Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The Exact Modulus of the Generalized Concave Kurdyka-Łojasiewicz Property

Xianfu Wang, Ziyuan Wang

MATHEMATICS OF OPERATIONS RESEARCH
Articles in Advance, pp. 1–19
ISSN 0364-765X (print), ISSN 1526-5471 (online)
http://pubsonline.informs.org/journal/moor

The Exact Modulus of the Generalized Concave Kurdyka-Łojasiewicz Property

Xianfu Wang,[a,*] **Ziyuan Wang**[a]

[a] Department of Mathematics, University of British Columbia, Kelowna, British Columbia V1V 1V7, Canada
*Corresponding author
Contact: shawn.wang@ubc.ca, https://orcid.org/0000-0001-6753-2659 (XW); ziyuan.wang@alumni.ubc.ca (ZW)

Received: August 26, 2020
Revised: March 8, 2021; July 10, 2021; September 17, 2021
Accepted: October 5, 2021
Published Online in Articles in Advance: February 10, 2022

MSC2020 Subject Classifications: Primary: 49J52, 26D10, 90C26; secondary: 26A51, 26B25

https://doi.org/10.1287/moor.2021.1227

Abstract. We introduce a generalized version of the concave Kurdyka-Łojasiewicz (KL) property by employing nonsmooth desingularizing functions. We also present the exact modulus of the generalized concave KL property, which provides an answer to the open question regarding the optimal concave desingularizing function. The exact modulus is designed to be the smallest among all possible concave desingularizing functions. Examples are given to illustrate this pleasant property. In turn, using the exact modulus, we provide the sharpest upper bound for the total length of iterates generated by the celebrated Bolte-Sabach-Teboulle proximal alternating linearized minimization algorithm.

Funding: X. Wang and Z. Wang were partially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grants.

Keywords: generalized concave Kurdyka-Łojasiewicz property • Kurdyka-Łojasiewicz property • optimal concave desingularizing function • Bolte-Daniilidis-Ley-Mazet desingularizing function • proximal alternating linearized minimization • nonconvex optimization

1. Introduction

The continuous optimization community has witnessed a surging interest of employing the concave Kurdyka-Łojasiewicz (KL) property (see Definition 1) to solve problems from various applications, such as image processing (Banert and Bot [2] and Ochs et al. [14]), compressed sensing (Liu et al. [11], Wen et al. [19], Yu et al. [22]), machine learning (Won et al. [20]), and many more. The aforementioned work, despite devoting to different proximal-type algorithms, share a common theme: Employing the concave KL property as a regularity condition to ensure the algorithm of interest has the finite length property; see, for example, Ochs et al. [14, theorem 4.9] and Won et al. [20, theorem 3.1]. This pleasant convergence methodology can be traced back to the fundamental work of Bolte et al. [4], Bolte et al. [7], and Attouch et al. [1].

In the concave KL property, the concave desingularizing function plays a central role in estimating both the convergence rate and total length of iterates generated by the algorithm of interest; see, for example, Banert and Bot [2, theorem 1, lemma 4]. However, the concave desingularizing functions are not necessarily unique. It is natural to ask what the optimal (minimal) one is. This question remains open in the current literature. Classic definition of the concave KL property requires continuous differentiability of desingularizing functions, precluding the infimum of all concave desingularizing functions from staying within the same class. This paper is devoted to answering the following open question:

$$\textit{What is the optimal concave desingularizing function?} \tag{1}$$

To this end, we introduce an extension of the concave KL property and its associated exact modulus by allowing nonsmooth desingularizing functions. This extended framework allows us to capture the optimal concave desingularizing function through the exact modulus, yet still compatible with the usual concave KL convergence technique employed by a vast amount of literature. Our work opens the door to improve convergence results of a broad range of algorithms that adopt the concave KL assumption.

Throughout this paper, $\mathbb{R}^n$ is the standard Euclidean space with inner product $\langle x,y\rangle = x^Ty$ and Euclidean norm $\|x\| = \sqrt{\langle x,x\rangle}$ for $x,y\in\mathbb{R}^n$. The open ball centered at $\bar{x}\in\mathbb{R}^n$ with radius $r>0$ is denoted by $\mathbb{B}(\bar{x};r)$. We let $\overline{\mathbb{R}} = (-\infty,\infty]$, $\mathbb{R}_+ = [0,\infty)$, and $\mathbb{N} = \{1,2,3,\dots\}$. The distance function of a subset $K\subseteq\mathbb{R}^n$ is $\mathrm{dist}(\cdot,K):\mathbb{R}^n\to[0,\infty]:$

$x \longmapsto \operatorname{dist}(x, K) = \inf\{\|x-y\| : y \in K\}$. For $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $r_1, r_2 \in [-\infty, \infty]$, we set $[r_1 < f < r_2] = \{x \in \mathbb{R}^n : r_1 < f(x) < r_2\}$. For $\eta \in (0, \infty]$, denote by $\mathcal{K}_\eta$ the class of functions $\varphi : [0, \eta) \to \mathbb{R}_+$ that satisfy the following three conditions: (i) $\varphi : [0, \eta) \to \mathbb{R}_+$ is continuous with $\varphi(0) = 0$; (ii) φ is C^1 on $(0, \eta)$; (iii) $\varphi'(t) > 0$ for all $t \in (0, \eta)$. The pointwise version[1] of the concave KL property is defined as follows.

Definition 1. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lower semicontinuous (lsc).

i. We say f has the KL property at $\bar{x} \in \operatorname{dom} \partial f$ if there exist a neighborhood $U \ni \bar{x}$, $\eta \in (0, \infty]$ and a function $\varphi \in \mathcal{K}_\eta$ such that for all $x \in U \cap [0 < f - f(\bar{x}) < \eta]$,

$$\varphi'(f(x) - f(\bar{x})) \cdot \operatorname{dist}(0, \partial f(x)) \geq 1, \tag{2}$$

where $\partial f(x)$ denotes the limiting subdifferential of f at x (see Definition 2). The function φ is called a desingularizing function of f at $\bar{x}$ with respect to U and η. We say f is a KL function if it has the KL property at every $\bar{x} \in \operatorname{dom} \partial f$.

ii. We say that f has the concave KL property at $\bar{x} \in \operatorname{dom} \partial f$ if it has the KL property at $\bar{x}$ with desingularizing function $\varphi \in \mathcal{K}_\eta$ being concave. Moreover, we say f is a concave KL function if it has the concave KL property at every $\bar{x} \in \operatorname{dom} \partial f$.

The pioneering work of Łojasiewicz [12] and Kurdyka [9] on differentiable functions laid the foundation of the KL property, which was extended to nonsmooth functions by Bolte et al. [4] and [6]. In their seminal work, Bolte et al. [7] coined the term "KL property," gave characterizations, and proposed the Bolte-Daniilidis-Ley-Mazet (BDLM) desingularizing function, which is the optimal desingularizing function under certain continuity and locally integrability conditions; see Fact 5 for details and Kurdyka [9, theorem 1] for a similar result in a different setting. However, the optimal concave desingularizing function associated with the concave KL property may not be captured by the BDLM desingularizing function when the continuity and integrability assumptions fail; see Section 3.3. The main contributions of this paper are listed below:

• Definition 5 generalizes the concave KL property. The main difference is that we allow the desingularizing function to be nondifferentiable.

• Proposition 2 shows that the exact modulus of the generalized concave KL property, given in Definition 6, is the optimal concave desingularizing function, provided the existence of concave desingularizing functions. This result answers the open question (1).

• Theorem 1 provides the sharpest upper bound on $\sum_{k=1}^{\infty} \|z_{k+1} - z_k\|$, where $(z_k)_{k \in \mathbb{N}}$ is a sequence generated by the PALM algorithm. This result improves Bolte et al. [5, theorem 1].

Although most published articles emphasize desingularizing functions of the form $\varphi(t) = c \cdot t^{1-\theta}$ for $c > 0$ and $\theta \in [0, 1)$, the exact modulus has various forms. Proposition 3 gives an explicit formula for the optimal concave desingularizing function of locally convex and C^1 functions on the real line, in which case the exact modulus coincides with the desingularizing function obtained from the BDLM integrability condition. However, examples are given to show that the exact modulus is indeed the smaller one, even for nondifferentiable convex functions on the real line; see Examples 4 and 5. More examples comparing these two objects are provided in Section 3.3. As a by-product concerning intersections of convex functions, we show in Example 2 that there exist distinct strictly increasing convex C^2 functions $f, g : \mathbb{R}_+ \to \mathbb{R}_+$ with $f(0) = g(0) = 0$ such that $\inf\{x > 0 : f(x) = g(x)\} = 0$. Using our technique in Theorem 1, one may improve other algorithms that adopt the concave KL property assumption.

The structure of this paper is as the following: Elements in variational analysis, classical analysis, and facts of the classical KL property are collected in Section 2. The generalized concave KL property, the exact modulus and their properties are studied in Section 3. Various examples and comparisons to the BDLM desingularizing functions are also given in that section. We revisit the celebrated PALM algorithm in Section 4. Concluding remarks and directions for future work are presented in Section 5.

2. Preliminaries

2.1. Elements of Variational and Classical Analysis

We will use frequently the following subgradients in the nonconvex setting; see, for example, Mordukhovich [13] and Rockafellar and Wets [17].

Definition 2. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper function. We say that

i. $v \in \mathbb{R}^n$ is a Fréchet subgradient of f at $\bar{x} \in \operatorname{dom} f$, denoted by $v \in \hat{\partial} f(\bar{x})$, if for every $x \in \operatorname{dom} f$,

$$f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle + o(\|x - \bar{x}\|). \tag{3}$$

ii. $v \in \mathbb{R}^n$ is a limiting subgradient of f at $\bar{x} \in \operatorname{dom} f$, denoted by $v \in \partial f(\bar{x})$, if

$$v \in \left\{ v \in \mathbb{R}^n : \exists x_k \xrightarrow{f} \bar{x}, \exists v_k \in \hat{\partial} f(x_k), v_k \to v \right\}, \tag{4}$$

where $x_k \xrightarrow{f} \bar{x} \Leftrightarrow x_k \to \bar{x}$ and $f(x_k) \to f(\bar{x})$. Moreover, we set $\operatorname{dom} \partial f = \{x \in \mathbb{R}^n : \partial f(x) \neq \emptyset\}$. We say that $\bar{x} \in \operatorname{dom} \partial f$ is a stationary point if $0 \in \partial f(\bar{x})$.

Definition 3 is the definition of proximal mapping; see, for example, Bolte et al. [5] and Rockafellar and Wets [17].

Definition 3. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc and let λ be a positive real number. The proximal mapping is defined by

$$\operatorname{Prox}_\lambda^f(x) = \arg\min_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{\lambda}{2} \|x - y\|^2 \right\}, \quad \forall x \in \mathbb{R}^n.$$

The following fact follows from Rockafellar and Wets [17, theorem 1.25].

