38, Algorithm 1], along with their specific local convergence rate. To the best of our knowledge, these results are new for DME-specific algorithms.
- For the first time, we establish the equivalence for the PLK exponent between DC functions and their DME functions under suitable assumptions. Armed with the known PLK exponent for F_λ in (1) and H_λ in (3), the linear or sublinear convergence rates for **GD** and **IGD** are derived based on the convergence rate results established in this paper. This further justifies the advantage for developing and designing DME-specific algorithms.
- The PLK equivalence also serves as a new tool for deducing the PLK exponent of certain DC functions. After demonstrating the PLK exponent of Lipschitz differentiable functions being 1/2 under a suitable assumption, we apply the established PLK equivalence to a specific DME function and determine that the PLK exponent of (5) is $1/2$, which was previously unknown.

We organize our paper as follows. Section 2 gives the notation and some preliminaries. Section 3 establishes the convergence rates for various DME-specific algorithms; Section 4 establishes the equivalence of the PLK exponent between DC functions and their DME functions. Section 5 shows that the PLK exponent of a nonconvex compressed sensing model with logistic penalty sparsity inducing term is $1/2$.

2 Notation and preliminaries

Throughout this paper, we use $\mathbb{R}^n$ and $\mathbb{R}^n_+$ to denote the n-dimensional Euclidean space and the nonnegative orthant, respectively. We let $\bar{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$. For a vector

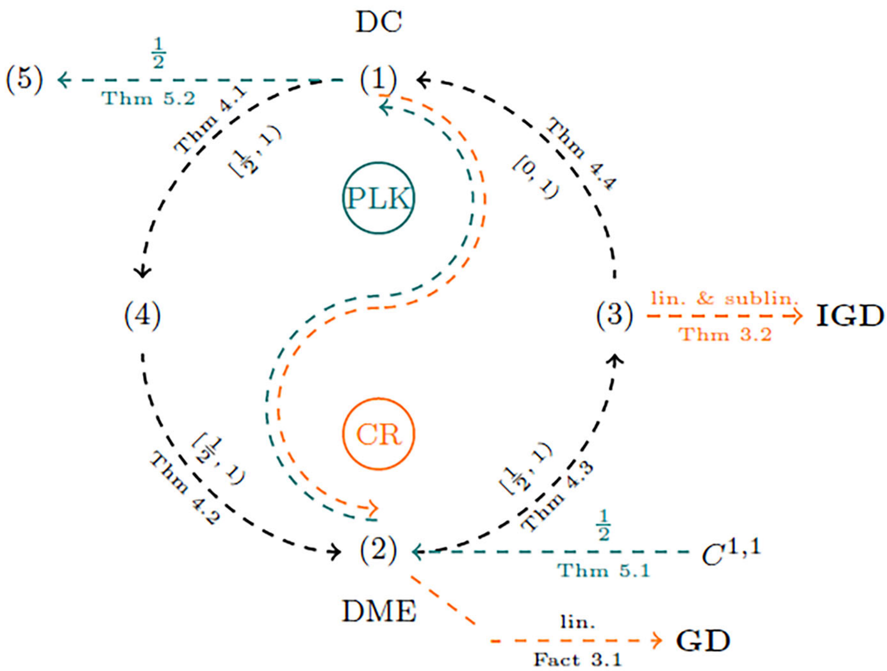

Fig. 1 The "Tai Chi"-like figure. Each black dashed edge, teal dashed line-segment, and orange dashed (poly)line-segment with direction represents a new directional relationship. The corresponding illustrated theorem obtained in this paper is marked near the path with the same color, where "Thm" refers to "Theorem". The labels within the black closed circle, with intervals $[\frac{1}{2}, 1)$ and $[0, 1)$, indicate the range of PLK exponents that can be passed on through the corresponding path; the two "eyes" labeled "PLK" and "CR" represent the "PLK exponent" and "local convergence rate" that can be deduced along the dashed path with the same color. Outside the closed circle, every number $\frac{1}{2}$ above the teal dashed line-segment indicates the PLK exponent of the LHS function that can be deduced as a special case of the RHS function; the labels "lin." and "sublin." above the orange dashed line-segment refer to the "linear" and "sublinear" convergence rates, respectively, that can be deduced for the RHS DME-specific algorithm from the specific PLK exponent of the LHS function

$x \in \mathbb{R}^n$, we denote the ℓ_2 norm and ℓ_1 norm of x as $\|x\|$ and $\|x\|_1$, respectively. The open neighborhood of x with radius $r > 0$ is represented as $B(x, r)$. For any two vectors $x, y \in \mathbb{R}^n$, we let $\langle x, y\rangle$ and $x \cdot y$ denote their inner product and Hadamard product, respectively. For $i = 1, 2, \cdots, n$, we use x_i and B_{ii} to denote the i-th element of vector $x \in \mathbb{R}^n$ and the i-th diagonal element of matrix $B \in \mathbb{R}^{n\times n}$, respectively. The identity matrix in $\mathbb{R}^{n\times n}$ is denoted as I. For a nonempty set $S \subset \mathbb{R}^n$, the distance from any $x \in \mathbb{R}^n$ to S is defined as $\operatorname{dist}(x, S) := \inf_{y\in S} \|x - y\|$. If $S = \emptyset$, then $\operatorname{dist}(x, S) := \infty$ by convention.

A function $h : \mathbb{R}^n \to \bar{\mathbb{R}}$ is considered proper if its domain $\operatorname{dom} h := \{x \in \mathbb{R}^n : h(x) < +\infty\}$ is nonempty. A proper function is said to be closed if it is lower semicontinuous. A proper closed function h is said to be level-bounded if for every $r \in \mathbb{R}$, the set $\{x \in \mathbb{R}^n : h(x) \le r\}$ is bounded and it is said to be level-coercive if there exist some $\alpha \in (0, +\infty)$ and $\tilde{\alpha} \in \mathbb{R}$ such that $F \ge \alpha\|\cdot\| + \tilde{\alpha}$; see [36, Theorem 3.26]. The regular subdifferential [36, Definition 8.3] and Mordukhovich limiting subdifferential

[27, section 4], [28, Definition 1.77] of h at $\bar{x} \in \operatorname{dom} h$ are defined respectively as

$$\widehat{\partial} h(\bar{x}) := \left\{ v \in \mathbb{R}^n : \liminf_{x \to \bar{x}, x \neq \bar{x}} \frac{h(x) - h(\bar{x}) - v^T(x - \bar{x})}{\|x - \bar{x}\|} \geq 0 \right\},$$
$$\partial h(\bar{x}) := \left\{ v \in \mathbb{R}^n : \exists x^k \to \bar{x} \text{ and } v^k \in \widehat{\partial} h(x^k) \text{ s.t. } h(x^k) \to h(\bar{x}) \text{ and } v^k \to v \right\}.$$

We denote $\operatorname{dom} \partial h := \{x \in \mathbb{R}^n : \partial h(x) \neq \emptyset\}$ and $\partial h(x) = \emptyset$ if $x \notin \operatorname{dom} \partial h$ by convention. When h is proper and convex, the Mordukhovich limiting subdifferential of h at $x \in \operatorname{dom} h$ reduces to the classical subdifferential in convex analysis [36, Proposition 8.12]:

$$\partial h(x) = \left\{ \xi \in \mathbb{R}^n : h(y) \geq h(x) + \langle \xi, y - x \rangle, \quad \forall y \in \mathbb{R}^n \right\}.$$

We use $C^{1,1}$ to denote the set of all Lipschitz differentiable functions on $\mathbb{R}^n$. For any function $h \in C^{1,1}$, the class of generalized Hessian matrix of h at $x \in \mathbb{R}^n$, denoted by $\partial^2 h(x)$, is defined in [18, Definition 2.1] as the convex hull of the set (in $\mathbb{R}^{n \times n}$)

$$\{J : \exists \{x^k\} \to x \text{ with } h \text{ is twice differentiable at } x^k \text{ and } \nabla^2 h(x^k) \to J\}.$$

This definition is equivalent to the generalized Jacobian $\partial_C \nabla h(x)$ as defined in [12, Definition 2.6.1]. For every $x \in \mathbb{R}^n$, the set $\partial^2 h(x)$ is a nonempty compact and convex set of symmetric matrices; as shown in [12, Proposition 2.6.2].

The convex conjugate of a proper closed convex function $h : \mathbb{R}^n \to \bar{\mathbb{R}}$ is defined as $h^*(x) = \sup_{y \in \mathbb{R}^n} \{\langle x, y \rangle - h(y)\}$. This definition implies that $h(x) + h^*(y) \geq \langle x, y \rangle$ for all $x, y \in \mathbb{R}^n$ and it is known that

$$h(x) + h^*(y) = \langle x, y \rangle \Longleftrightarrow y \in \partial h(x) \Longleftrightarrow x \in \partial h^*(y). \tag{6}$$

The proximal mapping $\operatorname{prox}_{\lambda h}$ and Moreau envelope $e_\lambda h$ of h with scaling parameters $\lambda > 0$ are defined respectively as the following two formulas:

$$\begin{aligned} \operatorname{prox}_{\lambda h}(x) &:= \arg\min_{u \in \mathbb{R}^n} \left\{ h(u) + \frac{1}{2\lambda} \|x - u\|^2 \right\}, \\ e_\lambda h(x) &:= \inf_{u \in \mathbb{R}^n} \left\{ h(u) + \frac{1}{2\lambda} \|x - u\|^2 \right\}. \end{aligned} \tag{7}$$

Here we gather some of their properties from the literature; see, e.g., [7, 40]:

Fact 2.1 (Some properties of proximal mapping and Moreau envelope) *Let $h : \mathbb{R}^n \to \bar{\mathbb{R}}$ be a proper closed convex function. Then, for all $\lambda > 0$, $e_\lambda h(x)$ always takes real numbers with the unique solution* $\operatorname{prox}_{\lambda h}(x)$ *for every $x \in \mathbb{R}^n$. Furthermore, the following properties hold:*

(i) $1/\lambda \left(x - \operatorname{prox}_{\lambda h}(x)\right) \in \partial h \left(\operatorname{prox}_{\lambda h}(x)\right)$;
(ii) $x - \lambda \operatorname{prox}_{h^*/\lambda}(x/\lambda) \in \partial h^* \left(\operatorname{prox}_{h^*/\lambda}(x/\lambda)\right)$;

(iii) $\mathrm{prox}_{\lambda h}(x) + \lambda\, \mathrm{prox}_{h^*/\lambda}(x/\lambda) = x$*; see [7, Theorem 14.3 (ii)];*
(iv) $e_\lambda h(x) + e_{1/\lambda} h^*(x/\lambda) = 1/(2\lambda)\|x\|^2$*; see [7, Theorem 14.3 (i)];*
(v) $\nabla e_\lambda h(x) = 1/\lambda \left(x - \mathrm{prox}_{\lambda h}(x)\right)$*; see [7, Theorem 12.30];*
(vi) $\mathrm{prox}_{\lambda h}$ *is nonexpansive; see [7, Theorem 12.28];*
(vii) $e_\lambda h$ *is convex and* $1/\lambda$*-Lipschitz differentiable; see [7, Proposition 12.15 and 12.30].*

Remark 2.1 In view of Fact 2.1(ii)-(iii), we can deduce that for all $\lambda > 0$,

$$\mathrm{prox}_{\lambda h}(x) \in \partial h^* \left(\mathrm{prox}_{h^*/\lambda}(x/\lambda)\right). \tag{8}$$

A point $\bar{x} \in \mathbb{R}^n$ is said to be a stationary point of (1) if

$$\partial f(\bar{x}) \cap \partial g(\bar{x}) \neq \emptyset. \tag{9}$$

In the next Fact, we review some properties of the DME function denoted as F_λ in (2), as presented in [40, Lemma 4].

Fact 2.2 *(Basic properties of DME function) Consider* (1) *and* (2)*. For all* $\lambda > 0$*, we have the following properties of* F_λ *in* (2)*:*

(i) F_λ *is* $1/\lambda$*-Lipschitz differentiable with gradient*

$$\nabla F_\lambda(x) = 1/\lambda \left(\mathrm{prox}_{\lambda g}(x) - \mathrm{prox}_{\lambda f}(x)\right), \ \forall x \in \mathbb{R}^n. \tag{10}$$

(ii) $\nabla F_\lambda(\bar{x}) = 0$ *if and only if* $\mathrm{prox}_{\lambda f}(\bar{x})$ *is a stationary point of* (1) *in the sense of* (9).

Fact 2.2 (ii) implies that the stationary points of the DC problem (1) in the sense of (9) can be obtained from the stationary points of its DME-based model (2). This observation serves as the main motivation behind the development of DME-specific algorithms; see [38, 40] for detailed discussion.

Now we recall the definition of metric regularity, which can be seen in [2, Definition 2.1] and [13, Subsection 3E], for example.