Fact 1. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc with $\inf_{\mathbb{R}^n} f > -\infty$. Then for $\lambda \in (0, \infty)$, $\operatorname{Prox}_\lambda^f(x)$ is nonempty for every $x \in \mathbb{R}^n$. Moreover, for every $v \in \mathbb{R}^n$ and $x \in \mathbb{R}^n$, we have

$$\operatorname{Prox}_\lambda^f\left(x - \frac{1}{\lambda} v\right) = \arg\min_y \left\{ \langle y - x, v \rangle + \frac{\lambda}{2} \|x - y\|^2 + f(y) \right\}.$$

Some well-known properties of convex functions on the real line are given in the following fact.

Fact 2 (Rockafellar [16, section 24], Bauschke and Combettes [3, chapter 17]). Let $I \subseteq \mathbb{R}$ be an open interval and let $\varphi : I \to \mathbb{R}$ be convex. Then the following statements hold true:

i. The side derivatives $\varphi'_-(t)$ and $\varphi'_+(t)$ are finite at every $t \in I$. Moreover, $\varphi'_-(t)$ and $\varphi'_+(t)$ are increasing.

ii. The function φ is differentiable except at countably many points of I, and $\varphi(s) - \varphi(t) = \int_t^s \varphi'_-(x) dx = \int_t^s \varphi'_+(x) dx$ for all $s, t \in I$.

iii. Let $t \in I$. Then for every $s \in I$, $\varphi(s) - \varphi(t) \geq \varphi'_-(t) \cdot (s - t)$.

The following result concerns the absolute continuity of integrals.

Fact 3 (Stromberg [18, theorem 6.79]). Let $f \in L^1$. Then for each $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\int_E |f(x)| ds < \varepsilon,$$

whenever $m(E) < \delta$, where E is a Lebesgue-measurable set and *m(E)* denotes its Lebesgue measure.

2.2. The Kurdyka-Łojasiewicz Property and Known Desingularizing Functions

In this section, we collect several facts about the KL property and desingularizing functions. We begin with a result asserting that the KL property at nonstationary points is automatic; see, for example, Attouch et al. [1, remark 3.2(b)] and also Li and Pong [10, lemma 2.1] for a detailed proof.

Fact 4. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc. Let $\bar{x} \in \operatorname{dom} \partial f$ be a nonstationary point. Then there exist c > 0 and $\theta \in [0, 1)$ such that f has the KL property at $\bar{x}$ with respect to $U = \mathbb{B}(\bar{x}; \varepsilon)$, $\eta = \varepsilon$ and $\varphi(t) = c \cdot t^{1-\theta}$.

We now recall some desingularizing functions described by Bolte et al. [7], which we will later compare with our main results. Recall that a proper and lsc function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is semiconvex if there exists $\alpha > 0$ such that $f + \frac{\alpha}{2} \| \cdot \|^2$ is convex.

Fact 5 (Bolte et al. [7, lemma 45, theorem 18]). Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be lsc and semiconvex. Let $\bar{x} \in [f = 0]$ and assume that there exist $\bar{r}, \bar{\varepsilon} > 0$ such that

$$x \in \mathbb{B}(\bar{x}; \bar{\varepsilon}) \cap [0 < f \leq \bar{r}] \Rightarrow 0 \notin \partial f(x). \tag{5}$$

Suppose there exist $r_0 \in (0, \bar{r})$ and $\varepsilon \in (0, \bar{\varepsilon})$ such that the function

$$u(r) = \frac{1}{\inf_{x \in \mathbb{B}(\bar{x};\varepsilon) \cap [f=r]} \operatorname{dist}(0, \partial f(x))}, \; r \in (0, r_0] \tag{6}$$

is finite valued and belongs to $L^1(0, r_0)$. Then the following statements hold:

i. There exists a continuous majorant $\bar{u} : (0, r_0] \to (0, \infty)$ such that $\bar{u} \in L^1(0, r_0)$ and $\bar{u}(r) \geq u(r)$ for all $r \in (0, r_0]$.

ii. Define for $t \in (0, \bar{r})$

$$\varphi(t) = \int_0^t \bar{u}(s) ds.$$

Then $\varphi \in \mathcal{K}_{r_0}$. For every $x \in \mathbb{B}(\bar{x};\varepsilon) \cap [0 < f \leq r_0]$, one has

$$\varphi'(f(x)) \operatorname{dist}(0, \partial f(x)) \geq 1.$$

Remark 1. Fact 5 is extracted from the implication $(v) \Rightarrow (i)$ in the proof of Bolte et al. [7, theorem 18], where the above desingularizing function $\varphi(t)$ was not stated explicitly in their theorem statement. Because results in this paper are on $\mathbb{R}^n$, we restrict Fact 5 to $\mathbb{R}^n$, in which case Assumption (24) of Bolte et al. [7, theorem 18] becomes superfluous.

Next we collect facts that ensure existence of concave desingularizing functions, which set the stage for our main results. With convexity, the following fact asserts that the desingularizing function given by Fact 5 can be taken to be concave with an enlarged domain $[0, \infty)$.

Fact 6 (Bolte et al. [7, lemma 45, theorem 29]). Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper, lsc, and convex function with $\inf f = 0$. Suppose that there exist $r_0 > 0$ and $\varphi \in \mathcal{K}_{r_0}$ such that for all $x \in [0 < f \leq r_0]$,

$$\varphi'(f(x)) \operatorname{dist}(0, \partial f(x)) \geq 1.$$

Then the following statements hold:

i. Define for $r \in (0, \infty)$ the function

$$u(r) = \frac{1}{\inf_{x \in [f=r]} \operatorname{dist}(0, \partial f(x))}.$$

Then u is finite valued, decreasing, and $u \in L^1(0, r_0)$. Moreover, there exists a decreasing continuous function $\tilde{u} \in L^1(0, r_0)$ such that $\tilde{u} \geq u$.

ii. Pick $\bar{r} \in (0, r_0)$ and define for $r \in (0, \infty)$

$$\varphi(r) = \begin{cases} \int_0^r \tilde{u}(s) ds, & \text{if } r \leq \bar{r}; \\ \int_0^r \tilde{u}(s) ds + \tilde{u}(\bar{r})(r - \bar{r}), & \text{otherwise.} \end{cases}$$

Then $\varphi \in \mathcal{K}_\infty$ is concave and for every $x \notin [f = 0]$,

$$\varphi'(f(x)) \operatorname{dist}(0, \partial f(x)) \geq 1.$$

Another celebrated result states that semialgebraic functions have the concave KL property.

Definition 4.

i. A set $E \subseteq \mathbb{R}^n$ is called semialgebraic if there exist finitely many polynomials $g_{ij}, h_{ij} : \mathbb{R}^n \to \mathbb{R}$ such that

$$E = \bigcup_{j=1}^{p} \bigcap_{i=1}^{q} \{x \in \mathbb{R}^n : g_{ij}(x) = 0 \text{ and } h_{ij}(x) < 0\}.$$

ii. A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called semialgebraic if its graph

$$\operatorname{gph} f = \{(x, y) \in \mathbb{R}^{n+1} : f(x) = y\}$$

is semialgebraic.

Fact 7 (Bolte et al. [6, corollary 16]). Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper and lsc function and let $\bar{x} \in \operatorname{dom} \partial f$. If f is semialgebraic, then it has the concave KL property at $\bar{x}$ with $\varphi(t) = c \cdot t^{1-\theta}$ for some c > 0 and $\theta \in (0, 1)$.

Remark 2.

i. Many useful functions in optimization are semialgebraic; see, for example, Attouch et al. [1] and Bolte et al. [5] and the references therein. Functions definable in o-minimal structure, which include semialgebraic functions, also satisfy the concave KL property; see Attouch et al. [1] and Bolte et al. [6].

ii. Although it is well known that real-polynomials are semialgebraic and thus have the KL property, only until very recently did Bolte et al. [8, corollary 9] provide an explicit formula for desingularizing functions of convex piecewise polynomials.

One can also determine the desingularizing function of the KL property for convex functions through the following growth condition.

Fact 8 (Bolte et al. [7, theorem 30]). Let $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ be a proper lsc convex function with $f(0)=\min f$. Let $S\subseteq\mathbb{R}^n$. Assume that there exists a function $m:\mathbb{R}_+\to\mathbb{R}_+$ that is continuous, strictly increasing, $m(0)=0$, $f\geq m(\mathrm{dist}(\cdot,\arg\min f))$ on $S\cap\mathrm{dom}\, f$ and

$$\exists\rho>0,\ \int_0^\rho \frac{m^{-1}(s)}{s}ds<\infty,$$

where m^{-1} denotes the inverse function of m. Then for all $x\in S\backslash\arg\min f$,

$$\varphi'(f(x))\mathrm{dist}(0,\partial f(x))\geq 1,$$

where for $t\in(0,\rho)$,

$$\varphi(t)=\int_0^t\frac{m^{-1}(s)}{s}ds.$$

Remark 3. Assuming there exists a concave desingularizing function, Facts 5, 6, and 8 may fail to capture the optimal one, even for convex functions on the real line; see Section 3.3.

3. The Generalized Concave KL Property and Its Exact Modulus

In this section, we provide an answer to the open question (1). The generalized concave Kurdyka-Łojasiewicz property and its exact modulus are introduced to provide the answer. Given existence of concave desingularizing functions, we shall see that the exact modulus is indeed optimal.

3.1. The Generalized Concave KL Property

For $\eta\in(0,\infty]$, denote by Φ_η the class of functions $\varphi:[0,\eta)\to\mathbb{R}_+$ satisfying the following conditions: (i) $\varphi(t)$ is right continuous at $t=0$ with $\varphi(0)=0$; (ii) φ is strictly increasing on $[0,\eta)$. Recall that the left derivative of $\varphi:[0,\infty)\to\mathbb{R}$ at $t\in(0,\infty)$ is defined by

$$\varphi'_-(t)=\lim_{s\to t^-}\frac{\varphi(s)-\varphi(t)}{s-t}.$$

Some useful properties of concave $\varphi\in\Phi_\eta$ are collected below.

Lemma 1. *For $\eta\in(0,\infty]$ and concave $\varphi\in\Phi_\eta$, the following assertions hold:*

i. *Let $t>0$. Then $\varphi(t)=\lim_{u\to0^+}\int_u^t\varphi'_-(s)ds=\int_0^t\varphi'_-(s)ds$.*

ii. *The function $t\longmapsto\varphi'_-(t)$ is decreasing and $\varphi'_-(t)>0$ for $t\in(0,\eta)$.*

iii. *For $0\leq s<t<\eta$, $\varphi'_-(t)\leq\frac{\varphi(t)-\varphi(s)}{t-s}$.*

Proof.

i. Invoking Fact 2(ii) yields

$$\varphi(t)=\lim_{u\to0^+}(\varphi(t)-\varphi(u))=\lim_{u\to0^+}\int_u^t\varphi'_-(s)ds<\infty,$$

where the first equality holds because φ is right-continuous at zero with $\varphi(0)=0$. Let $(u_n)_{n\in\mathbb{N}}$ be a decreasing sequence with $u_1<t$ such that $u_n\to0^+$ as $n\to\infty$. For each n, define $h_n:(0,t]\to\mathbb{R}_+$ by $h_n(s)=\varphi'_-(s)$ if $s\in(u_n,t]$ and

$h_n(s) = 0$ otherwise. Then the sequence $(h_n)_{n\in\mathbb{N}}$ satisfies (a) $h_n \le h_{n+1}$ for every $n \in \mathbb{N}$; (b) $h_n(s) \to \varphi'_-(s)$ pointwise on $(0,t)$; (c) The integral $\int_0^t h_n(s)ds = \int_{u_n}^t \varphi'_-(s)ds = \varphi(t) - \varphi(u_n) \le \varphi(t) - \varphi(0) < \infty$ for every $n \in \mathbb{N}$. Hence, the monotone convergence theorem implies that

$$\lim_{u\to 0^+} \int_u^t \varphi'_-(s)ds = \lim_{n\to\infty} \int_{u_n}^t \varphi'_-(s)ds = \lim_{n\to\infty} \int_0^t h_n(s)ds = \int_0^t \varphi'_-(s)ds.$$

ii. According to Fact 2(i), the function $t \longmapsto \varphi'_-(t)$ is decreasing. Suppose that $\varphi'_-(t_0) = 0$ for some $t_0 \in (0,\eta)$. Then by the monotonicity of φ'_- and (i), we would have $\varphi(t) - \varphi(t_0) = \int_{t_0}^t \varphi'_-(s)ds \le (t - t_0)\varphi'_-(t_0) = 0$ for $t > t_0$, which contradicts with the assumption that φ is strictly increasing.

iii. For $0 < s < t < \eta$, applying Fact 2(iii) to the convex function $-\varphi$ yields that $-\varphi(s) + \varphi(t) \ge -\varphi'_-(t)(s-t) \Longleftrightarrow \varphi'_-(t) \le (\varphi(t) - \varphi(s))/(t-s)$. The desired inequality then follows from the right-continuity of φ at zero. □

Now we introduce the pointwise generalized concave KL property and its setwise variant.

Definition 5. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc. Let $\bar{x} \in \operatorname{dom} \partial f$ and $\mu \in \mathbb{R}$, and let $V \subseteq \operatorname{dom} \partial f$ be a nonempty subset.

i. We say that f has the pointwise *generalized concave KL* at $\bar{x} \in \operatorname{dom} \partial f$ if there exist a neighborhood $U \ni \bar{x}$, $\eta \in (0,\infty]$ and concave $\varphi \in \Phi_\eta$, such that for all $x \in U \cap [0 < f - f(\bar{x}) < \eta]$,

$$\varphi'_-(f(x) - f(\bar{x})) \cdot \operatorname{dist}(0, \partial f(x)) \ge 1. \tag{7}$$

ii. Suppose that $f(x) = \mu$ on V. We say f has the setwise generalized concave KL property[2] on V if there exist $U \supset V$, $\eta \in (0,\infty]$ and concave $\varphi \in \Phi_\eta$ such that for every $x \in U \cap [0 < f - \mu < \eta]$,

$$\varphi'_-(f(x) - \mu) \cdot \operatorname{dist}(0, \partial f(x)) \ge 1. \tag{8}$$

Remark 4.

i. Evidently, the generalized concave KL property on a set V reduces to the generalized concave KL property at $\bar{x}$ if $V = \{\bar{x}\}$. This setwise definition will be useful in Section 4.

ii. Clearly, the concave KL property (see Definition 1) implies the generalized concave KL property. However, the generalized notion allows desingularizing functions to be nondifferentiable by using the left derivative, which is well defined thanks to Fact 2.

In the rest of this subsection, we work toward generalizing a result by Bolte et al. [5, lemma 6], whose proof we will follow. For nonempty subset $\Omega \subseteq \mathbb{R}^n$ and $\varepsilon \in (0,\infty]$, define $\Omega_\varepsilon = \{x \in \mathbb{R}^n : \operatorname{dist}(x,\Omega) < \varepsilon\}$. Let us recall the Lebesgue number lemma (Pugh [15, theorem 55]).

Lemma 2. *Let $\Omega \subseteq \mathbb{R}^n$ be a nonempty compact subset. Suppose that $\{U_i\}_{i=1}^p$ is a finite open cover of Ω. Then there exists $\varepsilon > 0$, which is called the Lebesgue number of Ω, such that*

$$\Omega \subseteq \Omega_\varepsilon \subseteq \bigcup_{i=1}^p U_i.$$

Proposition 1 connects the pointwise generalized concave KL property to its setwise counterpart, generalizes Bolte et al. [5, lemma 6], and will play a key role in Section 4.

Proposition 1. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper lsc and let $\mu \in \mathbb{R}$. Let $\Omega \subseteq \operatorname{dom} \partial f$ be a nonempty compact set on which $f(x) = \mu$ for all $x \in \Omega$. Suppose that f satisfies the pointwise generalized concave KL property at each $x \in \Omega$. Then there exist $\varepsilon > 0, \eta \in (0,\infty]$ and concave $\varphi(t) \in \Phi_\eta$ such that f has the setwise generalized concave KL property on Ω with respect to $U = \Omega_\varepsilon$, η and φ.*

Proof. For each $x \in \Omega$, there exist $\varepsilon = \varepsilon(x) > 0, \eta = \eta(x) \in (0,\infty]$ and concave $\varphi(t) = \varphi_x(t) \in \Phi_\eta$ such that for $y \in \mathbb{B}(x;\varepsilon) \cap [0 < f - f(x) < \eta]$,

$$\varphi'_-(f(y) - f(x)) \cdot \operatorname{dist}(0, \partial f(y)) \ge 1.$$

Note that $\Omega \subseteq \bigcup_{x\in\Omega} \mathbb{B}(x;\varepsilon)$. Because Ω is compact, there exist elements $x_1, \ldots, x_p \in \Omega$ such that $\Omega \subseteq \bigcup_{i=1}^p \mathbb{B}(x_i;\varepsilon_i)$. Moreover, for each i and $x \in \mathbb{B}(x_i;\varepsilon_i) \cap [0 < f - f(x_i) < \eta_i] = \mathbb{B}(x_i;\varepsilon_i) \cap [0 < f - \mu < \eta_i]$, one has

$$(\varphi_i)'_-(f(x) - \mu) \cdot \operatorname{dist}(0, \partial f(x)) \ge 1. \tag{9}$$

Define $\varphi(t) = \sum_{i=1}^{p}\varphi_i(t)$ and $\eta = \min_{1\le i\le p}\eta_i$. Evidently, φ is a concave function belonging to Φ_η. By Lemma 2, there exists $\varepsilon > 0$ such that $\Omega \subseteq \Omega_\varepsilon \subseteq \bigcup_{i=1}^{p}\mathbb{B}(x_i;\varepsilon_i)$, which by the fact that $\eta \le \eta_i$ for every i further implies that

$$x \in \Omega_\varepsilon \cap [0 < f - \mu < \eta] \Rightarrow \exists i_0,\ s.t.,\ x \in \mathbb{B}(x_{i_0};\varepsilon_{i_0}) \cap [0 < f - \mu < \eta_{i_0}].$$

Hence, for every $x \in \Omega_\varepsilon \cap [0 < f - \mu < \eta]$, one has

$$\varphi'_-(f(x) - \mu)\cdot \mathrm{dist}(0, \partial f(x)) \ge (\varphi_{i_0})'_-(f(x) - \mu)\cdot \mathrm{dist}(0, \partial f(x)) \ge 1,$$

where the first inequality holds because $(\varphi_i)'_-(t) > 0$ by Lemma 1 and the second inequality is implied by (9). □

3.2. The Exact Modulus of the Generalized Concave KL Property

Following the definition of the generalized concave KL property, we introduce its associated exact modulus.

Definition 6. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc. Let $\bar{x} \in \mathrm{dom}\,\partial f$ and let $U \subseteq \mathrm{dom}\,\partial f$ be a neighborhood of $\bar{x}$. Let $\eta \in (0,\infty]$. Furthermore, define $h : (0,\eta) \to \mathbb{R}$ by

$$h(s) = \sup\{\mathrm{dist}^{-1}(0, \partial f(x)) : x \in U \cap [0 < f - f(\bar{x}) < \eta], s \le f(x) - f(\bar{x})\}.$$

Suppose that $h(s) < \infty$ for $s \in (0,\eta)$. The exact modulus of the generalized concave KL property of f at $\bar{x}$ with respect to U and η is the function $\tilde{\varphi} : [0,\eta) \to \mathbb{R}_+ : t \longmapsto \int_0^t h(s)ds,\ \ \forall t \in (0,\eta)$, and $\tilde{\varphi}(0) = 0$. If $U \cap [0 < f - f(\bar{x}) < \eta] = \emptyset$ for given $U \ni \bar{x}$ and $\eta > 0$, then we set the exact modulus with respect to U and η to be $\tilde{\varphi}(t) \equiv 0$.

Remark 5.

i. The essential difference between the exact modulus and the BDLM desingularizing function in Fact 5 is that the exact modulus utilizes the set $U \cap [s \le f(x) - f(\bar{x})]$ instead of $U \cap [s = f(x) - f(\bar{x})]$. In addition, the exact modulus $\tilde{\varphi}$ is not necessarily differentiable, whereas the BDLM desingularizing requires differentiability. In order for the exact modulus to be well defined, however, it requires the existence of concave desingularizing functions, which is a strong assumption. Examples of such functions include convex functions satisfying the KL property and semialgebraic functions; cf. Facts 6 and 7, which are frequently treated in papers devoted to algorithmic applications of the concave KL property (Attouch et al. [1], Banert and Bot [2], Bolte et al. [5], Liu et al. [11], Ochs et al. [14], Wen et al. [19], Won et al. [20], Yu et al. [22]).

ii. Note that $\lim_{s\to 0^+} h(s)$ should be infinity if $\bar{x}$ is a stationary point, in which case the function $\tilde{\varphi}(t)$ represents a limit of Riemann or Lebesgue integrals.

iii. The assumption that $h(s) < \infty$ for $s \in (0,\eta)$ is necessary. For example, consider the exact modulus of the generalized concave KL property of the function $f(x) = 1 - e^{-|x|}$ at zero. Then one has for $x \ne 0$, $\mathrm{dist}^{-1}(0, \partial f(x)) = e^{|x|}$. Let $U = \mathbb{R}$ and $\eta_1 = 1$. Then

$$h_1(s) = h_{U,\eta_1}(s) = \sup\{\mathrm{dist}^{-1}(0, \partial f(x)) : \mathbb{R} \cap [0 < f < 1], s \le f(x)\} = \infty.$$

This can be avoided by shrinking the set $U \cap [0 < f - f(\bar{x}) < \eta]$. Let $\eta_2 \in (0,1)$. Then

$$h_2(s) = h_{U,\eta_2}(s) = \sup\left\{\mathrm{dist}^{-1}(0, \partial f(x)) : x \in \mathbb{R} \cap [0 < f < \eta_2], s \le f(x)\right\} = \frac{1}{1-\eta_2}.$$

The exact modulus $\tilde{\varphi}$ is designed to be the optimal concave desingularizing function. The following lemma is needed to prove this property.

Lemma 3. *Let $\eta \in (0,\infty]$ and let $h : (0,\eta) \to \mathbb{R}_+$ be a positive-valued decreasing function. Define $\varphi(t) = \int_0^t h(s)ds$ for $t \in (0,\eta)$ and set $\varphi(0) = 0$. Suppose that $\varphi(t) < \infty$ for $t \in (0,\eta)$. Then φ is a strictly increasing concave function on $[0,\eta)$ with*

$$\varphi'_-(t) \ge h(t)$$

for $t \in (0,\eta)$ and right-continuous at zero. If in addition h is a continuous function, then φ is C^1 on $(0,\eta)$.