Definition 2.1 A mapping $\mathcal{H} : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is said to be metrically regular at $\bar{x}$ for $\bar{y}$ when $\bar{y} \in \mathcal{H}(\bar{x})$ and there is a constant $k > 0$ together with neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that for all $x \in U$ and $y \in V$,

$$\mathrm{dist}(x, \mathcal{H}^{-1}(y)) \leq k\, \mathrm{dist}(y, \mathcal{H}(x)).$$

It can be observed that the metric regularity in above definition restricts the distance from x to $\mathcal{H}^{-1}(y)$, which is in turn bounded by the scaled distance between y and $\mathcal{H}(x)$ for any (x, y) around $(\bar{x}, \bar{y})$. More importantly, it requires V to be a subset of the range of $\mathcal{H}$. Notably, if a function $h : \mathbb{R}^n \to \mathbb{R}$ is locally strongly convex around $\bar{x}$, then its subdifferential ∂h satisfies metric regularity at the point $\bar{x}$ for any $\bar{y} \in \partial h(\bar{x})$. This can be verified using [3, Theorem 5.26(b)] and [36, Theorem 9.43]. To clarify, the local strong convexity here means that there exists $\epsilon > 0$ such that h is strongly convex in the neighborhood $B(\bar{x}, \epsilon)$ of $\bar{x}$; see [34, Equation (9)] for example.

Moreover, there exists a close connection between metric regularity and the quadratic growth condition. Here, the definition of quadratic growth condition is consistent with that in [2, 9], among others. Specifically, as shown in [2, Theorem 3.3], the metric regularity of ∂h at the critical point $\hat{x}$ of h for $\hat{y} \in \partial h(\hat{x})$ satisfies the quadratic growth condition. However, the converse does not necessarily hold.

In the rest of this section, we review the definitions of the PLK property and the PLK exponent; see also [4–6, 11, 23].

Definition 2.2 (PLK property and PLK exponent) We say that a proper closed function $h : \mathbb{R}^n \to \bar{\mathbb{R}}$ has the PLK property at $\bar{x} \in \operatorname{dom} \partial h$ if there exist a neighborhood $\mathcal{N}$ of $\bar{x}$, $v \in (0, +\infty]$ and a continuous concave function $\psi : [0, v) \to \mathbb{R}_+$ with $\psi(0) = 0$ as the desingularized function such that:

(i) ψ is continuously differentiable on $(0, v)$ with $\psi' > 0$ over $(0, v)$;
(ii) for all $x \in \mathcal{N}$ with $h(\bar{x}) < h(x) < h(\bar{x}) + v$, one has

$$\psi'(h(x) - h(\bar{x}))\, \operatorname{dist}(0, \partial h(x)) \geq 1.$$

We call h a PLK function if it satisfies the PLK property at all points of $\operatorname{dom} \partial h$.

If ψ can be chosen as $\psi(s) = \bar{c}s^{1-\alpha}$ for some $\bar{c} > 0$ and $\alpha \in [0, 1)$, then we say that h has the PLK property at $\bar{x}$ with an exponent of α. If h is a PLK function with the same exponent α at any $\bar{x} \in \operatorname{dom} \partial h$, then h is called a PLK function with exponent α.

As a generalization to nonsmooth settings, the above exponential PLK property originates from the Polyak-Łojasiewicz (PL) property, which was first introduced by Polyak from an optimization perspective in [32, Theorem 4] to establish the linear convergence rate of gradient descent method for Lipschitz differentiable functions in Hilbert spaces. For more formation and evolution process of exponential PLK property, we refer readers to, e.g., [10, Section 1].

3 Convergence Analysis for DME-Specific Algorithms

In this section, we present a convergence analysis of two DME-specific algorithms: **GD** (Gradient Descent) in [40, Algorithm 1] and **IGD** (Inexact Gradient Descent) in [38, Algorithm 1]. These algorithms are designed respectively to solve the DC problem (1) with and without new structured settings. Our analysis includes global convergence and the corresponding local convergence rate. Without any specific statement, we always assume that $\lambda > 0$ is any selected number in this section. To facilitate the convergence analysis, we also make the following general assumption.

Assumption 3.1 The DME function F_λ in (2) is level-bounded.

Note that the lower boundedness of F_λ can be derived from the lower boundedness of its original DC function F, and the level boundedness of F_λ can be satisfied according to [38, Proposition 2] by requiring that F in (1) satisfies one of the following sufficient conditions: (i) F is level-bounded and g is Lipschitz up to some constant, that is, there exist $L > 0$ and $M \geq 0$ such that $\|g(x) - g(y)\| \leq L\|x - y\| + M$ for all $x, z \in \mathbb{R}^n$, (ii) F is level-coercive, (iii) dom f is compact, (iv) F is level-bounded and dom $f = \mathbb{R}^n$. Moreover, the set of global minimizers for minimizing F in (1) is nonempty under Assumption 3.1, which follows from Fact 2.2. Therefore, it can be verified that Assumption 3.1 is suitable for the convergence analysis of DME-specific algorithms.

Let us consider **GD** from [40, Algorithm 1] as Algorithm 1 under Assumption 3.1. It has been well studied in [32, Theorem 4] and reference therein that the following fact holds.

Fact 3.1 *The iterates generated by* **GD** *under the PLK exponent of* F_λ *being* $1/2$ *are globally convergent with linear convergence rate.*

Remark 3.1 Under the PLK assumption of F_λ, the local convergence rate of **GD** cannot be improved from linear to superlinear. This limitation stems from the established result in [35, Remark 2.21] and [10, Theorem 4], which proves that the PLK exponent of $C^{1,1}$ functions at their local minimizers always lies outside the interval $(0, 1/2)$ – the key range known to enable superlinear convergence rates in first-order optimization methods (see [6, Theorem 3.2] for example).

Algorithm 1: GD [40, Algorithm 1]: Classical Gradient Descent on F_λ for solving (1) under Assumption 3.1

- **Select** $0 < \alpha < 2$;
- **Initialize** $s^0 \in \mathbb{R}^n$;
- **for** $t = 0, 1, \cdots$ **do**;
- $u^{t+1} = \text{prox}_{\lambda g}(s^t)$;
- $v^{t+1} = \text{prox}_{\lambda f}(s^t)$;
- $s^{t+1} = s^t + \alpha(v^{t+1} - u^{t+1})$;
- **end for**.

Now, we analyze the global convergence of the iterates generated by another DME-specific algorithm **IGD** below as in [38, Algorithm 1] and its local convergence rate. This algorithm is designed to solve the structured DC problem (1):

$$\min_{x \in \mathbb{R}^n} F(x) := f(x) - g(x), \tag{11}$$

where $f := p + q$, $q : \mathbb{R}^n \to \bar{\mathbb{R}}$ is proper, closed and convex, $p : \mathbb{R}^n \to \mathbb{R}$ has Lipschitz gradient with modulus L_p over dom q and $g : \mathbb{R}^n \to \mathbb{R}$ is continuous. To have the convergence analysis of **IGD**, we now present the following proposition as the first step towards our goal.

Proposition 3.1 *Consider problem* (11) *under Assumption 3.1. Let* $\{x^t\}$ *and* $\{z^t\}$ *be generated by* **IGD**. *Let* Ω *and* Γ *be the sets of all accumulation points of* $\{x^t\}$ *and* $\{(x^t, z^t)\}$, *respectively. Then we have*

Algorithm 2: IGD [38, Algorithm 1]: Inexact Gradient Descent on F_λ for solving (11) under Assumption 3.1

- **Let** $0 < \lambda < 1/L_p$ and $0 < \beta < 2$;
- **Initialize** $x^0 \in \operatorname{dom} q$ and $z^0 \in \mathbb{R}^n$;
- **for** $t = 0, 1, \cdots$ **do**;
- $x^{t+1} = \arg\min_{x\in\mathbb{R}^n} \left\{\langle \nabla p(x^t), x\rangle + q(x) + \frac{1}{2\lambda}\|x - z^t\|^2\right\}$;
- $\operatorname{prox}_{\lambda g}(z^t) = \arg\min_{x\in\mathbb{R}^n} \left\{g(x) + \frac{1}{2\lambda}\|x - z^t\|^2\right\}$;
- $z^{t+1} = z^k + \beta(x^{t+1} - \operatorname{prox}_{\lambda g}(z^t))$;
- **end for**.

(i) $\lim_{t\to\infty} \|x^{t+1} - x^t\| = 0$ *and* $\lim_{t\to\infty} \|z^{t+1} - z^t\| = 0$.
(ii) $\zeta := \lim_{t\to\infty} F(x^t)$ *exists.*
(iii) $F(\hat{x}) \equiv \zeta$ *whenever* $\hat{x} \in \Omega$.
(iv) $H_\lambda \equiv \zeta$ *on the set* Γ.

Proof (i): We have from [38, Lemma 2] that

$$\begin{aligned}\sum_{i=t}^{\infty} \|x^{i+1} - x^i\|^2 &\le \frac{1}{c_1}\sum_{i=t}^{\infty}\Big(H_\lambda(x^i, z^i) - H_\lambda(x^{i+1}, z^{i+1})\Big)\\ &\le \frac{1}{c_1}\left(H_\lambda(x^t, z^t) - \liminf_{i\to\infty} H_\lambda(x^i, z^i)\right) < +\infty,\end{aligned}$$

for some $c_1 > 0$, where the last inequality follows from Assumption 3.1 and $\inf H_\lambda = \inf F_\lambda = \inf F > -\infty$. This implies $\lim_{t\to\infty} \|x^{t+1} - x^t\| = 0$. Similarly, we can establish $\lim_{t\to\infty} \|z^{t+1} - z^t\| = 0$, which completes the proof.

(ii): According to [38, Lemma 2] and the lower boundedness of H_λ, we have

$$\lim_{t\to\infty} H_\lambda(x^t, z^t) = \zeta. \tag{12}$$

for some $\zeta \in \mathbb{R}$. Now we aim to prove the convergence of the sequence $\{F(x^t)\}$. Due to the boundedness of $\{x^t\}$ and $\{z^t\}$ (see [38, Theorem 1]) as well as the Lipschitz continuity of $\operatorname{prox}_{\lambda g}$, we can conclude that the sequence $\{\|2z^t\| + \|x^t\| + \|\operatorname{prox}_{\lambda g}(z^t)\|\}$ is also bounded. Let M_1 be an upper bound of this sequence and take $\xi^t \in \partial g(x^t)$. Then we have that

$$\begin{aligned}H_\lambda(x^t, z^t) =& F(x^t) + g(x^t) - g(\operatorname{prox}_{\lambda g}(z^t))\\ &+ \frac{1}{2\lambda}\left(\|z^t - x^t\|^2 - \|\operatorname{prox}_{\lambda g}(z^t) - z^t\|^2\right)\\ \le& F(x^t) + \langle \xi^t, x^t - \operatorname{prox}_{\lambda g}(z^t)\rangle\\ &+ \frac{1}{2\lambda}\|\operatorname{prox}_{\lambda g}(z^t) - x^t\|\left(\|z^t - x^t\| + \|z^t - \operatorname{prox}_{\lambda g}(z^t)\|\right)\\ \le& F(x^t) + \langle \xi^t, x^t - \operatorname{prox}_{\lambda g}(z^t)\rangle + \frac{M_1}{2\lambda}\|\operatorname{prox}_{\lambda g}(z^t) - x^t\|\end{aligned}$$

$$\leq F(x^t) + \langle \xi^t, x^t - x^{t+1} \rangle$$
$$+ \frac{1}{\beta}\langle \xi^t, z^{t+1} - z^t \rangle + \frac{M_1}{2\lambda}\left(\|x^{t+1} - x^t\| + \frac{1}{\beta}\|z^{t+1} - z^t\| \right),$$

where the first inequality follows from $\|a\|^2 - \|b\|^2 \leq \|a - b\|(\|a\| + \|b\|)$ for every pair $a, b \in \mathbb{R}^n$. Note that $H_\lambda(x^t, z^t) \geq F(x^t)$ for every t. This together with (12), item (i) and the above display shows that $\lim_{t\to\infty} F(x^t) = \zeta$ from the boundedness of $\{\bar{\xi}^t : \bar{\xi}^t \in \partial g(x^t), t = 0, 1, \cdots\}$. This completes the proof.

(iii): Let $\hat{x}$ be an arbitrary point in Ω. Then there exists a subsequence $\{x^{t_j}\}$ of $\{x^t\}$ such that $\lim_{j\to\infty} x^{t_j} = \hat{x}$. Now we aim to prove that $F(\hat{x}) = \zeta$.