Proof. Let $0 < t_0 < t_1 < \eta$. Then $\varphi(t_1) - \varphi(t_0) = \int_{t_0}^{t_1} h(s)ds \ge (t_1 - t_0)\cdot h(t_1) > 0$, which means φ is strictly increasing. Applying Fact 3, one concludes that $\varphi(t) \to \varphi(0) = 0$ as $t \to 0^+$. The concavity of φ and the inequality $\varphi'_-(t) \ge h(t)$

follow from a similar argument as in Rockafellar [16, theorem 24.2]. If in addition h is continuous, then by applying the fundamental theorem of calculus, one concludes that φ is C^1 on $(0,\eta)$. □

Proposition 2. *Let $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ be proper lsc and let $\bar{x}\in\operatorname{dom}\partial f$. Let U be a nonempty neighborhood of $\bar{x}$ and $\eta\in(0,\infty]$. Let $\varphi\in\Phi_\eta$ be concave and suppose that f has the generalized concave KL property at $\bar{x}$ with respect to U, η and φ. Then the exact modulus of the generalized concave KL property of f at $\bar{x}$ with respect to U and η, denoted by $\tilde{\varphi}$, is well defined, concave, and satisfies*

$$\tilde{\varphi}(t)\le\varphi(t),\quad \forall t\in[0,\eta).$$

Moreover, the function f has the generalized concave KL property at $\bar{x}$ with respect to U, η and $\tilde{\varphi}$. Consequently, the exact modulus $\tilde{\varphi}$ satisfies

$$\tilde{\varphi}=\inf\{\varphi\in\Phi_\eta:\varphi \text{ is a concave desingularizing function of } f \text{ at } \bar{x} \text{ with respect to } U \text{ and } \eta\}.$$

Proof. Let us show first that $\tilde{\varphi}(t)\le\varphi(t)$ on $[0,\eta)$, which implies immediately that $\tilde{\varphi}$ is well defined. If $U\cap[0<f-f(\bar{x})<\eta]=\emptyset$, then by our convention $\tilde{\varphi}(t)=0\le\varphi(t)$ for every $t\in[0,\eta)$. Therefore, we proceed with the assumption that $U\cap[0<f-f(\bar{x})<\eta]\neq\emptyset$. By assumption, one has for $x\in U\cap[0<f-f(\bar{x})<\eta]$,

$$\varphi'_{-}(f(x)-f(\bar{x}))\cdot\operatorname{dist}(0,\partial f(x))\ge 1,$$

which guarantees that $\operatorname{dist}(0,\partial f(x))>0$. Fix $s\in(0,\eta)$ and recall from Lemma 1(ii) that $\varphi'_{-}(t)$ is decreasing. Then for $x\in U\cap[0<f-f(\bar{x})<\eta]$ with $s\le f(x)-f(\bar{x})$, we have

$$\operatorname{dist}^{-1}(0,\partial f(x))\le\varphi'_{-}(f(x)-f(\bar{x}))\le\varphi'_{-}(s).$$

Taking the supremum over all $x\in U\cap[0<f-f(\bar{x})<\eta]$ satisfying $s\le f(x)-f(\bar{x})$ yields

$$h(s)\le\varphi'_{-}(s),$$

where $h(s)=\sup\{\operatorname{dist}^{-1}(0,\partial f(x)):x\in U\cap[0<f-f(\bar{x})<\eta],s\le f(x)-f(\bar{x})\}$. If $\lim_{s\to 0^+}h(s)=\infty$, then one needs to treat $\tilde{\varphi}(t)$ as an improper integral. For $t\in(0,\eta)$,

$$\tilde{\varphi}(t)=\lim_{u\to 0^+}\int_u^t h(s)ds\le\lim_{u\to 0^+}\int_u^t\varphi'_{-}(s)ds=\varphi(t)<\infty,$$

where the last equality follows from Lemma 1. If $\lim_{s\to 0^+}h(s)<\infty$, then the above argument still applies.

Recall that $\operatorname{dist}(0,\partial f(x))>0$ for every $x\in U\cap[0<f-f(\bar{x})<\eta]$. Hence, $h(s)$ is positive valued. Take $s_1,s_2\in(0,\eta)$ with $s_1\le s_2$. Then for $x\in U\cap[0<f-f(\bar{x})<\eta]$,

$$s_2\le f(x)-f(\bar{x})\Rightarrow s_1\le f(x)-f(\bar{x}),$$

implying that $h(s_2)\le h(s_1)$. Therefore, $h(s)$ is decreasing. Invoking Lemma 3, one concludes that $\tilde{\varphi}$ is a concave function belonging to Φ_η and $\varphi'_{-}(t)\ge h(t)$ for every $t\in(0,\eta)$.

Let $t\in(0,\eta)$. Then for $x\in U\cap[0<f-f(\bar{x})<\eta]$ with $t=f(x)-f(\bar{x})$,

$$\tilde{\varphi}'_{-}(f(x)-f(\bar{x}))\ge h(t)\ge\operatorname{dist}^{-1}(0,\partial f(x)),$$

where the last inequality is implied by the definition of $h(s)$, from which the generalized concave KL property readily follows because t is arbitrary.

Recall that φ is an arbitrary concave desingularizing function of f at $\bar{x}$ with respect to U and η and $\tilde{\varphi}(t)\le\varphi(t)$ for all $t\in[0,\eta)$. Hence,

$$\tilde{\varphi}\le\inf\{\varphi\in\Phi_\eta:\varphi \text{ is a concave desingularizing function of } f \text{ at } \bar{x} \text{ with respect to } U \text{ and } \eta\}.$$

On the other hand, the converse inequality holds as $\tilde{\varphi}$ is a concave desingularizing function of f at $\bar{x}$ with respect to U and η. □

Our next example shows that the exact modulus is not necessarily differentiable, which justifies the nonsmooth extension of desingularizing functions in Definition 5.

Example 1. Let $\rho > 0$. Consider the function given by

$$f(x) = \begin{cases} 2\rho|x| - 3\rho^2/2, & \text{if } |x| > \rho; \\ |x|^2/2, & \text{if } |x| \le \rho. \end{cases}$$

Then the function

$$\tilde{\varphi}_1(t) = \begin{cases} \sqrt{2t}, & \text{if } 0 \le t \le \rho^2/2; \\ t/(2\rho) + 3\rho/4, & \text{if } t > \rho^2/2, \end{cases}$$

is the exact modulus of the generalized concave KL property of f at $\bar{x} = 0$ with respect to $U = \mathbb{R}$ and $\eta = \infty$.

Proof. It is easy to see that for $s \in (0, \rho^2/2]$,

$$\begin{aligned} h_1(s) &= \sup\{\mathrm{dist}^{-1}(0, \partial f(x)) : x \in \mathbb{R} \cap [0 < f < \infty], s \le f(x)\} \\ &= \sup\{\mathrm{dist}^{-1}(0, \partial f(x)) : |x| \ge \sqrt{2s}\} = 1/\sqrt{2s}, \end{aligned}$$

and for $s > \rho^2/2$,

$$\begin{aligned} h_1(s) &= \sup\{\mathrm{dist}^{-1}(0, \partial f(x)) : x \in \mathbb{R} \cap [0 < f < \infty], s \le f(x)\} \\ &= \sup\{\mathrm{dist}^{-1}(0, \partial f(x)) : x \ne 0, |x| \ge s/(2\rho) + 3\rho/4\} = 1/(2\rho), \end{aligned}$$

from which the desired result readily follows. □

It is difficult to compute directly the exact modulus of the generalized concave KL property for multivariable functions because of its complicated definition. However, on the real line, we have the following pleasing formula.

Proposition 3. *Let $f : \mathbb{R} \to \overline{\mathbb{R}}$ be proper and lsc. Let $\bar{x}$ be a stationary point. Suppose that there exists an interval $(a, b) \subseteq \mathrm{int\,dom}\, f$, where $-\infty \le a < b \le \infty$, on which f is convex on (a, b) and C^1 on $(a, b)\backslash\{\bar{x}\}$. Set $\eta = \min\{f(a) - f(\bar{x}), f(b) - f(\bar{x})\}$, $f_1(x) = f(x + \bar{x}) - f(\bar{x})$ for $x \in (a - \bar{x}, 0]$ and $f_2(x) = f(x + \bar{x}) - f(\bar{x})$ for $[0, b - \bar{x})$. Furthermore, define $\tilde{\varphi} : [0, \eta) \to \mathbb{R}_+$,*

$$t \longmapsto \int_0^t \max\{(-f_1^{-1})'(s), (f_2^{-1})'(s)\} ds, \quad \forall t \in (0, \eta) \tag{10}$$

and $\tilde{\varphi}(0) = 0$. Then $\tilde{\varphi}(t)$ is the exact modulus of the generalized concave KL property at $\bar{x}$ with respect to $U = (a, b)$ and η. Note that we set $f(x) = \infty$ if $x = \pm\infty$ and $(f_i^{-1})' \equiv 0$ if f_i^{-1} does not exist.

Proof. Replacing $f(x)$ by $g(x) = f(x + \bar{x}) - f(\bar{x})$ if necessary, we assume without loss of generality that $\bar{x} = 0$ and $f(\bar{x}) = 0$. Then by the assumption that $\bar{x} = 0$ is a stationary point, we have $0 \in \partial f(0) = [f'_-(0), f'_+(0)]$, meaning that $f'_-(0) \le 0 \le f'_+(0)$. We learn from Fact 2 that $f'_-(x)$ and $f'_+(x)$ are increasing functions. Combining the C^1 assumption, we have $f'(x) = f'_-(x) \le f'_-(0) \le 0$ on $(a, 0)$ and $f'(x) = f'_+(x) \ge f'_+(0) \ge 0$ on $(0, b)$. Hence, for $x \in (a, b)\backslash\{0\}$,

$$\mathrm{dist}\,(0, \partial f(x)) = |f'(x)| = \begin{cases} -f'(x) = -f'_1(x), & \text{if } x \in (a, 0); \\ f'(x) = f'_2(x), & \text{if } x \in (0, b), \end{cases}$$

meaning that the function $x \longmapsto \mathrm{dist}(0, \partial f(x))$ is decreasing on $(a, 0)$ and increasing on $(0, b)$.

Now we work toward showing that $h(s) = \max\{(-f_1^{-1})'(s), (f_2^{-1})'(s)\}$, where $h(s)$ is the function given in Definition 6. Recall that $f'(x)$ is increasing on $(a, b)\backslash\{0\}$ with $f'(x) \le 0$ on $(a, 0)$ and $f'(x) \ge 0$ on $(0, b)$. Shrinking the interval (a, b) if necessary, we only need to consider the following four cases.

Case 1. Suppose that $f'_1(x) < 0$ for $x \in (a, 0)$ and $f'_2(x) > 0$ for $x \in (0, b)$. Then both f_1 and f_2 are invertible and

$$\mathrm{dist}^{-1}(0, \partial f(x)) = \begin{cases} -1/f'_1(x), & \text{if } a < x < 0; \\ 1/f'_2(x), & \text{if } 0 < x < b. \end{cases}$$

Fix $s \in (0, \eta)$. For $x \in (a, 0)$, on which f_1 is decreasing,

$$s \le f(x) = f_1(x) \Longleftrightarrow f_1^{-1}(s) \ge x. \tag{11}$$

Similarly for $x \in (0,b)$,

$$s \le f(x) = f_2(x) \Longleftrightarrow f_2^{-1}(s) \le x. \tag{12}$$

Hence, one concludes that for $x \in (a,b)$,

$$s \le f(x) \Longleftrightarrow x \in (a, f_1^{-1}(s)] \bigcup [f_2^{-1}(s), b).$$

On the other hand, we have $0 < f(x) < \eta \Longleftrightarrow x \in (f_1^{-1}(\eta), f_2^{-1}(\eta)) \backslash \{0\}$, where $f_1^{-1}(\eta) > a$ and $f_2^{-1}(\eta) < b$, which means $(a,b) \cap [0 < f < \eta] = (f_1^{-1}(\eta), f_2^{-1}(\eta)) \backslash \{0\}$.

Altogether, we conclude that the function $h : (0,\eta) \to \mathbb{R}$ given in Definition 6 satisfies

$$\begin{aligned} h(s) &= \sup\{\mathrm{dist}^{-1}(0, \partial f(x)) : x \in (a,b) \cap [0 < f < \eta], s \le f(x)\} \\ &= \sup\{\mathrm{dist}^{-1}(0, \partial f(x)) : x \in (f_1^{-1}(\eta), f_1^{-1}(s)] \bigcup [f_2^{-1}(s), f_2^{-1}(\eta))\} \\ &= \max\{-1/(f_1')(f_1^{-1}(s)), 1/(f_2')(f_2^{-1}(s))\} \\ &= \max\{(-f_1^{-1})'(s), (f_2^{-1})'(s)\}, \end{aligned}$$

where the third equality is implied by the fact that $x \longmapsto \mathrm{dist}^{-1}(0, \partial f(x))$ is increasing on $(a,0)$ and decreasing on $(0,b)$.

Case 2. If $f'(x) = 0$ on $(a,0)$ and $f'(x) > 0$ on $(0,b)$, then f_2 is invertible and $(f_2^{-1})'(s) = 1/f'(f_2^{-1}(s)) > 0$ on $(0,\eta)$. Note that by our convention $(f_1^{-1})'(s)$ is set to be zero for all s. Hence, it suffices to prove $h(s) = (f_2^{-1})'(s)$. For $s \in (0,\eta)$,

$$\begin{aligned} h(s) &= \sup\{\mathrm{dist}^{-1}(0, \partial f(x)) : x \in U \cap [0 < f < \eta], s \le f(x)\} \\ &= \sup\{1/f_2'(x) : x \in [f_2^{-1}(s), b)\} = 1/f_2'(f_2^{-1}(s)) = (f_2^{-1})'(s), \end{aligned}$$

where the second equality is implied by (12), $U \cap [0 < f < \eta] = (0,b)$, and the fact that $1/f_2'(x)$ is decreasing on $(0,b)$.

Case 3. If $f'(x) < 0$ on $(a,0)$ and $f'(x) = 0$ on $(0,b)$, then f_1 is invertible. A similar argument proves that $h(s) = (-f_1^{-1})'(s)$.

Case 4. Now we consider the case where $f'(x) = 0$ on (a, b), in which case $U \cap [0 < f < \eta] = \emptyset$ and the corresponding exact modulus is $\tilde{\varphi} \equiv 0$ by our convention. Moreover, $(-f_1^{-1})'(s)$ and $(f_2^{-1})'(s)$ are set to be constant zero. Hence, we have $\tilde{\varphi}(t) = \int_0^t 0\, ds = 0$, which completes the proof. □

Remark 6. In the setting of Proposition 3, it is easy to see that the exact modulus satisfies

$$\tilde{\varphi}(t) = \int_0^t \frac{ds}{\inf_{(a,b)\cap[f=s]} \mathrm{dist}(0, \partial f(x))},$$

which means that the BDLM desingularizing function given by Fact 5 coincides with the exact modulus. However, this is not true without the C^1 assumption in Proposition 3; see Examples 4 and 5.

Combining Fact 4 and Proposition 3, we immediately obtain the following corollary.

Corollary 1. *Let $f : (a,b) \to \mathbb{R}$ be a differentiable convex function. Then f is a concave KL function, that is, f satisfies the concave KL-property at every point of (a, b).*

When proving Proposition 3, our initial attempt is to take $\eta > 0$ sufficiently small so that $t \longmapsto \max\{-f_1^{-1}(t), f_2^{-1}(t)\}$ becomes either $-f_1^{-1}$ or f_2^{-1} on $[0,\eta)$. This attempt leads to a question of independent interest: Let f and g be two smooth strictly increasing convex functions defined on $[0,\infty)$ with $f(0) = g(0)$.

$$\textit{Is}\inf\{x > 0 : f(x) = g(x)\}\ \textit{always positive}?$$

The answer is negative, as our next example shows.

Example 2. There exist strictly increasing convex C^2 functions $f, g : \mathbb{R}_+ \to \mathbb{R}_+$ with $f(0) = g(0) = 0$ such that $\inf\{x > 0 : f(x) = g(x)\} = 0$. To be specific, let $h : \mathbb{R} \to \mathbb{R}$ be given by

$$h(x) = \begin{cases} \sin\left(\frac{1}{x}\right)e^{-\frac{1}{x^2}}, & \text{if } x \neq 0; \\ 0, & \text{if } x = 0. \end{cases}$$

Define $h''_+(s) = \max\{h''(s), 0\}$ and $h''_-(s) = -\min\{h''(s), 0\}$. Furthermore, set $f_1(x) = \int_0^x h''_-(t)dt$ and $g_1(x) = \int_0^x h''_+(t)dt$. Then the functions $f, g : \mathbb{R}_+ \to \mathbb{R}_+$ given by

$$g(x) = \int_0^x g_1(t)dt, \; f(x) = \int_0^x f_1(t)dt$$

are strictly increasing convex and C^2 functions with $f(0) = g(0) = 0$ and satisfy $g(x) - f(x) = h(x)$. Hence, $\inf\{x > 0 : f(x) = g(x)\} = \inf\{x > 0 : h(x) = 0\} = 0$.

Proof. Note that $h(x) \in C^\infty$. We now show that h is a difference of convex functions. Observe from the definition that $h''_+(s)$ and $h''_-(s)$ are positive valued and continuous. Then by the fundamental theorem of calculus, $g_1(x)$ and $f_1(x)$ are both increasing C^1 functions with $f_1'(x) = h''_-(x)$ and $g_1'(x) = h''_+(x)$. Because $h''(x) = h''_+(x) - h''_-(x)$,

$$g_1(x) - f_1(x) = \int_0^x h''(s)ds = h'(x) - h'(0) = h'(x).$$

Suppose that there exists $x_0 > 0$ such that $g_1(x_0) = 0$. Then $g_1(x) = 0$ for $x \in (0, x_0)$, which implies $h''(x) \leq 0$ on $(0, x_0)$. This is impossible because $h''(x)$ oscillates between positive and negative infinitely many times when $x \to 0^+$. Hence, g_1 is strictly positive. A similar argument shows that f_1 is also strictly positive. Then applying the fundamental theorem of calculus, one concludes that f and g are strictly increasing C^2 functions with $f'(x) = f_1(x)$ and $g'(x) = g_1(x)$. Functions f and g are both convex because $f'' = f_1' = h''_- \geq 0$ and $g'' = g_1' = h''_+ \geq 0$. Furthermore,

$$g(x) - f(x) = \int_0^x h'(s)ds = h(x) - h(0) = h(x).$$

Hence, $\inf\{x > 0 : f(x) = g(x)\} = \inf\{x > 0 : h(x) = 0\} = \inf\{1/(n\pi) : n \in \mathbb{N}\} = 0$. □

3.3. Comparison with the BDLM Desingularizing Functions

In this subsection, we compare the exact modulus to the BDLM desingularizing functions in Facts 5 and 6. A comparison with the growth condition in Fact 8 is also carried out. Examples will be given to show that the exact modulus is the optimal concave desingularizing function, provided that such functions exist.

Below, we use φ for the BDLM desingularizing function. By picking a decreasing and continuous majorant[3] of u in the integrability Condition (6) and integrating, one can get a concave BDLM desingularizing function, which may not be the smallest.

Example 3. Let $U = \mathbb{R}$ and $\eta = \infty$. Define $f : \mathbb{R} \to \mathbb{R}_+$ by

$$f(x) = \begin{cases} \frac{1}{2}x, & \text{if } 0 \leq x \leq \frac{1}{4}; \\ \frac{3}{2}\left(x - \frac{1}{4}\right) + \frac{1}{8}, & \text{if } \frac{1}{4} < x \leq \frac{1}{2}; \\ x, & \text{if } x > \frac{1}{2}; \\ 0 & \text{otherwise.} \end{cases}$$

Let $u : (0, \eta) \to \mathbb{R}_+$ be the function given in the integrability Condition (6) with $\varepsilon = \bar{r} = \infty$, and let h be given in Definition 6. Then the following statements hold:

i. Functions u and h are given by

$$u(s) = \begin{cases} 2, & \text{if } 0 < s \leq \frac{1}{8}; \\ \frac{2}{3}, & \text{if } \frac{1}{8} < s < \frac{1}{2}; \\ 1, & \text{if } s \geq \frac{1}{2}. \end{cases} \text{ and } h(s) = \begin{cases} 2, & \text{if } 0 < s \leq \frac{1}{8}; \\ 1, & \text{if } s > \frac{1}{8}. \end{cases}$$

ii. The function h satisfies

$$h = \inf\{\overline{u} : \overline{u} : (0,\eta) \to \mathbb{R}_+ \text{ is a continuous and decreasing function with } \overline{u} \geq u\}.$$

iii. The exact modulus of f at $\overline{x} = 0$ with respect to U and η is

$$\tilde{\varphi}(t) = \begin{cases} 2t, & \text{if } 0 \leq t \leq \dfrac{1}{8}; \\ t + \dfrac{1}{8}, & \text{if } t > \dfrac{1}{8}. \end{cases}$$

Let $\overline{u}$ be a continuous and decreasing majorant of u and define $\varphi(t) = \int_0^t \overline{u}(s)ds$. Then $\tilde{\varphi} \leq \varphi$ on $(0, 1/8]$ and $\tilde{\varphi} < \varphi$ on $(1/8, \infty)$.

Proof.

i. The desired results follow from simple calculations.