From $x^{t_j} = \arg\min_{x\in\mathbb{R}^n} \left\{ \langle \nabla p(x^{t_j-1}), x \rangle + q(x) + \frac{1}{2\lambda}\|x - z^{t_j-1}\|^2 \right\}$, we have that

$$\langle \nabla p(x^{t_j-1}), x^{t_j} - \hat{x} \rangle + q(x^{t_j}) + \frac{1}{2\lambda}\|x^{t_j} - z^{t_j-1}\|^2 \leq q(\hat{x}) + \frac{1}{2\lambda}\|\hat{x} - z^{t_j-1}\|^2. \quad (13)$$

Taking M_2 as the upper bound of $\{\|x^{t_j} - z^{t_j-1}\| + \|\hat{x} - z^{t_j-1}\|\}$ by using the boundedness of $\{x^t\}$ and $\{z^t\}$ again from [38, Theorem 1], we have that

$$\|\hat{x} - z^{t_j-1}\|^2 - \|x^{t_j} - z^{t_j-1}\|^2 \leq \left(\|x^{t_j} - z^{t_j-1}\| + \|\hat{x} - z^{t_j-1}\| \right) \|x^{t_j} - \hat{x}\|$$
$$\leq M_2\|x^{t_j} - \hat{x}\|,$$

where the first inequality follows from $\|a\|^2 - \|b\|^2 \leq \|a - b\|(\|a\| + \|b\|)$ for every pair $a, b \in \mathbb{R}^n$. This together with (13) implies that

$$\langle \nabla p(x^{t_j-1}), x^{t_j} - \hat{x} \rangle + q(x^{t_j}) \leq q(\hat{x}) + \frac{M_2}{2\lambda}\|\hat{x} - x^{t_j}\|. \quad (14)$$

Therefore, we have that

$$\zeta = \lim_{j\to\infty} F(x^{t_j}) = \lim_{j\to\infty} \left[p(x^{t_j}) + q(x^{t_j}) - g(x^{t_j}) + \langle \nabla p(x^{t_j-1}), x^{t_j} - \hat{x} \rangle \right]$$
$$\leq \lim_{j\to\infty} \left[p(x^{t_j}) + q(\hat{x}) - g(x^{t_j}) + \frac{M_2}{2\lambda}\|\hat{x} - x^{t_j}\| \right] = F(\hat{x}),$$

where the second equality uses $\lim_{j\to\infty} x^{t_j} = \hat{x}$ and the Lipschitz continuity of ∇p, the inequality follows from (14) and the last equality uses $\lim_{j\to\infty} x^{t_j} = \hat{x}$. This together with $F(\hat{x}) \leq \liminf_{j\to\infty} F(x^{t_j}) = \lim_{j\to\infty} F(x^{t_j}) = \zeta$ (thanks to the lower semicontinuity of F in the first inequality) implies that $F(\hat{x}) = \zeta$. This completes the proof.

(iv): For any $(x^*, z^*) \in \Gamma$, we have from [38, Theorem 1] that $x^* = \mathrm{prox}_{\lambda g}(z^*)$. This gives us that

$$H_\lambda(x^*, z^*) = f(\mathrm{prox}_{\lambda g}(z^*)) + \frac{1}{2\lambda}\|\mathrm{prox}_{\lambda g}(z^*) - z^*\|^2 - e_\lambda g(z^*)$$

$$= f(\mathrm{prox}_{\lambda g}(z^*)) - g(\mathrm{prox}_{\lambda g}(z^*)) = F(x^*) = \zeta,$$

which implies that $H_\lambda \equiv \zeta$ on the set Γ. This completes the proof. □

With the theorem above, we can establish the global convergence of $\{x^k\}$ generated by **IGD** in the following theorem. The proof follows standard procedures as in [42, Theorem 4.2 (iv)] and is omitted for brevity.

Theorem 3.1 (Global convergence of $\{x^k\}$ and $\{z^k\}$ generated by **IGD**) *Consider* (11) *under Assumption 3.1. Let $\{x^k\}$ and $\{z^k\}$ be generated by* **IGD**. *Suppose that H_λ is a PLK function. Then the sequences $\{x^k\}$ and $\{z^k\}$ are convergent. Moreover, we have*

$$\sum_{t=0}^{\infty}\left(\|x^t - x^{t+1}\| + \|z^t - z^{t+1}\|\right) < +\infty.$$

Now, we are ready to show the local convergence rate of **IGD** in the following theorem. The proof is similar to that of [42, Theorem 4.3] by taking x^t in [42, Theorem 4.3] replaced by $Y^t := \begin{bmatrix} x^t \\ z^t \end{bmatrix}$ generated by **IGD**. For brevity, we omit the proof.

Theorem 3.2 [Local convergence rate of **IGD**] *Let $\{x^t\}$ be a sequence generated by* **IGD** *for solving* (11). *Suppose that $\lim_{t\to\infty} x^t = \bar{x}$ for some $\bar{x} \in \mathbb{R}^n$, and H_λ is a PLK function with $\psi = cs^{1-\theta}$ being the desingularized function for some $\theta \in [0, 1)$ and $c > 0$. Then one has that*

(i) If $\theta = 0$, then there exists $t_0 > 0$ so that x^t is constant for $t > t_0$;
(ii) If $\theta \in (0, \frac{1}{2}]$, then there exist $c_1 > 0$, $t_1 > 0$ and $\eta \in (0, 1)$ such that $\|x^t - \bar{x}\| < c_1\eta^t$ for $t > t_1$;
(iii) If $\theta \in (\frac{1}{2}, 1)$, then there exist $c_2 > 0$, $t_2 > 0$ such that $\|x^t - \bar{x}\| < c_2 t^{-\frac{1-\theta}{2\theta-1}}$ for $t > t_2$.

4 The Equivalence for PLK Exponent Between DC Functions and Their DME Functions

In this section, we establish the equivalence for the PLK exponent between the DC function F in (1) and its DME function F_λ in (2) for any fixed $\lambda > 0$. We begin by listing our equivalence results from Theorem 4.1 to Theorem 4.4, and then followed by more detailed explanations of their proofs.

We first establish in Theorem 4.1 that the PLK exponent in the range $[1/2, 1)$ can be inherited from F in (1) to H in (4). The inheritance of the PLK exponent from F to H has been deeply discussed in [24, Theorem 4.1]. However [24, Theorem 4.1] only addresses the case where the PLK exponent is $1/2$ under the assumption that g is Lipschitz differentiable. Our result extends this conclusion to cover a broader range of exponents, from $1/2$ to 1, under the metric regularity of ∇g.

Theorem 4.1 (Deriving the PLK exponent of H from that of F) *Consider* (1) *and* (4). *We have the following two statements regarding the transmission of the PLK exponent from F in* (1) *to H in* (4):

(i) *Suppose that* $g : \mathbb{R}^n \to \mathbb{R}$ *is Lipschitz differentiable. If* F *is a PLK function with an exponent of* $\frac{1}{2}$, *then* H *is also a PLK function with the same exponent of* $\frac{1}{2}$.

(ii) *If we assume that* g *is locally Lipschitz differentiable and its gradient* ∇g *is metrically regular at every stationary point* x *of* F *in the sense of* (9) *for* $\nabla g(x)$, *then we have the following implication: if* F *has the PLK exponent of* $\alpha \in [\frac{1}{2}, 1)$ *at* $\bar{x} \in \mathrm{dom}\partial F$, *then* H *satisfies the PLK property at* $(\bar{x}, \nabla g(\bar{x})) \in \mathrm{dom}\partial H$ *with the same exponent* α.

In the second theorem of this section, we prove that the PLK exponent in the range $[1/2, 1)$ can be inherited from H in (4) to F_λ in (2) without imposing any additional assumptions. To the best of our knowledge, this combining with Theorem 4.1 provides a new way to determine the PLK exponent of DME functions based on the known one of their original DC functions, as illustrated in examples from [24, Corollary 4.1]. This result does not assume that F_λ is (sub)analytic compared to earlier results; see [11] and references therein. Moreover, it is closely related to the study of global convergence and local convergence rates of certain DME-type algorithms to solve DC programming problems, as exemplified by Fact 3.1 in Section 3.

Theorem 4.2 (Deriving the PLK exponent of F_λ from that of H)*Consider* (4) *and* (2). *If* H *in* (4) *has the PLK property at*

$$(\bar{x}, \bar{y}) := \left(\mathrm{prox}_{\lambda f}(\bar{\xi}), \mathrm{prox}_{g^*/\lambda}(\bar{\xi}/\lambda)\right)$$

for some $\bar{\xi} \in \mathbb{R}^n$ *with an exponent of* $\alpha \in [\frac{1}{2}, 1)$, *then* F_λ *in* (2) *has the PLK property at* $\bar{\xi}$ *with the same exponent* α. *Specifically, if* H *is a PLK function with an exponent of* $\alpha \in [\frac{1}{2}, 1)$, *then* F_λ *is also a PLK function with the exponent* α.

Note that F_λ in (2) can be constructed as an inf-projection of H_λ in (3); see [44]. This establishes a “reverse" inf-projection relationship from F_λ to H_λ. Consequently, our next Theorem 4.3 not only provides a link to connect the PLK exponents but also yields an interesting converse result to [44, Theorem 3.1] without any additional assumptions beyond the continuity of f. Moreover, Theorem 4.3 also offers a way to establish the linear or sublinear convergence rate of the first-order DME-type algorithm, **IGD**, as demonstrated in Theorem 3.2 of this paper.

Theorem 4.3 (Deriving the PLK exponent of H_λ from that of F_λ)*Consider* (2) *and* (3). *Suppose* f *is continuous on* $\mathrm{dom} f$. *If* F_λ *is a PLK function with exponent of* $\alpha \in [\frac{1}{2}, 1)$. *Then* H_λ *is also a PLK function with the exponent* α.

Finally, as the last piece of our closed circle-puzzle, Theorem 4.4 aims to deduce the PLK exponent of DC function F in (1) from that of H_λ in (3) under suitable assumptions. This problem has been deeply studied in [44, Theorem 3.1]. However, the result in [44, Theorem 3.1] relies on the assumption of locally uniform level boundedness, which can be ensured by strong convexity or by convexity and compactness; see [44, Corollary 3.3]. Whether H_λ, as a nonconvex function in general, satisfies locally uniform level boundedness remains unclear, posing a main challenge to directly applying the proof of [44, Theorem 3.1] to prove Theorem 4.4. To overcome this obstacle, we assume that g in H_λ is differentiable. This specific structure of H_λ enables us to establish the desired properties, similar to those in the conclusion of [44, Lemma 2.2] under locally uniform level boundedness, which were used to prove [44, Theorem 3.1] under the PLK assumption.

Theorem 4.4 (Deriving the PLK exponent of F from that of H_λ) *Consider* (3) *and* (1). *Suppose g is differentiable. If H_λ is a PLK function with exponent of $\alpha \in [0, 1)$, then F is also a PLK function with the same exponent α.*

In the remainder of this section, we give our proofs of above four theorems. Let us show the proof of Theorem 4.1 first.

Proof of Theorem 4.1 (i): Note that if (x, y) satisfies $0 \in \partial H(x, y)$, then $y = \nabla g(x)$ holds from the differentiability of g. This implies that the first part of this theorem holds, as demonstrated in the proof of [24, Theorem 4.1].

(ii): Suppose that $\bar{x} \in \operatorname{dom} \partial F$ satisfies $0 \in \partial F(\bar{x})$, which means that $\bar{x}$ is a stationary point of F, as defined in (9). Then the local Lipschitz differentiability of g implies that

$$\nabla g(\bar{x}) \in \partial f(\bar{x}). \tag{15}$$

Since g is also assumed to be closed and convex, one has that $\bar{x} \in \partial g^*(\nabla g(\bar{x}))$. This together with (15) gives $0 \in \partial H(\bar{x}, \nabla g(\bar{x}))$. Moreover, we have that

$$H(\bar{x}, \nabla g(\bar{x})) = f(\bar{x}) - \langle \bar{x}, \nabla g(\bar{x})\rangle + g^*(\nabla g(\bar{x})) = f(\bar{x}) - g(\bar{x}) = F(\bar{x}), \tag{16}$$

where the second equality follows from (6).