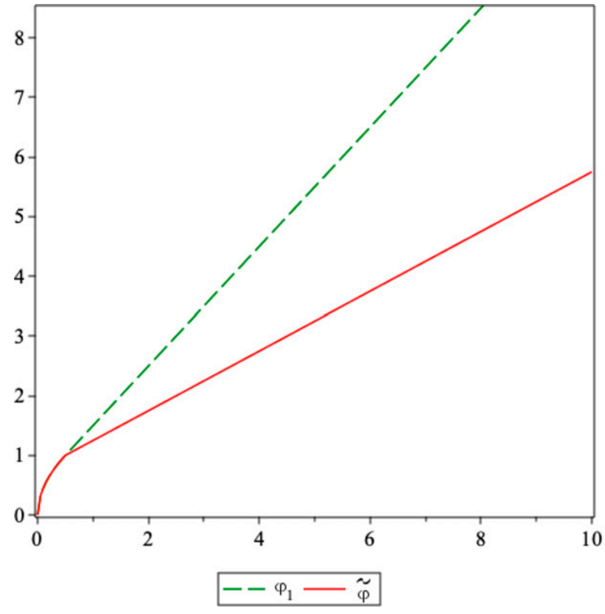
ii. Define $w = \inf\{\overline{u} : \overline{u} : (0,\eta) \to \mathbb{R}_+$ is a continuous and decreasing function with $\overline{u} \geq u\}$. Let $n \in \mathbb{N}$ and define $w_n : (0,\eta) \to \mathbb{R}_+$ by

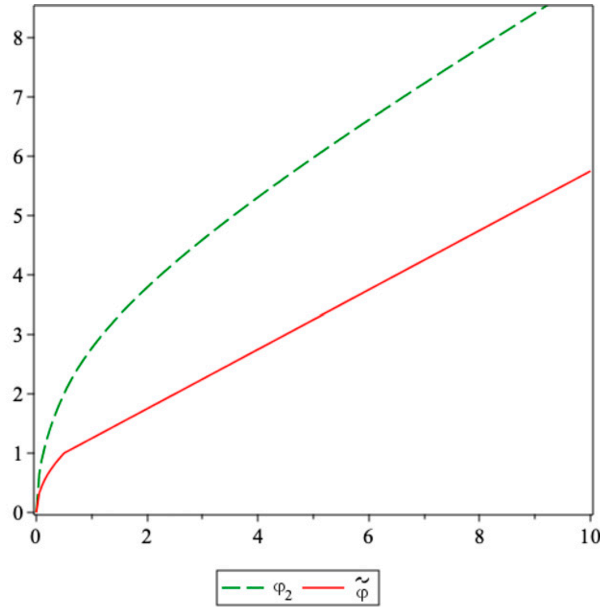
$$w_n(s) = \begin{cases} 2, & \text{if } 0 < s \leq \dfrac{1}{8}; \\ -n\left(s - \dfrac{1}{8}\right) + 2, & \text{if } \dfrac{1}{8} < s \leq \dfrac{1}{8} + \dfrac{1}{n}; \\ 1, & \text{if } s > \dfrac{1}{8} + \dfrac{1}{n}. \end{cases}$$

Then w_n is a decreasing and continuous majorant of u and $w_n \geq w$. Pick $s > 1/8$ and note that $\lim_{n\to\infty} w_n(s) = 1$. Then $w(s) \leq \lim_{n\to\infty} w_n(s) = 1$, which together with the fact that $w \geq 1$ yields $w(s) = 1$. On the other hand, for $s \leq 1/8$, we have $2 \leq w(s) \leq w_n(s) = 2$, which means $w(s) = 2$. Therefore, $w = h$ by (i).

iii. Integrating h yields the desired formula of $\tilde{\varphi}$. Statement(ii) implies that any continuous and decreasing majorant $\overline{u}$ of u satisfies $\overline{u} \geq h$ and there exists some $\varepsilon > 0$ such that $\overline{u}(s) > h(s)$ on $(\frac{1}{8}, \frac{1}{8} + \varepsilon)$. If there was $t_0 > 1/8$ such that $\varphi(t_0) = \tilde{\varphi}(t_0)$, then we would have $\overline{u} = h$ almost everywhere on $(0, t_0)$, which is absurd. □

Figure 1. (Color online) Plots of Example 4. Left: The exact modulus $\tilde{\varphi}$ and φ_1 in the limiting case, where $\overline{r} = r_0$. Right: The exact modulus $\tilde{\varphi}$ and φ_2.





We now compare the exact modulus with Facts 6 and 8 by recycling Example 1. On one hand, we shall see that the exact modulus is smaller than any BDLM desingularizing function given by Fact 6. On the other hand, we will show that the smallest desingularizing function obtained from the growth condition in Fact 8 is still bigger than the exact modulus.

Example 4. Consider the function f given in Example 1 with $\rho = 1$. Recall from Example 1 that the exact modulus of f at $\bar{x}$ with respect to $U = \mathbb{R}$ and $\eta = \infty$ is

$$\tilde{\varphi}(t) = \begin{cases} \sqrt{2t}, & \text{if } 0 \le t \le 1/2; \\ t/2 + 3/4, & \text{if } t > 1/2. \end{cases}$$

Moreover, the following statements hold:

i. Applying Fact 6 with $r_0 = 1/2$ and $\bar{r} \in (0, r_0)$ gives that f satisfies the concave KL property at $\bar{x} = 0$ with respect to $U = \mathbb{R}$, $\eta = \infty$ and

$$\varphi_1(t) = \begin{cases} \sqrt{2t}, & \text{if } t \le \bar{r}; \\ \sqrt{2\bar{r}} + \dfrac{1}{\sqrt{2\bar{r}}}(t - \bar{r}), & \text{if } t > \bar{r}. \end{cases}$$

Evidently $\tilde{\varphi} \le \varphi_1$, even in the limiting case where $\bar{r} = r_0$; see the left plot in Figure 1. There are other constructions; however, they are all bigger than the exact modulus $\tilde{\varphi}$; see Remark 7 for a detailed discussion.

ii. Define $m : \mathbb{R}_+ \to \mathbb{R}_+$ by

$$m(t) = \begin{cases} \dfrac{1}{2}t^2, & \text{if } 0 \le t \le 1; \\ 2t - \dfrac{3}{2}, & \text{if } t > 1. \end{cases}$$

Then Fact 8 implies that f has the concave KL property at zero with respect to $U = \mathbb{R}$, $\eta = \infty$ and

$$\varphi_2(t) = \begin{cases} 2\sqrt{2t}, & \text{if } t \le \dfrac{1}{2}; \\ \dfrac{t}{2} + \dfrac{3}{4}\ln(t) + \dfrac{7}{4} + \dfrac{3}{4}\ln(2), & \text{if } t > \dfrac{1}{2}, \end{cases}$$

the smallest one by Fact 8. However, $\tilde{\varphi} \le \varphi_2$; see the right plot in Figure 1.

Proof.

i. For $r \in (0, 1/2]$, $f(x) = r \Longleftrightarrow |x| = \sqrt{2r}$ and for $r \in (1/2, \infty)$, we have $f(x) = r \Longleftrightarrow |x| = 3/4 + r/2$. Then

$$u(r) = \frac{1}{\inf_{x \in [f = r]} \operatorname{dist}(0, \partial f(x))} = \begin{cases} \dfrac{1}{\sqrt{2r}}, & \text{if } 0 < r \le \dfrac{1}{2}; \\ \dfrac{1}{2}, & \text{if } r > \dfrac{1}{2}. \end{cases}$$

Noticing that u is continuous on $(0, r_0)$, we set the continuous majorant $\tilde{u}$ in Fact 6 to be u. The desired φ_1 then follows from applying Fact 6.

ii. Clearly all conditions in Fact 8 are satisfied. In particular, the equality $f(x) = m(\operatorname{dist}(x, \arg\min f)) = m(|x|)$ holds for all x, which means m is the largest possible modulus of the growth condition. The larger function m is, the smaller its inverse. Hence, $\varphi_2(t) = \int_0^t m^{-1}(s)/s \, ds$ is the smallest possible desingularizing function that one can get from Fact 8. The rest of the statement follows from a simple calculation. □

Remark 7. The function φ_1 given in Example 4 is indeed $\varphi_1(t) = \int_0^t \bar{u}(s) ds$, where

$$\bar{u}(r) = \begin{cases} u(r), & \text{if } r \le \bar{r}; \\ u(\bar{r}), & \text{if } r > \bar{r}, \end{cases}$$

which is a continuous and decreasing majorant of u, where u is given in the proof above. Replacing $\overline{u}$ by other such majorant of u certainly yields a different φ_1. However, notice that for this example, we have $\tilde{\varphi}(t) = \int_0^t u(s)ds$. Therefore, $\tilde{\varphi}(t) \le \int_0^t \overline{u}(s)ds = \varphi_1(t)$, no matter which majorant $\overline{u}$ we choose.

Despite the exact modulus $\tilde{\varphi}$ in Example 4 is smaller than the desingularizing function obtained from Fact 6, there is still some overlap. In what follows, we construct an example where the exact modulus of a nondifferentiable convex function is the strictly smaller one everywhere, except at the origin. Note that we shall compare the exact modulus to Fact 5 instead of Fact 6, as the former is more general.

Example 5. Let $r_1 = \pi^2/6 - 1$ and $r_{k+1} = r_k - 1/[k^2(k+1)]$ for $k \in \mathbb{N}$. Define for $k \in \mathbb{N}$ and $x > 0$

$$f(x) = \frac{1}{k}\left(x - \frac{1}{k}\right) + r_k, \ \forall x \in \left(\frac{1}{k+1}, \frac{1}{k}\right].$$

Let $f(-x) = f(x)$ for $x < 0$ and $f(0) = 0$. Then the following statements hold:

i. The function $f : [-1,1] \to [0, r_1]$ is continuous and convex with $\arg\min f = \{0\}$.

ii. The exact modulus of f at zero with respect to $U = [-1,1]$ and $\eta = r_1$ is a piecewise linear function $\tilde{\varphi} : [0, r_1] \to \mathbb{R}_+$ satisfying

$$\tilde{\varphi}(t) = k(t - r_{k+1}) + \sum_{i=k+1}^{\infty} i(r_i - r_{i+1}), \ \forall t \in (r_{k+1}, r_k], \ \forall k \in \mathbb{N},$$

and $\tilde{\varphi}(0) = 0$. Furthermore, every desingularizing function $\varphi : [0, r_1] \to \mathbb{R}_+$ obtained from Fact 5 satisfies $\varphi(t) > \tilde{\varphi}(t)$ on $(0, r_1]$.