Let $1 > \epsilon_1 > 0$ be such that ∇g is Lipschitz continuous in $B(\bar{x}, \epsilon_1)$ with $l \geq 0$ being its Lipschitz modulus. Since ∇g is metrically regular at $\bar{x}$ for $\nabla g(\bar{x})$, there exist $\epsilon_2 > 0$ and $\mathcal{M} > 0$, such that for any $x \in B(\bar{x}, \epsilon_2)$ and $y \in B(\nabla g(\bar{x}), \epsilon_2)$, one has that $\operatorname{dist}(x, \partial g^*(y)) \leq \mathcal{M} \operatorname{dist}(y, \nabla g(x))$. Now, take $0 < \epsilon_3 < \min\{\frac{\epsilon_1}{2}, \epsilon_2, \frac{\epsilon_1}{2\mathcal{M}(l+1)}\}$. Then for any $(x, y) \in \operatorname{dom}\partial H$ satisfying $x \in B(\bar{x}, \epsilon_3)$ and $y \in B(\nabla g(\bar{x}), \epsilon_3)$, one has that

$$\begin{aligned}\operatorname{dist}(x, \partial g^*(y)) \leq \mathcal{M}\|\nabla g(x) - y\| &\leq \mathcal{M}\|\nabla g(x) - \nabla g(\bar{x})\| + \mathcal{M}\|\nabla g(\bar{x}) - y\| \\ &\leq \mathcal{M}(l+1)\epsilon_3 < \frac{\epsilon_1}{2} < 1.\end{aligned} \tag{17}$$

Moreover, for any $0 < \tilde{\epsilon} < \frac{\epsilon_1}{2} - \operatorname{dist}(x, \partial g^*(y))$, there exists $u_{\tilde{\epsilon}} \in \partial g^*(y)$ such that

$$\|u_{\tilde{\epsilon}} - x\| < \operatorname{dist}(x, \partial g^*(y)) + \tilde{\epsilon} < \frac{\epsilon_1}{2}.$$

This means $\|u_{\tilde{\epsilon}} - \bar{x}\| \leq \|u_{\tilde{\epsilon}} - x\| + \|x - \bar{x}\| < \frac{\epsilon_1}{2} + \epsilon_3 < \epsilon_1$, which is $u_{\tilde{\epsilon}} \in B(\bar{x}, \epsilon_1) \cap \partial g^*(y)$. Recalling that ∇g is Lipschitz continuous in $B(\bar{x}, \epsilon_1)$ with constant l, one follows from (6) and [7, Lemma 2.64] that

$$\begin{aligned}H(x, y) &= F(x) + g(x) - \langle x, y\rangle + g^*(y) \\ &= F(x) + g(x) - g(u_{\tilde{\epsilon}}) + \langle u_{\tilde{\epsilon}} - x, y\rangle\end{aligned}$$

$$\leq F(x) + \frac{l}{2}\|u_{\tilde{\epsilon}} - x\|^2 \leq F(x) + \frac{l}{2}\left(\text{dist}(x, \partial g^*(y)) + \tilde{\epsilon}\right)^2,$$

By letting $\tilde{\epsilon} \to 0$, we have from the boundedness of $\text{dist}(x, \partial g^*(y))$ based on (17) that

$$H(x, y) \leq F(x) + \frac{l}{2}\text{dist}^2(x, \partial g^*(y)). \tag{18}$$

Since F satisfies PLK property at $\bar{x} \in \text{dom}\partial F = \text{dom}\partial f$ with exponent $\alpha \in [\frac{1}{2}, 1)$, there exist $c \in (0, 1)$ and $0 < \epsilon < \epsilon_3$ such that for every $x \in \text{dom}\partial F$ satisfying $\|x - \bar{x}\| \leq \epsilon$ and $F(x) < F(\bar{x}) + \epsilon$, we have

$$\text{dist}^{\frac{1}{\alpha}}(0, \partial F(x)) \geq c\left(F(x) - F(\bar{x})\right). \tag{19}$$

Since g is Lipschitz differentiable in $B(\bar{x}, \epsilon)$, we have from [36, Theorem 9.43] that ∂g^* is metrically regular at $\nabla g(\bar{x})$ for $\bar{x}$. This means by shrinking $\epsilon > 0$ and c if necessary that one can infer that for any (x, y) satisfying $\|x - \bar{x}\| \leq \epsilon$ and $\|y - \nabla g(\bar{x})\| \leq \epsilon$,

$$\text{dist}(x, \partial g^*(y)) \geq c\|y - \nabla g(x)\|. \tag{20}$$

By using $\epsilon, c > 0$ as in (20), we consider $(x, y) \in \text{dom}\partial H$ with $\max\{\|x - \bar{x}\|, \|y - \nabla g(\bar{x})\|\} \leq \epsilon$ and $H(\bar{x}, \nabla g(\bar{x})) < H(x, y) < H(\bar{x}, \nabla g(\bar{x})) + \epsilon$. Then $x \in \text{dom}\partial f,\ y \in \text{dom}\partial g^*$ and furthermore,

$$F(\bar{x}) + \epsilon > H(x, y) \geq F(x), \tag{21}$$

where the first inequality follows from (16) and the last one follows from Fenchel-Young inequality. Now, for any such (x, y), one has that

$$\begin{aligned}
2\,\text{dist}^{\frac{1}{\alpha}}(0, \partial H(x, y)) &\geq \text{dist}^{\frac{1}{\alpha}}\left(0, \begin{bmatrix} \partial f(x) - y \\ -x + \partial g^*(y) \end{bmatrix}\right) + \text{dist}^{\frac{1}{\alpha}}(x, \partial g^*(y)) \\
&\overset{(a)}{\geq} \bar{c}\left(\text{dist}^{\frac{1}{\alpha}}(y, \partial f(x)) + \text{dist}^{\frac{1}{\alpha}}(x, \partial g^*(y)) + \text{dist}^{\frac{1}{\alpha}}(x, \partial g^*(y))\right) \\
&\overset{(b)}{\geq} \bar{c}\left(\text{dist}^{\frac{1}{\alpha}}(y, \partial f(x)) + c^{\frac{1}{\alpha}}\|y - \nabla g(x)\|^{\frac{1}{\alpha}} + \text{dist}^{\frac{1}{\alpha}}(x, \partial g^*(y))\right) \\
&\overset{(c)}{\geq} \tilde{c}\left(\text{dist}^{\frac{1}{\alpha}}(y, \partial f(x)) + \|y - \nabla g(x)\|^{\frac{1}{\alpha}} + \text{dist}^{\frac{1}{\alpha}}(x, \partial g^*(y))\right) \\
&\overset{(d)}{\geq} \frac{1}{2^{\frac{1}{\alpha}-1}}\tilde{c}\left(\text{dist}^{\frac{1}{\alpha}}(\nabla g(x), \partial f(x)) + \text{dist}^{\frac{1}{\alpha}}(x, \partial g^*(y))\right) \\
&\overset{(e)}{=} \hat{c}\left(\text{dist}^{\frac{1}{\alpha}}(0, \partial F(x)) + \text{dist}^{\frac{1}{\alpha}}(x, \partial g^*(y))\right) \\
&\overset{(f)}{\geq} c^*\left(\frac{1}{c}\text{dist}^{\frac{1}{\alpha}}(0, \partial F(x)) + \frac{l}{2}\text{dist}^{\frac{1}{\alpha}}(x, \partial g^*(y))\right) \\
&\overset{(g)}{\geq} c^*\left(F(x) - F(\bar{x}) + \frac{l}{2}\text{dist}^{\frac{1}{\alpha}}(x, \partial g^*(y))\right) \\
&\overset{(h)}{\geq} c^*\left(F(x) - F(\bar{x}) + \frac{l}{2}\text{dist}^2(x, \partial g^*(y))\right) \overset{(k)}{\geq} c^*\left(H(x, y) - H(\bar{x}, \bar{y})\right),
\end{aligned}$$

where (a) follows from [23, Lemma 2.2] with the existence of $\bar{c} \in (0, 1)$ and $\frac{1}{2} \leq \alpha < 1$, (b) follows from (20), (c) follows from $0 < c < 1$ and $\tilde{c} := \bar{c}c^{\frac{1}{\alpha}} \in (0, \bar{c})$, (d) holds because of $\alpha < 1$ and the convexity of function $\|\cdot\|^{\frac{1}{\alpha}}$, (e) follows from $\hat{c} := \frac{1}{2^{\frac{1}{\alpha}-1}}\tilde{c}$, (f) follows from $c^* := \frac{\hat{c}}{\frac{1}{c}+\frac{l}{2}}$, (g) holds by using (19) and (21), (h) follows from (17) and $1 < \frac{1}{\alpha} \leq 2$, (k) follows from (16) and (18). This completes the proof. □

Remark 4.1 It can be seen from the second part of the above proof that the assumption of metric regularity as in Theorem 4.1 is crucial, along with the local Lipschitz differentiability of g. This is because for any (x, y) locally around $(\bar{x}, \nabla g(\bar{x}))$, we first utilize this assumption to restrict the distance from x to $\partial g^*(y)$ to be sufficiently small in (17). Then, we leverage it to ensure the applicability of g's local Lipschitz differentiability, thereby allowing us to upper bound $H(x, y)$ by $F(x)$ and $\operatorname{dist}(x, \partial g^*(y))$ as in (18). Within the current proof framework, these two relationships are crucial to our conclusion. However, if we consider g to be a constant function on $\mathbb{R}^n$, then it satisfies all assumptions on g in Theorem 4.1 except for the required metric regularity. By using [3, Corollary 4.21], this can be seen from $\partial g^*(x) = \begin{cases} \emptyset & \text{if } x \neq 0, \\ \mathbb{R}^n & \text{if } x = 0. \end{cases}$ Note that this subdifferential of g^* implies that the above crucial relationship in (17) fails to hold. This demonstrates the assumption of metric regularity in Theorem 4.1 is indispensable under the current framework of its proof.

Now, we give the proofs of Theorem 4.2 and Theorem 4.3 successively.

Proof of Theorem 4.2 We first claim that if $\bar{\xi}$ satisfies $\nabla F_\lambda(\bar{\xi}) = 0$, then $0 \in \partial H(\bar{x}, \bar{y})$ holds by letting

$$\bar{x} := \operatorname{prox}_{\lambda f}(\bar{\xi}) \quad \text{and} \quad \bar{y} := \operatorname{prox}_{g^*/\lambda}(\bar{\xi}/\lambda). \tag{22}$$

In fact, note that

$$\partial H(x, y) = \begin{bmatrix} -y + \partial f(x) \\ -x + \partial g^*(y) \end{bmatrix} \tag{23}$$

holds for every $(x, y) \in \operatorname{dom} \partial H$. We only need to prove $\bar{y} \in \partial f(\bar{x})$ and $\bar{x} \in \partial g^*(\bar{y})$. To do so, we first observe from (10) that the equation $\nabla F_\lambda(\bar{\xi}) = 0$ is equivalent to

$$\operatorname{prox}_{\lambda f}(\bar{\xi}) = \operatorname{prox}_{\lambda g}(\bar{\xi}). \tag{24}$$

Then, by applying Fact 2.1 (i), we deduce that

$$\frac{1}{\lambda}\left(\bar{\xi} - \operatorname{prox}_{\lambda g}(\bar{\xi})\right) \in \partial f\left(\operatorname{prox}_{\lambda f}(\bar{\xi})\right).$$

This combining with Fact 2.1 (iii) implies that $\operatorname{prox}_{g^*/\lambda}(\bar{\xi}/\lambda) \in \partial f\left(\operatorname{prox}_{\lambda f}(\bar{\xi})\right)$, which means $\bar{y} \in \partial f(\bar{x})$ from (22). In addition, together (24) with (8), one has that $\operatorname{prox}_{\lambda f}(\bar{\xi}) \in \partial g^*\left(\operatorname{prox}_{g^*/\lambda}(\bar{\xi}/\lambda)\right)$, which means $\bar{x} \in \partial g^*(\bar{y})$ from (22). This completes the claim.

Next, since H is a PLK function with exponent α at $(\bar{x}, \bar{y})$, there exist $c > 0$ and $\epsilon > 0$ such that

$$\operatorname{dist}^{\frac{1}{\alpha}}(0, \partial H(x, y)) \geq c(H(x, y) - H(\bar{x}, \bar{y})) \tag{25}$$

holds whenever $\|(x, y) - (\bar{x}, \bar{y})\| \leq (1 + \lambda)/\lambda\epsilon$, $(x, y) \in \operatorname{dom}\partial H = \operatorname{dom}\partial f \times \operatorname{dom}\partial g^*$, and $H(x, y) < H(\bar{x}, \bar{y}) + \epsilon$. Moreover, observe that, for any ξ, the definitions of Proximal mapping and Moreau envelope in (7) imply that

$$\begin{aligned} f(\operatorname{prox}_{\lambda f}(\xi)) &= e_\lambda f(\xi) - 1/(2\lambda)\|\xi - \operatorname{prox}_{\lambda f}(\xi)\|^2, \\ g^*\left(\operatorname{prox}_{g^*/\lambda}(\xi/\lambda)\right) &= e_{1/\lambda} g^*(\xi/\lambda) - \lambda/2\|\xi/\lambda - \operatorname{prox}_{g^*/\lambda}(\xi/\lambda)\|^2 \\ &= \frac{1}{2\lambda}\|\xi\|^2 - e_\lambda g(\xi) - \frac{\lambda}{2}\|\xi/\lambda - \operatorname{prox}_{g^*/\lambda}(\xi/\lambda)\|^2, \end{aligned} \tag{26}$$

where the last equality follows from Fact 2.1 (iv). We can then sum the relations in (26) and rearrange the terms to obtain

$$\begin{aligned} F_\lambda(\xi) = {} & H(\operatorname{prox}_{\lambda f}(\xi), \operatorname{prox}_{g^*/\lambda}(\xi/\lambda)) + 1/(2\lambda)\|\xi\|^2 - 1/\lambda\langle \xi, \operatorname{prox}_{\lambda f}(\xi) \\ & + \lambda \operatorname{prox}_{g^*/\lambda}(\xi/\lambda)\rangle + 1/(2\lambda)\|\operatorname{prox}_{\lambda f}(\xi) + \lambda \operatorname{prox}_{g^*/\lambda}(\xi/\lambda)\|^2 \\ = {} & H(\operatorname{prox}_{\lambda f}(\xi), \operatorname{prox}_{g^*/\lambda}(\xi/\lambda)) + 1/(2\lambda)\|\operatorname{prox}_{\lambda g}(\xi) - \operatorname{prox}_{\lambda f}(\xi)\|^2, \end{aligned}$$

where the last equality follows from Fact 2.1 (iii). This means that

$$F_\lambda(\xi) = H(\operatorname{prox}_{\lambda f}(\xi), \operatorname{prox}_{g^*/\lambda}(\xi/\lambda)) + 1/(2\lambda)\|\operatorname{prox}_{\lambda g}(\xi) - \operatorname{prox}_{\lambda f}(\xi)\|^2. \tag{27}$$

Furthermore, for any ξ, it holds by letting $\tilde{\xi} = 1/\lambda(\xi - \operatorname{prox}_{\lambda g}(\xi))$ that

$$\begin{aligned} & (\lambda + 1)/\lambda\|\operatorname{prox}_{\lambda f}(\xi) - \operatorname{prox}_{\lambda g}(\xi)\| \\ & \overset{(a)}{\geq} \operatorname{dist}\left(\tilde{\xi}, \partial f(\operatorname{prox}_{\lambda f}(\xi))\right) + \operatorname{dist}\left(\operatorname{prox}_{\lambda f}(\xi), \partial g^*\left(\tilde{\xi}\right)\right) \\ & \overset{(b)}{\geq} \operatorname{dist}\left(0, \partial H\left(\operatorname{prox}_{\lambda f}(\xi), \tilde{\xi}\right)\right) \overset{(c)}{=} \operatorname{dist}\left(0, \partial H(\operatorname{prox}_{\lambda f}(\xi), \operatorname{prox}_{g^*/\lambda}(\xi/\lambda))\right), \end{aligned} \tag{28}$$

where (a) follows from Fact 2.1 (i), Fact 2.1 (iii) and (8), (b) follows from (23) and (c) holds because of Fact 2.1 (iii).