Proof.

i. Note that we have

$$\lim_{k\to\infty} r_{k+1} = r_1 - \sum_{i=1}^{\infty} \frac{1}{i^2(i+1)} = r_1 - \sum_{i=1}^{\infty}\left(\frac{1}{i^2} - \frac{1}{i(i+1)}\right) = r_1 - r_1 = 0,$$

which implies that f is well defined and continuous at zero. To see the continuity at $x = 1/(k+1)$ for $k \in \mathbb{N}$, it suffices to observe that

$$\lim_{x\to\frac{1}{k+1}^+} f(x) = \frac{1}{k}\left(\frac{1}{k+1} - \frac{1}{k}\right) + r_k = r_{k+1} = f\left(\frac{1}{1+k}\right),$$

where the second last equality follows from the definition of r_{k+1}. Moreover, f is piecewise linear with increasing slope, then it is convex.

ii. Clearly we have $\partial f(1/k) = [1/k, 1/(k-1)]$ for every $k \in \mathbb{N}$ with $k \ge 2$. It follows easily that for $x \in [-1,1]$ with $1/(k+1) < |x| \le 1/k$, one has $\operatorname{dist}(0, \partial f(x)) = 1/k$. For $r \in (r_{k+1}, r_k]$, $r = f(x) \Longleftrightarrow |x| = k(r - r_k) + 1/k$. Elementary calculation yields

$$u(r) = \frac{1}{\inf_{x\in U\cap[f=r]} \operatorname{dist}(0, \partial f(x))} = k.$$

Note that $u \in L^1(0, r_1)$; hence, all conditions in Fact 5 are satisfied. Indeed,

$$\int_0^{r_1} u(s)ds = \sum_{k=1}^{\infty} k \cdot (r_k - r_{k+1}) = \sum_{k=1}^{\infty} \frac{1}{k(k+1)} \le \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty.$$

Then Bolte et al. [7, lemma 44] implies there exists a continuous and decreasing majorant $\overline{u}$ of u. The continuity of $\overline{u}$ ensures that for every k, there exists $\varepsilon_k > 0$ such that $\overline{u} > u$ on $(r_k, r_k + \varepsilon_k)$. Evidently the exact modulus $\tilde{\varphi}$ satisfies $\tilde{\varphi}(t) = \int_0^t u(s)ds$. Altogether, we conclude that the desingularizing function given by Fact 5 satisfies

$$\varphi(t) = \int_0^t \overline{u}(s)ds > \int_0^t u(s)ds = \tilde{\varphi}(t), \ \forall t \in (0, r_1].$$

Indeed, if there was $t_0 \in (0, r_1]$ such that $\varphi(t_0) = \tilde{\varphi}(t_0)$, then we would have $\overline{u} = u$ almost everywhere on $(0, t_0]$, which is absurd. □

4. The PALM Algorithm Revisited

In this section, we revisit the celebrated proximal alternating linearized minimization (PALM) algorithm. We will show that the exact modulus of the generalized concave KL property leads to the sharpest upper bound on the total length of trajectory of iterates generated by PALM.

4.1. The PALM Algorithm

Consider the following nonconvex and nonsmooth optimization model:

$$\min_{(x,y)\in\mathbb{R}^n\times\mathbb{R}^m} \Psi(x,y) = f(x) + g(y) + F(x,y),$$

where $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ and $g:\mathbb{R}^m\to\overline{\mathbb{R}}$ are proper and lsc and $F:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ is C^1. This model covers many optimization problems in practice; see Bolte et al. [5]. Bolte et al. [5] proposed the following algorithm to solve the problem.

PALM: Proximal Alternating Linearized Minimization

1. Initialization: Start with arbitrary $z_0 = (x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$.

2. For each $k = 0, 1, \ldots,$ generate a sequence $(z_k)_{k\in\mathbb{N}} = (x_k, y_k)_{k\in\mathbb{N}}$ as follows, where quantities $L_1(y_k)$ and $L_2(x_{k+1})$ will be given in (A2):

2.1. Take $\gamma_1 > 1$, set $c_k = \gamma_1 L_1(y_k)$ and compute

$$x_{k+1} \in \operatorname{Prox}^f_{c_k}\left(x_k - \frac{1}{c_k}\nabla_x F(x_k, y_k)\right). \tag{13}$$

2.2. Take $\gamma_2 > 1$, set $d_k = \gamma_2 L_2(x_{k+1})$ and compute

$$y_{k+1} \in \operatorname{Prox}^g_{d_k}\left(y_k - \frac{1}{d_k}\nabla_y F(x_{k+1}, y_k)\right). \tag{14}$$

The PALM algorithm is analyzed under the following blanket assumptions in Bolte et al. [5].

A1. Suppose that $\inf_{\mathbb{R}^n\times\mathbb{R}^m}\Psi > -\infty$, $\inf_{\mathbb{R}^n} f > -\infty$ and $\inf_{\mathbb{R}^m} g > -\infty$.

A2. For every fixed $y\in\mathbb{R}^m$, the function $x\longmapsto F(x,y)$ is $C^{1,1}_{L_1(y)}$, that is,

$$\|\nabla_x F(x_1, y) - \nabla_x F(x_2, y)\| \le L_1(y)\|x_1 - x_2\|,\ \forall x_1, x_2 \in \mathbb{R}^n.$$

Assume similarly that for every $x\in\mathbb{R}^n$, $y\longmapsto F(x,y)$ is $C^{1,1}_{L_2(x)}$.

A3. For $i = 1, 2$, there exist $\lambda_i^-, \lambda_i^+ > 0$ such that

$$\inf\{L_1(y_k) : k\in\mathbb{N}\} \ge \lambda_1^- \text{ and } \inf\{L_2(x_k) : k\in\mathbb{N}\} \ge \lambda_2^-,$$
$$\sup\{L_1(y_k) : k\in\mathbb{N}\} \le \lambda_1^+ \text{ and } \sup\{L_2(x_k) : k\in\mathbb{N}\} \le \lambda_2^+.$$

A4. Suppose that ∇F is Lipschitz continuous on bounded subsets of $\mathbb{R}^n\times\mathbb{R}^m$, that is, on every bounded subset $B_1\times B_2$ of $\mathbb{R}^n\times\mathbb{R}^m$, there exists $M > 0$ such that for all $(x_i, y_i) \in B_1\times B_2$, $i = 1, 2$,

$$\|\nabla F(x_1, y_1) - \nabla F(x_2, y_2)\| \le M\|(x_1 - x_2, y_1 - y_2)\|.$$

Fact 1 shows that PALM is well defined. Bolte et al. [5] showed that the PALM algorithm enjoys the following properties.

Lemma 4 (Bolte et al. [5, lemma 3]). *Suppose that (A1)–(A4) hold. Let $(z_k)_{k\in\mathbb{N}}$ be a sequence generated by PALM. Then the following hold:*

i. *The sequence $(\Psi(z_k))_{k\in\mathbb{N}}$ is decreasing and in particular*

$$\frac{\rho_1}{2}\|z_{k+1} - z_k\|^2 \le \Psi(z_k) - \Psi(z_{k+1}),\ \forall k \ge 0, \tag{15}$$

where $\rho_1 = \min\{(\gamma_1 - 1)\lambda_1^-, (\gamma_2 - 1)\lambda_2^-\}$.

ii. *We have $\sum_{k=1}^{\infty}\|z_{k+1} - z_k\|^2 < \infty$, and hence $\lim_{k\to\infty}\|z_{k+1} - z_k\| = 0$.*

Lemma 5 (Bolte et al. [5, lemma 4]). *Suppose that (A1)–(A4) hold and that $M > 0$ is the Lipschitz constant given in (A4). Let $(z_k)_{k\in\mathbb{N}}$ be a sequence generated by PALM, which is assumed to be bounded. For $k\in\mathbb{N}$, define*

$$A_x^k = c_{k-1}(x_{k-1} - x_k) + \nabla_x F(x_k, y_k) - \nabla_x F(x_{k-1}, y_{k-1}),$$
$$A_y^k = d_{k-1}(y_{k-1} - y_k) + \nabla_y F(x_k, y_k) - \nabla_y F(x_k, y_{k-1}).$$

Then $(A_x^k, A_y^k) \in \partial\Psi(x_k, y_k)$, *and*

$$\|(A_x^k, A_y^k)\| \le \|A_x^k\| + \|A_y^k\| \le (2M + 3\rho_2)\|z_k - z_{k-1}\|,\ \forall k \in \mathbb{N},$$

where $\rho_2 = \max\{\gamma_1\lambda_1^+, \gamma_2\lambda_2^+\}$.

Denote the set of subsequential limit points of $(z_k)_{k\in\mathbb{N}}$ by $\omega(z_0) = \{z \in \mathbb{R}^n \times \mathbb{R}^m : \exists (z_{k_q})_{q\in\mathbb{N}} \subseteq (z_k)_{k\in\mathbb{N}}, z_{k_q} \to z \text{ as } q \to \infty\}$. The following lemma summarizes useful properties of $(z_k)_{k\in\mathbb{N}}$ and $\omega(z_0)$, where Lemma 6(i) follows from the proof of Bolte et al. [5, lemma 5(i)].

Lemma 6 (Bolte et al. [5, lemma 5]). *Suppose that (A1)–(A4) hold. Let $(z_k)_{k\in\mathbb{N}}$ be a sequence generated by PALM, which is assumed to be bounded. Then the following assertions hold:*

i. *For every $z^* \in \omega(z_0)$ and $(z_{k_q})_{q\in\mathbb{N}}$ converging to z^*,*

$$\lim_{q\to\infty} \Psi(z_{k_q}) = \Psi(z^*).$$

Moreover, $\omega(z_0) \subseteq \operatorname{stat}\Psi$, where $\operatorname{stat}\Psi$ denotes the set of stationary points of Ψ.

ii. *We have $\lim_{k\to\infty}\operatorname{dist}(z_k, \omega(z_0)) = 0$.*

iii. *The set $\omega(z_0)$ is nonempty, compact, and connected.*

iv. *The objective function is constant on $\omega(z_0)$.*

4.2. The Sharpest Upper Bound for the Total Length of Trajectory of Iterates

In this subsection, we improve a result by Bolte et al. [5, theorem 1]. We begin with a technical lemma, which is a sharper version of Bolte et al. [5, lemma 6].

Lemma 7. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper lsc and let $\mu \in \mathbb{R}$. Let $\Omega \subseteq \operatorname{dom}\partial f$ be a nonempty compact set on which $f(x) = \mu$ for all $x \in \Omega$. Suppose that f has the pointwise generalized concave KL property at each $x \in \Omega$. Let ε, $\eta > 0$ and concave $\varphi \in \Phi_\eta$ be those given in Proposition 1. Set $U = \Omega_\varepsilon$ and define $h : (0, \eta) \to \mathbb{R}_+$ by*

$$h(s) = \sup\{\operatorname{dist}^{-1}(0, \partial f(x)) : x \in U \cap [0 < f - \mu < \eta], s \le f(x) - \mu\}.$$

Then the function $\tilde{\varphi} : [0, \eta) \to \mathbb{R}_+ : t \mapsto \int_0^t h(s)ds,\ \ \forall t \in (0, \eta)$, with $\tilde{\varphi}(0) = 0$, is well defined, concave and belongs to Φ_η. The function f has the setwise generalized concave KL property on Ω with respect to U, η, and $\tilde{\varphi}$. Consequently,

$$\tilde{\varphi} = \inf\{\varphi \in \Phi_\eta : \varphi \text{ is a concave desingularizing function of } f \text{ on } \Omega \text{ with respect to } U \text{ and } \eta\}.$$

We say $\tilde{\varphi}$ is the exact modulus of the setwise generalized concave KL property of f on Ω with respect to U and η.

Proof. Apply a similar argument as in Proposition 2. □

The following theorem provides the "sharpest" upper bound for the total length of the trajectory of iterates generated by PALM, which improves Bolte et al. [5, theorem 1]. The notion of sharpest will be specified later in Remark 8. Our proof follows a similar approach as in Bolte et al. [5, theorem 1] but makes use of the exact modulus of the setwise generalized concave KL property.