Consider any ξ with $\|\xi - \bar{\xi}\| \leq \epsilon$ and $F_\lambda(\bar{\xi}) < F_\lambda(\xi) < F_\lambda(\bar{\xi}) + \epsilon$, where ϵ is defined as in (25). We have

$$\|(\operatorname{prox}_{\lambda f}(\xi), \operatorname{prox}_{g^*/\lambda}(\xi/\lambda)) - (\bar{x}, \bar{y})\| \leq (\lambda + 1)/\lambda\|\xi - \bar{\xi}\| \leq (\lambda + 1)/\lambda\epsilon, \tag{29}$$

where the first inequality follows from (22), the triangle inequality and Fact 2.1 (vi). Since $F_\lambda(\bar{\xi}) < F_\lambda(\xi) < F_\lambda(\bar{\xi}) + \epsilon$, one has that

$$H(\operatorname{prox}_{\lambda f}(\xi), \operatorname{prox}_{g^*/\lambda}(\xi/\lambda)) \leq F_\lambda(\xi) < F_\lambda(\bar{\xi}) + \epsilon = H(\bar{x}, \bar{y}) + \epsilon, \tag{30}$$

where the first inequality follows from (27) and the last equality holds because of (22), (24) and (27). Furthermore, since $\nabla F_\lambda(\bar{\xi}) = 0$, by shrinking $\epsilon > 0$ further if necessary, the continuity of ∇F_λ implies that

$$\|\mathrm{prox}_{\lambda f}(\xi) - \mathrm{prox}_{\lambda g}(\xi)\| = \lambda\|\nabla F_\lambda(\xi)\| < 1. \tag{31}$$

Then, using the facts $\mathrm{prox}_{\lambda f}(\xi) \in \mathrm{dom}\partial f$ and $\mathrm{prox}_{g^*/\lambda}(\xi/\lambda) \in \mathrm{dom}\partial g^*$, we deduce that for any ξ satisfying $\|\xi - \bar{\xi}\| \le \epsilon$ and $F_\lambda(\bar{\xi}) < F_\lambda(\xi) < F_\lambda(\bar{\xi}) + \epsilon$,

$$\begin{aligned}
(1+1/\lambda)^{\frac{1}{\alpha}}\|\nabla F_\lambda(\xi)\|^{\frac{1}{\alpha}} &\overset{\text{(a)}}{=} (1+1/\lambda)^{\frac{1}{\alpha}}(1/\lambda)^{\frac{1}{\alpha}}\|\mathrm{prox}_{\lambda f}(\xi) - \mathrm{prox}_{\lambda g}(\xi)\|^{\frac{1}{\alpha}}\\
&\overset{\text{(b)}}{\ge} (1/\lambda)^{\frac{1}{\alpha}}\,\mathrm{dist}^{\frac{1}{\alpha}}\left(0, \partial H(\mathrm{prox}_{\lambda f}(\xi), \mathrm{prox}_{g^*/\lambda}(\xi/\lambda))\right)\\
&\overset{\text{(c)}}{\ge} \bar{c}\left(H(\mathrm{prox}_{\lambda f}(\xi), \mathrm{prox}_{g^*/\lambda}(\xi/\lambda)) - H(\bar{x}, \bar{y})\right)\\
&\overset{\text{(d)}}{=} \bar{c}\left(F_\lambda(\xi) - F_\lambda(\bar{\xi}) - 1/(2\lambda)\|\mathrm{prox}_{\lambda g}(\xi) - \mathrm{prox}_{\lambda f}(\xi)\|^2\right)\\
&\overset{\text{(e)}}{\ge} \bar{c}\left(F_\lambda(\xi) - F_\lambda(\bar{\xi}) - \lambda/2\|\nabla F_\lambda(\xi)\|^{\frac{1}{\alpha}}\right),
\end{aligned}$$

where (a) follows from (10), (b) follows from (28), (c) holds because of (25), (29), (30) and $\bar{c} := c(1/\lambda)^{\frac{1}{\alpha}}$, (d) follows from (27) and (30), and (e) holds because of the inequality $\alpha \ge \frac{1}{2}$ and (31). This completes the proof. □

Proof of Theorem 4.3 According to [23, Lemma 2.1], we only need to prove that H_λ has the PLK exponent of $\alpha \in [\frac{1}{2}, 1)$ at every point $(\bar{x}, \bar{y})$ satisfying $0 \in \partial H_\lambda(\bar{x}, \bar{y})$. In order to do so, we first claim that if $(\bar{x}, \bar{y})$ satisfies $0 \in \partial H_\lambda(\bar{x}, \bar{y})$, then $0 = \nabla F_\lambda(\bar{y})$. In fact, if $0 \in \partial H_\lambda(\bar{x}, \bar{y})$, then from $\partial H_\lambda(x, y) = \begin{bmatrix} 1/\lambda(x-y) + \partial f(x) \\ 1/\lambda(-x + \mathrm{prox}_{\lambda g}(y)) \end{bmatrix}$, we have that $1/\lambda(\bar{y} - \bar{x}) \in \partial f(\bar{x})$ and $\bar{x} = \mathrm{prox}_{\lambda g}(\bar{y})$. The latter implies that $\bar{y} \in \bar{x} + \partial\lambda g(\bar{x})$ by using Fact 2.1 (i). Consequently, we have $1/\lambda(\bar{y} - \bar{x}) \in \partial f(\bar{x}) \cap \partial g(\bar{x})$. Let $\bar{\xi} \in \partial f(\bar{x}) \cap \partial g(\bar{x})$ be such that $\bar{y} - \bar{x} = \lambda\bar{\xi}$. Then we obtain

$$\bar{y} = \mathrm{prox}_{\lambda g}(\bar{y}) + \lambda\bar{\xi} \in \mathrm{prox}_{\lambda g}(\bar{y}) + \lambda\partial f(\bar{x}) = \mathrm{prox}_{\lambda g}(\bar{y}) + \lambda\partial f(\mathrm{prox}_{\lambda g}(\bar{y})),$$

which means $\bar{y} \in \mathrm{prox}_{\lambda g}(\bar{y}) + \lambda\partial f(\mathrm{prox}_{\lambda g}(\bar{y}))$. This is equivalent to saying $\mathrm{prox}_{\lambda g}(\bar{y}) = \mathrm{prox}_{\lambda f}(\bar{y})$. Therefore, we can conclude that $0 = \nabla F_\lambda(\bar{y})$ by using (10), and $H_\lambda(\bar{x}, \bar{y}) = F_\lambda(\bar{y})$.

Let $G(x, y) := f(x) + 1/(2\lambda)\|x-y\|^2 - f(\mathrm{prox}_{\lambda f}(y)) - 1/(2\lambda)\|y - \mathrm{prox}_{\lambda f}(y)\|^2$. Then we have from (7) that

$$0 \le H_\lambda(x, y) - F_\lambda(y) = G(x, y). \tag{32}$$

Now, we claim that

$$G(x, y) \le 1/(2\lambda)\,\mathrm{dist}^2(y - x, \lambda\partial f(x)). \tag{33}$$

In fact, let $\xi \in \partial f(x)$. Then we have

$$G(x, y) \le \langle \xi, x - \mathrm{prox}_{\lambda f}(y)\rangle + 1/(2\lambda)\|x - y\|^2 - 1/(2\lambda)\|y - \mathrm{prox}_{\lambda f}(y)\|^2$$
$$= 1/(2\lambda)\|x - y + \lambda\xi\|^2 - 1/(2\lambda)\|y - \mathrm{prox}_{\lambda f}(y) - \lambda\xi\|^2 \le 1/(2\lambda)\|x - y + \lambda\xi\|^2,$$

where the first inequality follows from the convexity of f and $\xi \in \partial f(x)$, the equality follows from (7). Taking the infimum over $\xi \in \partial f(x)$ in above display, we obtain (33).

Since F_λ is a PLK function with an exponent of α at $\bar{y}$, there exist $\epsilon > 0$ and $c > 0$ such that when $\|y - \bar{y}\| \le \epsilon$ and $0 < F_\lambda(y) - F_\lambda(\bar{y}) < \epsilon$, we have

$$\|\nabla F_\lambda(y)\|^{\frac{1}{\alpha}} \ge c\,(F_\lambda(y) - F_\lambda(\bar{y}))\,. \tag{34}$$

Now, for such ϵ, consider any (x, y) satisfying $\|(x, y) - (\bar{x}, \bar{y})\| \le \epsilon$, $(x, y) \in \mathrm{dom}\,\partial H_\lambda$ and $H_\lambda(\bar{x}, \bar{y}) < H_\lambda(x, y) < H_\lambda(\bar{x}, \bar{y}) + \epsilon$. Then $\|y - \bar{y}\| \le \|(x, y) - (\bar{x}, \bar{y})\| \le \epsilon$ and $F_\lambda \le H_\lambda(x, y) < F_\lambda(\bar{y}) + \epsilon$. Moreover, it holds that

$$\begin{aligned}\|x - \mathrm{prox}_{\lambda f}(y)\| &\le \|x - \bar{x}\| + \|\bar{x} - \mathrm{prox}_{\lambda f}(y)\| \\ &= \|x - \bar{x}\| + \|\mathrm{prox}_{\lambda f}(\bar{y}) - \mathrm{prox}_{\lambda f}(y)\| \le \|x - \bar{x}\| + \|y - \bar{y}\| \le 2\epsilon,\end{aligned} \tag{35}$$

where the equality follows from $\bar{x} = \mathrm{prox}_{\lambda g}(\bar{y})$ and $\mathrm{prox}_{\lambda g}(\bar{y}) = \mathrm{prox}_{\lambda f}(\bar{y})$, the second inequality follows from Fact 2.1 (vi) and the last inequality follows from $\max\{\|x - \bar{x}\|, \|y - \bar{y}\|\} \le \|(x, y) - (\bar{x}, \bar{y})\| \le \epsilon$. By shrinking ϵ again if necessary, the above display with the continuity of f on $\mathrm{dom} f$ implies that

$$\begin{aligned}G(x, y) \le\, & f(x) - f(\mathrm{prox}_{\lambda f}(y)) \\ & + 1/(2\lambda)\|x - \mathrm{prox}_{\lambda f}(y)\|(\|x\| + 2\|y\| + \|\mathrm{prox}_{\lambda f}(y)\|) \le 1,\end{aligned} \tag{36}$$

where the first inequality follows from $\|a\|^2 - \|b\|^2 \le \|a - b\|(\|a\| + \|b\|)$ for every pair $a, b \in \mathbb{R}^n$ and the triangle inequality, and the last inequality follows from the continuity of $\mathrm{prox}_{\lambda f}$ and (35). Therefore, for any such (x, y), one has that

$$\begin{aligned}2\lambda^{\frac{1}{\alpha}}\,\mathrm{dist}^{\frac{1}{\alpha}}(0, \partial H_\lambda(x, y)) &\ge \mathrm{dist}^{\frac{1}{\alpha}}\left(0, \begin{bmatrix}(I + \lambda\partial f)(x) - y \\ -x + \mathrm{prox}_{\lambda g}(y)\end{bmatrix}\right) + \mathrm{dist}^{\frac{1}{\alpha}}(y - x, \lambda\partial f(x)) \\ &\overset{\text{(a)}}{\ge} \tilde{c}\left(\mathrm{dist}^{\frac{1}{\alpha}}(y, (I + \lambda\partial f)(x)) + \|x - \mathrm{prox}_{\lambda g}(y)\|^{\frac{1}{\alpha}}\right) + (2\lambda)^{\frac{1}{2\alpha}} G^{\frac{1}{2\alpha}}(x, y) \\ &\overset{\text{(b)}}{\ge} \bar{c}\left(\mathrm{dist}^{\frac{1}{\alpha}}(x, (I + \lambda\partial f)^{-1}(y)) + \|x - \mathrm{prox}_{\lambda g}(y)\|^{\frac{1}{\alpha}}\right) + (2\lambda)^{\frac{1}{2\alpha}} G^{\frac{1}{2\alpha}}(x, y) \\ &\overset{\text{(c)}}{\ge} \bar{c}\frac{1}{2^{\frac{1}{\alpha}-1}}\|\mathrm{prox}_{\lambda f}(y) - \mathrm{prox}_{\lambda g}(y)\|^{\frac{1}{\alpha}} + (2\lambda)^{\frac{1}{2\alpha}} G^{\frac{1}{2\alpha}}(x, y) \\ &\overset{\text{(d)}}{\ge} \frac{\bar{c}}{2}\|\mathrm{prox}_{\lambda f}(y) - \mathrm{prox}_{\lambda g}(y)\|^{\frac{1}{\alpha}} + (2\lambda)^{\frac{1}{2\alpha}} G(x, y)\end{aligned}$$

$$
\begin{aligned}
&\overset{(e)}{\geq} c^* \left((1/\lambda)^{\frac{1}{\alpha}} \|\mathrm{prox}_{\lambda f}(y) - \mathrm{prox}_{\lambda g}(y)\|^{\frac{1}{\alpha}} + cG(x, y) \right) \\
&\overset{(f)}{\geq} c^* c \left(F_\lambda(y) - F_\lambda(\bar{y}) + G(x, y) \right) \overset{(g)}{=} \hat{c} \left(H_\lambda(x, y) - H_\lambda(\bar{x}, \bar{y}) \right),
\end{aligned}
$$

where (a) holds for some $\tilde{c} > 0$ because of [23, Lemma 2.2], (33) and $\frac{1}{2} \leq \alpha < 1$, (b) holds for some $\bar{c} > 0$ by using the metric regularity of $I + \lambda\partial f$ (thanks to [28, Theorem 1.49] and the Lipschitz continuity of $(I + \lambda\partial f)^{-1}$), (c) follows from $\alpha < 1$, the convexity of function $x^{\frac{1}{\alpha}}$ and (7), (d) follows from $\frac{1}{2} \leq \alpha$ and (36), (e) follows from the notation $c^* := \min\{\frac{\bar{c}}{2}\lambda^{1/\alpha}, \frac{1}{c}(2\lambda)^{\frac{1}{2\alpha}}\}$ with $\lambda, c > 0$ and the nonnegativity of $G(x, y)$ as in (32), (f) follows from (34), and (g) follows from (32), $H_\lambda(\bar{x}, \bar{y}) = F_\lambda(\bar{y})$ and $\hat{c} := c^* c$. This completes the proof. □

It is worth noting that the continuity of H_λ plays a crucial role in ensuring that the PLK exponent of $(1/2, 1)$ is inherited from F_λ to H_λ in the above proof. Moreover, the PLK exponent being $1/2$ is closely related to the linear convergence rate, which is a desirable convergence rate for many first-order optimization methods; see, for example, [5], [6] and references therein. Therefore, when our primary concern is the PLK exponent being $1/2$, we can drop the continuity of f in Theorem 4.3 to accommodate the more general lower semicontinuity hypotheses commonly seen in DC programming and lead to the following result.

Corollary 4.1 *Suppose f and g are proper closed and convex. If F_λ is a PLK function with an exponent of $\alpha = \frac{1}{2}$. Then H_λ is also a PLK function with the same exponent α.*

Proof Without considering $H_\lambda(x, y) - F_\lambda(y) \leq 1$ whenever (x, y) is sufficiently closed to $(\bar{x}, \bar{y})$ with $0 \in \partial H_\lambda(\bar{x}, \bar{y})$, we can follow a similar argument as in the proof of Theorem 4.3 to complete this proof. The details are omitted for brevity. □

In the remainder of this section, we give the proof of Theorem 4.4. In order to do so, we require the following fact regarding H_λ in (3) and F in (1). This fact is a direct consequence by using the equivalent representation of H_λ, namely $H_\lambda = f - (-e_\lambda(-e_\lambda g))$, and [36, Ex. 1.44].

Fact 4.1 *Consider H_λ in (3) and F in (1). One has that for every $x \in \mathbb{R}^n$, $\inf_y\{H_\lambda(x, y)\} = F(x)$.*

Proof of Theorem 4.4 Let $Y(x) := \arg\min_y\{H_\lambda(x, y)\}$. Since f is proper and closed, we have that $f(x) + 1/(2\lambda)\|x - y\|^2$ is closed and proper with respect to variable x for any fixed y. This together with the differentiability of g implies that H_λ is also a proper and closed function.

For arbitrary $\bar{x} \in \mathrm{dom}\partial F$, we have that $Y(\bar{x}) = \{\bar{x} + \lambda\nabla g(\bar{x})\}$ by using the differentiability of g again. Let $\Omega(\bar{x}) := \{\bar{x}\} \times Y(\bar{x})$. Then $\Omega(\bar{x}) \subseteq \mathrm{dom}\partial H_\lambda$ and the mapping Ω is upper semicontinuous on $\mathrm{dom}\partial F$.

Since $H_\lambda(\bar{x}, \bar{x} + \lambda\nabla g(\bar{x})) \equiv F(\bar{x})$ for every $\bar{x} \in \mathrm{dom}\partial F$, we have from the PLK property of H_λ and [44, Lemma 2.2] that there exist $\nu, c > 0$ and $a \in [0, 1)$ such that

$$
\mathrm{dist}(0, \partial H_\lambda(x, y)) \geq c(H_\lambda(x, y) - F(\bar{x}))^\alpha \tag{37}
$$

holds for every $(x, y) \in \mathrm{dom}\,\partial H_\lambda$ satisfying $F(\bar{x}) < H_\lambda(x, y) < F(\bar{x}) + a$ and $\mathrm{dist}((x, y), \Omega(\bar{x})) < \nu$.

Now, from the upper semicontinuity of Ω, for such ν, there exists $\epsilon \in (0, \frac{1}{2}\nu)$ such that for every $x \in B(\bar{x}, \epsilon) \cap \mathrm{dom}\partial F$, we have $Y(x) \subseteq Y(\bar{x}) + \frac{1}{2}\nu B(0, 1)$ and thus $\mathrm{dist}((x, y), \Omega(\bar{x})) \leq \|x - \bar{x}\| + \|y - \bar{x} - \lambda\nabla g(\bar{x})\| < \nu$ whenever $y \in Y(x)$. If we further require that $F(\bar{x}) < F(x) < F(\bar{x}) + a$, then for every $x \in B(\bar{x}, \epsilon) \cap \mathrm{dom}\partial F$ and $y \in Y(x)$, by shrinking $\nu > 0$ if necessary, we have that

$$F(\bar{x}) < H_\lambda(x, y) < F(\bar{x}) + a.$$

Thus, for any such $x \in B(\bar{x}, \epsilon) \cap \mathrm{dom}\partial F$ and $y = x + \lambda\nabla g(x) \in Y(x)$, one has

$$\begin{aligned}\partial H_\lambda(x, y) &= \begin{bmatrix} 1/\lambda(x - y) + \partial f(x) \\ 1/\lambda(-x + \mathrm{prox}_{\lambda g}(y)) \end{bmatrix} = \begin{bmatrix} -\nabla g(x) + \partial f(x) \\ 1/\lambda(-x + \mathrm{prox}_{\lambda g}(y)) \end{bmatrix} \\ &= \begin{bmatrix} -\nabla g(x) + \partial f(x) \\ 0 \end{bmatrix},\end{aligned}$$

where the last equality follows from Fact 2.1 (i). This implies that

$$\begin{aligned}\mathrm{dist}(0, \partial F(x)) = \mathrm{dist}\left(0, \begin{bmatrix} \partial f(x) - \nabla g(x) \\ 0 \end{bmatrix}\right) &= \mathrm{dist}(0, \partial H_\lambda(x, y)) \\ &\geq c(H_\lambda(x, y) - F(\bar{x}))^\alpha \geq c(F(x) - F(\bar{x}))^\alpha,\end{aligned}$$

where the first inequality follows from (37), and the last inequality follows from Fact 4.1. This completes the proof. □

5 Deducing New PLK Exponent for the Least Squares Model with Logistic Penalty

In this section, we use the equivalence result of the PLK exponent discussed in Section 4 to derive the new PLK exponent of a specific DC function arising in compressed sensing. The key contributions are Theorem 4.3 and Theorem 4.4 in this equivalence result, which offer a new approach to deducing the PLK exponent of a DC function from that of its DME function. Note that the DME function is Lipschitz differentiable. We begin by exploring the PLK exponent of functions that belong to $C^{1,1}$ under the nonsingularity of their generalized Hessian matrices.

5.1 The PLK Exponent of Lipschitz Differentiable Functions

In this subsection, for any $x, y \in \mathbb{R}^n$, we use the notation $[x, y]$ to denote the set $\{tx + (1 - t)y : t \in [0, 1]\}$, and (x, y) to denote the set $\{tx + (1 - t)y : t \in (0, 1)\}$. It is worth noting that these notations differ from those used in previous sections which are used to denoting vector pairs.

Now, we present a theorem that provides a sufficient condition for a Lipschitz differentiable function to be a PLK function with exponent $1/2$.

Theorem 5.1 (The PLK exponent for functions that belong to $C^{1,1}$) *Let $h \in C^{1,1}$. If all elements of $\partial^2 h(x)$ are nonsingular at every critical point x of h, then h is a PLK function with an exponent of* $1/2$.

Proof Take any critical point $\bar{x}$ of h, i.e. $\nabla h(\bar{x}) = 0$. Since $\partial^2 h$ is locally bounded; see [18, Subsection 2.1] for the discussion, there exist $\epsilon > 0$ and $K > 0$ such that

$$\sup\{\|V\| : V \in \partial^2 h(y),\ y \in B(\bar{x}, \epsilon)\} \le K. \tag{38}$$

Let $x \in B(\bar{x}, \epsilon)$ with $h(x) > h(\bar{x})$. Then it follows from [18, Theorem 2.3], (38) and $\nabla h(\bar{x}) = 0$ that there exist $c \in (\bar{x}, x)$ and $\bar{V} \in \partial^2 h(c)$ such that

$$h(x) - h(\bar{x}) = 1/2(x - \bar{x})^T \bar{V}(x - \bar{x}) \le 1/2K\|x - \bar{x}\|^2. \tag{39}$$

Moreover, according to [12, Theorem 7.1.1], we have from the nonsingularity of $\partial^2 h(\bar{x})$ that ∇h admits a Lipschitz inverse function ∇h^{-1} over the neighborhood $B(\bar{x}, \epsilon)$ by shrinking ϵ if necessary. Take $\bar{\delta} > 0$ as the Lipschitz modulus. Then we have for arbitrary $x, y \in B(\bar{x}, \epsilon)$ that

$$\|x - y\| = \|\nabla h^{-1}(\nabla h(x)) - \nabla h^{-1}(\nabla h(y))\| \le \bar{\delta}\|\nabla h(x) - \nabla h(y)\|.$$

The above display together with (39) implies

$$\|\nabla h(x) - \nabla h(\bar{x})\| \ge c(h(x) - h(\bar{x}))^{1/2},$$

where $c := \bar{\delta}^{-1}\sqrt{2K}$. By using [23, Lemma 2.1], we obtain the conclusion. □

Remark 5.1 Consider Theorem 5.1. Notably, the assumption in this theorem coincides with the hypotheses of Clarke's inverse function theorem [12, Theorem 7.1.1] at the critical point. However, verifying this assumption can be challenging in general cases; see [14, Section 2] and [8, Section 1] for a detailed discussion. To overcome this difficulty, we utilize the second-order expansion of h as stated in [18, Theorem 2.3]. This allows us to follow the argument in the proof of [39, Lemma 4.16] and establish that if every critical point $\bar{x}$ of h in the aforementioned theorem is a locally strong minimizer, meaning that there exist positive numbers $\tilde{\mu}$ and $\tilde{\epsilon}$ such that for all $x \in B(\bar{x}, \tilde{\epsilon})$, $h(x) - h(\bar{x}) \ge \tilde{\mu}\|x - \bar{x}\|^2$, then every matrix in $\partial^2 h(\bar{x})$ is nonsingular. Moreover, the converse also holds true.

Recall that for any $\lambda > 0$, the DME function F_λ defined in (2) belongs to $C^{1,1}$. The following corollary provides a new way for deducing the PLK exponent of F in (1) by verifying the nonsingularity of generalized Hessian matrices at every critical point of its DME function F_λ for any fixed $\lambda > 0$.

Corollary 5.1 *Consider* (1) *and* (2). *Suppose for any fixed* $\lambda > 0$ *that*

(i) f is continuous on $\operatorname{dom} f$ *and g is differentiable;*
(ii) all elements of $\partial^2 F_\lambda(x)$ are nonsingular at every critical point x of F_λ.

Then, F in (1) *is a PLK function with exponent of* $1/2$.

Proof The claim follows directly from Theorems 4.3, 4.4 and 5.1. □

5.2 The PLK Exponent of a Specific Example

In this second part of Section 5, we explore the PLK exponent of function F in (5) with matrix $A = I$ as an example of applying Corollary 5.1. It can be verified that F satisfies the item (i) of Corollary 5.1. In order to simplify the verification that F satisfies the item (ii) of Corollary 5.1, we set $\lambda = 1$, which means that we now verify the nonsingularity of the generalized Hessian matrix of $F_1 = e_1 f - e_1 g$ at each of its critical points.

Note that for every $x \in \operatorname{dom} \partial^2 F_1$, we have from Fact 2.1 (vi), [12, Proposition 2.3.3] and [12, Theorem 2.5.1] that $\partial^2 F_1(x) \subseteq \partial_C \operatorname{prox}_g(x) - \partial_C \operatorname{prox}_f(x)$, where f and g are defined as in (5). We begin by deriving the expression for $\partial_C \operatorname{prox}_g(x) - \partial_C \operatorname{prox}_f(x)$.

Lemma 5.1 *Consider* (5) *with $A = I$. We have that*

(i) The proximal mapping of f at $x \in$ $\operatorname{dom} f$ *is*

$$\operatorname{prox}_f(x) = \frac{1}{2}\operatorname{sign}(b + x) \cdot (\|b + x\|_1 - \mu/\sigma)_+ .$$

(ii) For every $i = 1, 2, \cdots, n$, the i-th element of $\operatorname{prox}_g(x)$ *at $x \in$* $\operatorname{dom} g$ *is*

$$\left(\operatorname{prox}_g(x)\right)_i = \operatorname{sign}(x_i)\frac{\sigma|x_i| - \sigma^2 - \mu + \sqrt{\left(\sigma|x_i| - \sigma^2 - \mu\right)^2 + 4\sigma^3|x_i|}}{2\sigma}.$$

(iii) For every $i = 1, 2, \cdots, n$, we have that

$$\begin{aligned}&\left(\partial_C \operatorname{prox}_g(x)\right)_{ii} - \left(\partial_C \operatorname{prox}_f(x)\right)_{ii}\\ &= \begin{cases} \frac{\sigma^2-\mu+\sigma|x_i|}{2\sqrt{(\sigma^2+\mu-\sigma|x_i|)^2+4\sigma^3|x_i|}} & \text{if } |b_i + x_i| > \frac{\mu}{\sigma},\\ \frac{1}{2} + \frac{\sigma^2-\mu+\sigma|x_i|}{2\sqrt{(\sigma^2+\mu-\sigma|x_i|)^2+4\sigma^3|x_i|}} & \text{if } |b_i + x_i| < \frac{\mu}{\sigma},\\ \frac{1}{2} + \frac{\sigma^2-\mu+\sigma|x_i|}{2\sqrt{(\sigma^2+\mu-\sigma|x_i|)^2+4\sigma^3|x_i|}} - \left[0, \frac{1}{2}\right] & \text{if } |b_i + x_i| = \frac{\mu}{\sigma}.\end{cases}\end{aligned}$$

Proof (i): From the definition of f in (5), we have that for arbitrary $x \in \mathbb{R}^n$,

$$\operatorname{prox}_f(x) = \arg\min_y \left\{\frac{\mu}{\sigma}\|y\|_1 + \left\|y - \frac{b+x}{2}\right\|^2\right\} = \operatorname{prox}_{\frac{\mu}{2\sigma}\|\cdot\|_1}\left(\frac{b+x}{2}\right).$$

Then the conclusion holds from the proximal mapping of scaled ℓ_1-norm in [3, Example 6.8].

(ii): As for the computation of $\mathrm{prox}_g(x)$, we have from the definition of g in (5) that for arbitrary $x \in \mathbb{R}^n$ and every $i = 1, 2, \cdots, n$,

$$\mathrm{prox}_g(x) = \arg\min_y \left\{ \sum_{i=1}^{n} \left(\frac{\mu}{\sigma}|y_i| - \mu\log\left(1 + \frac{|y_i|}{\sigma}\right) + \frac{1}{2}(x_i - y_i)^2 \right) \right\}.$$

Note that for every i, $\frac{\mu}{\sigma}|y_i| - \mu\log\left(1 + \frac{|y_i|}{\sigma}\right) + \frac{1}{2}(x_i - y_i)^2$ is nonnegative with related to y_i for any fixed point x_i. We have that

$$\min_y \left(g(y) + \frac{1}{2}\|x - y\|^2 \right) = \sum_{i=1}^{n} \min_{y_i} \left(\frac{\mu}{\sigma}|y_i| - \mu\log\left(1 + \frac{|y_i|}{\sigma}\right) + \frac{1}{2}(x_i - y_i)^2 \right),$$

and therefore,

$$(\mathrm{prox}_g(x))_i = \arg\min_{y_i} \left\{ \frac{\mu}{\sigma}|y_i| - \mu\log\left(1 + \frac{|y_i|}{\sigma}\right) + \frac{1}{2}(x_i - y_i)^2 \right\}. \tag{40}$$

Taking the first order optimality condition on the right-hand side of (40), one has that

$$\left(1 + \frac{\mu}{\sigma(\sigma + |\left(\mathrm{prox}_g(x)\right)_i|)}\right)\left(\mathrm{prox}_g(x)\right)_i = x_i. \tag{41}$$

By inversely solving $\left(\mathrm{prox}_g(x)\right)_i$ from (41), we obtain the conclusion.

(iii): Taking the generalized Jacobian of $\mathrm{prox}_{\frac{\mu}{2\sigma}\|\cdot\|_1}$ as in [39, Subsection 6.3], we have from Lemma 5.1 (i) that for every $i = 1, 2, \cdots, n$,

$$\left(\partial_C \,\mathrm{prox}_f(x)\right)_{ii} = \begin{cases} \frac{1}{2} & \text{if } |b_i + x_i| > \frac{\mu}{\sigma}, \\ 0 & \text{if } |b_i + x_i| < \frac{\mu}{\sigma}, \\ [0, \frac{1}{2}] & \text{if } |b_i + x_i| = \frac{\mu}{\sigma}. \end{cases} \tag{42}$$

After a further calculation on $\left(\mathrm{prox}_g(x)\right)_i$, we have that

$$\left(\partial_C \,\mathrm{prox}_g(x)\right)_{ii} = \frac{1}{2} + \frac{\sigma^2 - \mu + \sigma|x_i|}{2\sqrt{(\sigma^2 + \mu - \sigma|x_i|)^2 + 4\sigma^3|x_i|}}. \tag{43}$$

Therefore, it follows from (42) and (43) that we complete the proof. □

Remark 5.2 Combining the definition of DME function F_1 of F in (5) with Lemma 5.1 (i) and (ii), we can conclude that F_1 is a separable function.

Now we are ready to derive the PLK exponent of F_1 in the following theorem.

Theorem 5.2 *Consider F in (5) with $A = I$. Suppose that the vector b and two parameters μ and σ satisfy that for every $i = 1, 2, \cdots, n$,*

$$|b_i| \notin \{2\sqrt{\mu} - \sigma, \sigma + 2\mu/\sigma - 2\sqrt{\mu}\}. \tag{44}$$

Then the PLK exponent of the DME function F_1, as defined in (2), of F is $\frac{1}{2}$.

Proof Let x^* be the critical point of F_1, which means that $\mathrm{prox}_f(x^*) = \mathrm{prox}_g(x^*)$. According to [23, Lemma 2.1], we only need to prove that the PLK exponent of F_1 at x^* is $\frac{1}{2}$ under our assumption. Since F_1 is separable, we examine each dimension to determine its PLK exponent by using [23, Theorem 3.3]. To simplify the subsequent analysis, we use x, x^* and b to denote x_i, x_i^* and b_i, respectively, for every $i = 1, 2, \cdots, n$. We now divide our discussion into three cases:

Case 1: $|b + x^*| < \frac{\mu}{\sigma}$. In this case, we note that

$$1 > \frac{\sigma^2 - \mu + \sigma|x^*|}{\sqrt{(\sigma^2 + \mu - \sigma|x^*|)^2 + 4\sigma^3|x^*|}} > -1.$$

This together with Lemma 5.1 (iii) implies that $\partial^2 F_1(x^*) > 0$.

Case 2: $|b+x^*| > \frac{\mu}{\sigma}$. In this case, we have from Lemma 5.1 (iii) and the assumption $\mu \gg \sigma$ that if $x^* = 0$, then $0 \notin \partial^2 F_1(x^*)$. If $x^* \neq 0$, then based on (44), we first consider the subcase: $b + x^* > \frac{\mu}{\sigma}$. In this subcase, if $x^* > 0$, then for every $x \in \mathbb{R}$, we have from

$$F_1'(x) = \mathrm{prox}_g(x) - \mathrm{prox}_f(x) = \frac{-\sigma^2 - \sigma b + \sqrt{(\sigma x + \sigma^2 - \mu)^2 + 4\sigma^2\mu}}{2\sigma},$$

that the critical point x^* exists only if $\sigma + b > 2\sqrt{\mu}$ holds. Under this condition, one has $\sigma x^* = \pm\sqrt{(\sigma^2 + \sigma b)^2 - 4\sigma^2\mu} + \mu - \sigma^2$, which implies that $0 \notin \partial^2 F_1(x^*)$ by using Lemma 5.1 (iii) and (44). If $x^* < 0$, then for every $x \in \mathbb{R}$, we have from

$$F_1'(x) = \mathrm{prox}_g(x) - \mathrm{prox}_f(x) = \frac{\sigma^2 - \sigma b + 2\mu - \sqrt{(\sigma x - \sigma^2 + \mu)^2 + 4\sigma^2\mu}}{2\sigma}$$

that the critical point x^* exists only if $\sigma^2 - \sigma b + 2\mu > 2\sigma\sqrt{\mu}$ holds. In this context, one has $\sigma x^* = \pm\sqrt{(\sigma^2 - \sigma b + 2\mu)^2 - 4\sigma^2\mu} - \mu + \sigma^2$, which implies $0 \notin \partial^2 F_1(x^*)$ because of Lemma 5.1 (iii) and (44).

Similarly, for the subcase $b + x^* < -\frac{\mu}{\sigma}$, we can deduce that $0 \notin \partial^2 F_1(x^*)$ from Lemma 5.1 (iii) and (44), regardless of the sign of x^*. Therefore, the PLK exponent of F_1 at x^* is $\frac{1}{2}$ whenever $|b + x^*| > \frac{\mu}{\sigma}$.

Case 3: $|b + x^*| = \frac{\mu}{\sigma}$. In this case, we can deduce from (10) and Lemma 5.1 (i), (ii) that $\mathrm{prox}_g(x^*) = 0$. This implies that $x^* = 0$ and $|b| = \mu/\sigma$. To complete the proof, in the followings, we rely on the definition of the PLK exponent in Definition 2.2 .

First, let us consider the subcase that $b = \frac{\mu}{\sigma}$. In this subcase, if $x \to 0^+$, then according to Lemma 5.1 (i), (ii), the proximal mapping of f and g at x are respectively

$$\operatorname{prox}_f(x) = \frac{x}{2} \text{ and } \operatorname{prox}_g(x) = \frac{\sigma x - \sigma^2 - \mu + \sqrt{\left(\sigma x - \sigma^2 - \mu\right)^2 + 4\sigma^3 x}}{2\sigma}. \tag{45}$$

The first equality in (45) implies that

$$\begin{aligned} F_1(x) - F_1(0) &= \frac{x^2}{4} - \frac{\mu}{\sigma} \operatorname{prox}_g(x) + \mu \log(1 + \frac{\operatorname{prox}_g(x)}{\sigma}) - \frac{(x - \operatorname{prox}_g(x))^2}{2} \\ &\leq \frac{x^2}{4} - \frac{\mu \operatorname{prox}_g^2(x)}{2\sigma^2} + \frac{\mu \operatorname{prox}_g^3(x)}{3\sigma^3} - \frac{1}{2}(x - \operatorname{prox}_g(x))^2. \end{aligned} \tag{46}$$

Let

$$\iota(x) := \sigma x - \sigma^2 - \mu + \sqrt{\left(\sigma x - \sigma^2 - \mu\right)^2 + 4\sigma^3 x}$$

and

$$\bar{\iota}(x) := \sqrt{\left(\sigma x - \sigma^2 - \mu\right)^2 + 4\sigma^3 x} - (\sigma x - \sigma^2 - \mu).$$

Then one has $\lim_{x \to 0^+} \bar{\iota}(x) = 2(\sigma^2 + \mu)$, which means that for sufficiently small $x > 0$, we have that $\bar{\iota}(x) > 0$ and $\iota(x) = 4\sigma^3 x / \bar{\iota}(x)$. This together with (46) implies that

$$\begin{aligned} F_1(x) - F_1(0) &\leq \frac{6x^2\sigma^6 - 3\sigma^2 \mu \iota^2(x) + \mu \iota^3(x) - 3\sigma^4 (2\sigma x - \iota(x))^2}{24\sigma^6} \\ &\leq \frac{\left(-6\sigma^6 \left(\bar{\iota}^3(x) + 8\sigma^2 \mu \bar{\iota}(x) - 8\sigma^2 \bar{\iota}^2(x) + 8\sigma^4 \bar{\iota}(x)\right)\right) x^2 + 64\sigma^9 \mu x^3}{24\sigma^6 \bar{\iota}^3(x)}, \end{aligned}$$

where the first inequality follows from $\operatorname{prox}_g(x) = \frac{\iota(x)}{2\sigma}$ and the second equality in (45). Note that for $x > 0$ sufficiently close to 0, the inequality $\bar{\iota}^3(x) + 8\sigma^2 \mu \bar{\iota}(x) - 8\sigma^2 \bar{\iota}^2(x) + 8\sigma^4 \bar{\iota}(x) < 0$ holds due to $\mu \gg \sigma$. Therefore $F_1(x) - F_1(0) < 0$.

If on the other hand that $x \to 0^-$, then $0 < b + x < \frac{\mu}{\sigma}$ holds for $x < 0$ sufficiently close to 0. Moreover, we have from Lemma 5.1 (i) and (ii) that

$$\operatorname{prox}_f(x) = 0 \text{ and } \operatorname{prox}_g(x) = \frac{\sigma x + \sigma^2 + \mu - \sqrt{\left(\sigma x + \sigma^2 + \mu\right)^2 - 4\sigma^3 x}}{2\sigma} < 0. \tag{47}$$

The first equality in (47) means that

$$\begin{aligned} F_1(x) - F_1(0) &= \frac{x^2}{2} + \frac{\mu}{\sigma}\operatorname{prox}_g(x) + \mu\log\Big(1 - \frac{\operatorname{prox}_g(x)}{\sigma}\Big) - \frac{1}{2}(x - \operatorname{prox}_g(x))^2 \\ &\le \frac{x^2}{2} - \frac{\mu\operatorname{prox}_g^2(x)}{2\sigma^2} - \frac{\mu\operatorname{prox}_g^3(x)}{3\sigma^3} - \frac{1}{2}(x - \operatorname{prox}_g(x))^2. \end{aligned} \tag{48}$$

Let

$$\kappa(x) := \sigma x + \sigma^2 + \mu - \sqrt{\left(\sigma x + \sigma^2 + \mu\right)^2 - 4\sigma^3 x}$$

and

$$\bar{\kappa}(x) := \sqrt{\left(\sigma x + \sigma^2 + \mu\right)^2 - 4\sigma^3 x} + \sigma x + \sigma^2 + \mu.$$

Then one has that $\operatorname{prox}_g(x) = \frac{\kappa(x)}{2\sigma}$. Furthermore, combining (47) with (10), we obtain $F_1'(x) = \operatorname{prox}_g(x) - \operatorname{prox}_f(x) = \operatorname{prox}_g(x)$. This together with (48) implies that

$$\frac{F_1'^2(x)}{F_1(x) - F_1(0)} \ge 6\sigma^4 \frac{\kappa(x)}{-3\sigma^2\mu\kappa(x) - \mu\kappa^2(x) + 12x\sigma^5 - 3\sigma^4\kappa(x)}. \tag{49}$$

Since $\lim_{x\to 0^-}\bar{\kappa}(x) = 2(\sigma^2 + \mu)$, it follows that for $x < 0$ sufficiently close to 0, we have $\kappa(x) = 4\sigma^3 x/\bar{\kappa}(x)$. This combining with (49) implies that

$$\begin{aligned} \frac{F_1'^2(x)}{F_1(x) - F_1(0)} &\ge \frac{2\sigma^{10}\bar{\kappa}(x)x^2}{\left(-\mu\sigma^8\bar{\kappa}(x) + \sigma^8\bar{\kappa}^2(x) - \sigma^{10}\bar{\kappa}(x)\right)x^2 - \frac{4}{3}\mu\sigma^9 x^3} \\ &= \frac{2\sigma^{10}\bar{\kappa}(x)}{-\mu\sigma^8\bar{\kappa}(x) + \sigma^8\bar{\kappa}^2(x) - \sigma^{10}\bar{\kappa}(x) - \frac{4}{3}\mu\sigma^9 x} \\ &= \frac{2\sigma^{10}}{-\mu\sigma^8 + \sigma^8\bar{\kappa}(x) - \sigma^{10} - \frac{4}{3\bar{\kappa}(x)}\mu\sigma^9 x} > 0, \end{aligned}$$

where the last inequality follows from $\mu \gg \sigma$.

Therefore, when $b = \frac{\mu}{\sigma}$, we have that the PLK exponent of F_1 at $x^* = 0$ is $\frac{1}{2}$. Similarly, the same result holds when $b = -\frac{\mu}{\sigma}$. Consequently, the PLK exponent of F_1 at the unique critical point $x^* = 0$ is $\frac{1}{2}$ whenever $|b| = \frac{\mu}{\sigma}$, thus completing the proof. □

6 Conclusions

This paper establishes the equivalence for the Polyak-Łojasiewicz-Kurdyka (PLK) exponent between the difference-of-convex (DC) function F in (1) and its difference-of-Moreau-envelope (DME) function F_λ in (2). This equivalence is demonstrated through the use of two potential functions, H in (4) and H_λ in (3), which have been previously utilized in DC programming research. As shown in this paper, the equivalence not only confirms the global convergence and the linear or sublinear convergence rates for recently developed DME-specific algorithms, thereby reinforcing the rationale behind their development and design, but also introduces a approach to uncovering the previously unknown PLK exponent of certain DC functions. Here, we show as an example that the PLK exponent of a nonconvex compressed sensing model (5) that incorporates a logistic penalty term is 1/2, which was previously unknown.

Acknowledgements The author would like to thank Ting Kei Pong in the Hong Kong Polytechnic University, Ying Lin in Hong Kong University, and Xiaozhou Wang in South China Normal University for their inspiring discussions and valuable suggestions regarding this paper sincerely, particularly for Ting Kei Pong's original manuscript for Theorem 4.2 in this paper. Moreover, I must not forget to express my gratitude to all the reviewers for their profound insights and enlightening comments, which have greatly enriched this paper.

Funding: This work is supported partly by the funding of Visual Computing and Virtual Reality Key Laboratory of Sichuan Province under Grant SCVCVR2023.03VS.

Declarations

Competing interests Not Applicable.

References

1. Alvarado, A., Scutari, G., Pang, J.-S.: A new decomposition method for multiuser DC-programming and its applications. IEEE Trans. Signal Process. **62**, 2984–2998 (2014)
2. Aragón Artacho, F.J., Geoffroy, M.H.: Characterizations of metric regularity of subdifferentials. J. Convex Anal. **15**, 365–380 (2008)
3. Amir, B.: First-Order Methods in Optimization. SIAM, Philadelphia (2017)
4. Attouch, H., Bolte, J.: On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Math. Program. **116**, 5–16 (2009)
5. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. Math. Oper. Res. **35**, 438–457 (2010)
6. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. Math. Program. **137**, 91–129 (2013)
7. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, Cham (2017)
8. Behr, F., Dolzmann, G.: A note on Clarke's generalized Jacobian for the inverse of bi-Lipschitz maps. J. Optim. Theory Appl. **200**, 852–857 (2024)
9. Bello-Cruz, Y., Li, G., Nghia, T.T.A.: Quadratic growth conditions and uniqueness of optimal solution to Lasso. J. Optim. Theory Appl. **194**, 167–190 (2022)
10. Bento, G. C., Mordukhovich, B. S., Mota, T. S., Nesterov, Y.: Convergence of descent optimization algorithms under Polyak-Łojasiewicz-Kurdyka conditions, to appear in J. Optim. Theory Appl. (2025); https://optimization-online.org/2025/02
11. Bolte, J., Daniilidis, A., Lewis, A.: The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J. Optim. **17**, 1205–1223 (2007)
12. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
13. Dontchev, A.L., Rockafellar, R.T.: Implicit Functions and Solution Mappings. Springer, New York (2014)

14. Fabian, M., Hiriart-Urruty, J.-B., Pauwels, E.: On the generalized Jacobian of the inverse of a Lipschitzian mapping. Set-Valued Var. Anal. **30**, 1443–1451 (2022)
15. Gong, P., Zhang, C., Lu, Z., Huang, J., Ye, J.: A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In: Proceedings of the 30th International Conference on Machine Learning (ICML 2013), 37-45 (2013)
16. Gotoh, J.Y., Takeda, A., Tono, K.: DC formulations and algorithms for sparse optimization problems. Math. Program. **169**, 141–176 (2018)
17. Hiriart-Urruty, J.-B.: How to regularize a difference of convex functions. J. Math. Anal. Appl. **162**, 196–209 (1991)
18. Hiriart-Urruty, J.-B., Strodiot, J.-J., Nguyen, V.H.: Generalized Hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data. Appl. Math. Optim. **11**, 43–56 (1984)
19. Hoang, T.: Convex Analysis and Global Optimization. Springer, New York (1998)
20. Horst, R., Thoai, N.V.: DC programming: overview. J. Optim. Theory Appl. **103**, 1–43 (1999)
21. Le, T.H.A., Pham, D.T.: The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. Ann. Oper. Res. **133**, 23–46 (2005)
22. Li, G., Pong, T.K.: Douglas-Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems. Math. Program. **159**, 371–401 (2016)
23. Li, G., Pong, T.K.: Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods. Found. Comput. Math. **18**, 1199–1232 (2018)
24. Liu, T., Pong, T.K., Takeda, A.: A refined convergence analysis of $pDCA_e$ with applications to simultaneous sparse recovery and outlier detection. Comput. Optim. Appl. **73**, 69–100 (2019)
25. Lu, Z., Zhou, Z., Sun, Z.: Enhanced proximal DC algorithms with extrapolation for a class of structured nonsmooth DC minimization. Math. Program. **176**, 369–401 (2019)
26. Luo, Z.-Q., Pang, J.-S., Ralph, D.: Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge (1996)
27. Mordukhovich, B.S.: Maximum principle in problems of time-optimal control with nonsmooth constraints. J. Appl. Math. Mech. **40**, 960–969 (1976)
28. Mordukhovich, B. S.: Variational Analysis and Generalized Differentiation, Vol. I: Basic Theory. Springer, Berlin/Heidelberg (2006)
29. Nikolova, M., Ng, M.K., Zhang, S., Ching, W.-K.: Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization. SIAM J. Imaging Sci. **1**, 2–25 (2008)
30. Ouyang, W., Liu, Y., Pong, T.K., Wang, H.: Kurdyka-Łojasiewicz exponent via Hadamard parametrization. SIAM J. Optim. **35**, 62–91 (2025)
31. Pham, D.T., Le, T.H.A.: Convex analysis approach to DC programming: theory, algorithm and applications. Acta Math. Vietnam **22**, 289–355 (1997)
32. Polyak, B.T.: Gradient methods for the minimization of functionals. USSR Comput. Math. Math. Phys. **3**, 864–878 (1963)
33. Prater-Bennette, A., Shen, L., Tripp, E. E.: The proximity operator of the Log-Sum penalty. J. Sci. Comput. 93 (2022)
34. Pun, Y.-M., So, A.M.-C.: Local strong convexity of source localization and error bound for target tracking under time-of-arrival measurements. IEEE Trans. Signal Process. **70**, 190–201 (2022)
35. Rebjock, Q., Boumal, N.: Fast convergence to non-isolated minima: four equivalent conditions for C^2 functions. Math. Program. **213**, 151–199 (2025)
36. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)
37. Sanjabi, M., Razaviyayn, M., Luo, Z.-Q.: Optimal joint base station assignment and beamforming for heterogeneous networks. IEEE Trans. Signal Process. **62**, 1950–1961 (2014)
38. Sun, K., Sun, X.A.: Algorithms for Difference-of-Convex Programs Based on Difference-of-Moreau-Envelopes Smoothing. INFORMS J. Optim. **5**, 321–339 (2023)
39. Themelis, A., Ahookhosh, M., Patrinos P.: On the acceleration of forward-backward splitting via an inexact Newton method. In: Splitting Algorithms, Modern Operator Theory, and Applications. Chapter 15, 363-412 (2019)
40. Themelis, A., Hermans, B., Patrinos, P.: A new envelope function for nonsmooth DC optimization. In: 2020 59th IEEE Conference on Decision and Control (CDC), 4697-4702 (2020)
41. Wang, Y.: New improved penalty methods for sparse reconstruction based on difference of two norms. Eprint. https://api.semanticscholar.org/Corpus, ID: 6733127 (2015)
42. Wen, B., Chen, X., Pong, T.K.: A proximal difference-of-convex algorithm with extrapolation. Comput. Optim. Appl. **69**, 297–324 (2018)

43. Yin, P., Lou, Y., He, Q., Xin, J.: Minimization of ℓ_{1-2} for compressed sensing. SIAM J. Sci. Comput. **37**, A536–A563 (2015)
44. Yu, P., Li, G., Pong, T.K.: Kurdyka-Łojasiewicz exponent via inf-projection. Found. Comput. Math. **22**, 1171–1217 (2022)
45. Zeng, L., Yu, P., Pong, T.K.: Analysis and algorithms for some compressed sensing models based on ℓ_1/ℓ_2 minimization. SIAM J. Optim. **31**, 1576–1603 (2021)