Theorem 1. *Suppose that the objective function Ψ is a generalized concave KL function such that (A1)–(A4) hold. Let $(z_k)_{k\in\mathbb{N}}$ be a sequence generated by PALM, which is assumed to be bounded. Then the following assertions hold:*

i. *The sequence $(z_k)_{k\in\mathbb{N}}$ converges to a stationary point z^* of objective function Ψ.*

ii. *The sequence $(z_k)_{k\in\mathbb{N}}$ has finite length. More precisely, there exist $l \in \mathbb{N}$, $\eta \in (0, \infty]$ and $\tilde{\varphi} \in \Phi_\eta$ such that for $p \ge l + 1$ and every $q \in \mathbb{N}$,*

$$\sum_{k=p}^{p+q} \|z_{k+1} - z_k\| \le C \cdot \tilde{\varphi}(\Psi(z_p) - \Psi(z^*)) + \|z_p - z_{p-1}\|. \tag{16}$$

Therefore,

$$\sum_{k=1}^{\infty} \|z_{k+1} - z_k\| \le A + C \cdot \tilde{\varphi}(\Psi(z_{l+1}) - \Psi(z^*)) < \infty, \tag{17}$$

where $A = \|z_{l+1} - z_l\| + \sum_{k=1}^{l}\|z_{k+1} - z_k\| < \infty$ and $C = 2(2M + 3\rho_2)/\rho_1$.

Proof. Because $(z_k)_{k\in\mathbb{N}}$ is bounded, there exists a convergent subsequence, say $z_{k_q} \to z^* \in \omega(z_0)$. Then Lemma 6(i) implies $\lim_{q\to\infty}\Psi(z_{k_q}) = \Psi(z^*)$ and $z^* \in \operatorname{crit}\Psi$. Because $(\Psi(z_k))_{k\in\mathbb{N}}$ is a decreasing sequence by Lemma 4, we have $\lim_{k\to\infty}\Psi(z_k) = \lim_{q\to\infty}\Psi(z_{k_q}) = \Psi(z^*)$.

We will show that $(z_k)_{k\in\mathbb{N}}$ converges to z^*, and along the way we also establish (16) and (17). We proceed by considering two cases.

Case 1. If there exits l such that $\Psi(z_l) = \Psi(z^*)$, then by the decreasing property of $(\Psi(z_k))_{k\in\mathbb{N}}$, one has $\Psi(z_{l+1}) = \Psi(z_l)$ and therefore $z_l = z_{l+1}$ by (15). Hence, by induction, we conclude that $\lim_{k\to\infty} z_k = z^*$. The desired assertion follows immediately.

Case 2. Now we consider the case where $\Psi(z^*) < \Psi(z_k)$ for all $k \in \mathbb{N}$. By Lemma 6 and assumption, Ψ is a generalized concave KL function that is constant on compact set $\omega(z_0)$. Invoking Lemma 7 shows that there exist $\varepsilon > 0$ and $\eta > 0$ such that the exact modulus of the setwise generalized concave KL property on $\Omega = \omega(z_0)$ with respect to $U = \Omega_\varepsilon$ and η exists, which is denoted by $\tilde{\varphi}$. Hence, for every $z \in \Omega_\varepsilon \cap [0 < \Psi - \Psi(z^*) < \eta]$,

$$\tilde{\varphi}'_-(\Psi(z) - \Psi(z^*)) \cdot \operatorname{dist}(0, \partial\Psi(z)) \geq 1. \tag{18}$$

Because $\lim_{k\to\infty}\Psi(z_k) = \Psi(z^*)$, there exists some $l_1 > 0$ such that $0 < \Psi(z_k) - \Psi(z^*) < \eta$ for $k > l_1$. On the other hand, Lemma 6(ii) shows that there exists $l_2 > 0$ such that $\operatorname{dist}(z_k, w(z_0)) < \varepsilon$ for $k > l_2$. Altogether, we conclude that for $k > l = \max\{l_1, l_2\}$, $z_k \in \Omega_\varepsilon \cap [0 < \Psi - \Psi(z^*) < \eta]$ and

$$\tilde{\varphi}'_-(\Psi(z_k) - \Psi(z^*)) \cdot \operatorname{dist}(0, \partial\Psi(z_k)) \geq 1. \tag{19}$$

It follows from Lemma 5 that $\operatorname{dist}(0, \partial\Psi(z_k)) \leq \|(A_x^k, A_y^k)\| \leq (2M + 3\rho_2)\|z_k - z_{k-1}\|$. Hence, one has from (19) that for $k > l$,

$$\tilde{\varphi}'_-(\Psi(z_k) - \Psi(z^*)) \geq \operatorname{dist}^{-1}(0, \partial\Psi(z_k)) \geq \frac{1}{2M + 3\rho_2}\|z_k - z_{k-1}\|^{-1}. \tag{20}$$

Note that $\|z_k - z_{k-1}\| \neq 0$. Otherwise Lemma 5 would imply that

$$\operatorname{dist}(0, \partial\Psi(z_k)) \leq (2M + 3\rho_2)\|z_k - z_{k-1}\| = 0,$$

which contradicts with (19). Applying Lemma 1(ii) to $\tilde{\varphi}$ with $s = \Psi(z_{k+1}) - \Psi(z^*)$ and $t = \Psi(z_k) - \Psi(z^*)$, one obtains for $k > l$

$$\begin{aligned} \frac{\tilde{\varphi}(\Psi(z_k) - \Psi(z^*)) - \tilde{\varphi}(\Psi(z_{k+1}) - \Psi(z^*))}{\Psi(z_k) - \Psi(z_{k+1})} &\geq \tilde{\varphi}'_-(\Psi(z^k) - \Psi(z^*)) \\ &\geq \frac{1}{2M + 3\rho_2}\|z_k - z_{k-1}\|^{-1}. \end{aligned} \tag{21}$$

For the sake of simplicity, we set

$$\Delta_{p,q} = \tilde{\varphi}(\Psi(z_p) - \Psi(z^*)) - \tilde{\varphi}(\Psi(z_q) - \Psi(z^*)).$$

Then (21) can be rewritten as

$$\Psi(z_k) - \Psi(z_{k+1}) \leq \|z_k - z_{k-1}\| \cdot \Delta_{k,k+1} \cdot (2M + 3\rho_2). \tag{22}$$

Furthermore, Lemma 4(i) gives

$$\|z_{k+1} - z_k\|^2 \leq \frac{2}{\rho_1}[\Psi(z_k) - \Psi(z_{k+1})] \leq C\Delta_{k,k+1}\|z_k - z_{k-1}\|,$$

where $C = 2(2M + 3\rho_2)/\rho_1 \in (0, \infty)$. By the geometric mean inequality $2\sqrt{\alpha\beta} \leq \alpha + \beta$ for $\alpha, \beta \geq 0$, one gets for $k > l$

$$2\|z_{k+1} - z_k\| \leq C\Delta_{k,k+1} + \|z_k - z_{k-1}\|. \tag{24}$$

Let $p \geq l+1$. For every $q \in \mathbb{N}$, summing up (24) from p to $p+q$ yields

$$\begin{aligned}2\sum_{k=p}^{p+q}\|z_{k+1}-z_k\| &\leq C\sum_{k=p}^{p+q}\Delta_{k,k+1}+\sum_{k=p}^{p+q}\|z_k-z_{k-1}\|+\|z_{p+q+1}-z_{p+q}\|\\ &= C\Delta_{p,p+q+1}+\sum_{k=p}^{p+q}\|z_{k+1}-z_k\|+\|z_p-z_{p-1}\|\\ &\leq C\tilde{\varphi}(\Psi(z_p)-\Psi(z^*))+\sum_{k=p}^{p+q}\|z_{k+1}-z_k\|+\|z_p-z_{p-1}\|,\end{aligned}$$

where the last inequality holds because $\tilde{\varphi} \geq 0$. Hence, for $q \in \mathbb{N}$

$$\sum_{k=p}^{p+q}\|z_{k+1}-z_k\| \leq C\cdot\tilde{\varphi}(\Psi(z_p)-\Psi(z^*))+\|z_p-z_{p-1}\|,$$

which proves (16). By taking $q \to \infty$,

$$\sum_{k=1}^{\infty}\|z_{k+1}-z_k\| \leq \sum_{k=1}^{p-1}\|z_{k+1}-z_k\|+C\cdot\tilde{\varphi}(\Psi(z_p)-\Psi(z^*))+\|z_p-z_{p-1}\|,$$

from which (17) readily follows by setting $p = l+1$.

Now let $p \geq l+1$, where l is the index given in assertion (i), and let $q \in \mathbb{N}$. Then

$$\|z_{p+q}-z_p\| \leq \sum_{k=p}^{p+q-1}\|z_{k+1}-z_k\| \leq \sum_{k=p}^{p+q}\|z_{k+1}-z_k\|.$$

Recall that $\tilde{\varphi}(t) \to 0$ as $t \to 0^+$, $\Psi(z_k)-\Psi(z^*) \to 0$ and $\|z_{k+1}-z_k\| \to 0$ as $k \to \infty$. Invoking (16), one obtains that

$$\|z_{p+q}-z_p\| \leq C\cdot\tilde{\varphi}(\Psi(z_p)-\Psi(z^*))+\|z_p-z_{p-1}\| \to 0, p \to \infty,$$

meaning that $(z_k)_{k\in\mathbb{N}}$ is Cauchy and hence convergent. Because $z_{k_q} \to z^*$, we conclude that $z_k \to z^*$. □

Remark 8. The bound for the total length of iterates (17) is the sharpest, in the sense that it is the smallest one can get by using the usual KL convergence analysis. Assuming that the objective function Ψ has the concave KL property, Bolte et al. [5, theorem 1] showed that

$$\sum_{k=1}^{\infty}\|z_{k+1}-z_k\| \leq A+C\cdot\varphi(\Psi(z_{l+1})-\Psi(z^*)), \tag{25}$$

where $\varphi(t)$ is in our terminology a concave desingularizing function for the setwise concave KL property of Ψ on $\Omega = w(z_0)$ with respect to Ω_ε and $\eta > 0$; see Bolte et al. [5, lemma 6]. Note that A and C are fixed. Then we learn from (25) that the smaller the $\varphi(t)$ is, the sharper the upper bound becomes. According to Lemma 7, $\tilde{\varphi}$ is the smallest among all possible φ. Hence, the upper bound given by (17) is the sharpest.

5. Conclusion

In this work, we introduced the generalized concave KL property and its exact modulus, which answers the open question (1). Our results open the door for obtaining sharp results of algorithms that adopt the concave KL assumption. We conclude this paper with some future directions:

- Compute or at least estimate the exact modulus of the generalized concave KL property for concrete optimization models.
- One way to estimate the exact modulus is applying calculus rules of the generalized concave KL property. Li and Pong [10] and Yu et al. [21] developed several calculus rules of the concave KL property, in the case where desingularizing functions take the specific form $\varphi(t) = c\cdot t^{1-\theta}$, where $c > 0$ and $\theta \in [0,1)$. However, the exact modulus has various forms depending on the given function, which requires us to obtain general calculus rules without assuming desingularizing functions of any specific form.

Acknowledgments
The authors thank Dr. Heinz H. Bauschke for many useful discussions and suggesting this research problem. The authors are also grateful to editor Michael P. Friedlander and anonymous referees for their valuable comments and suggestions on terminologies that improved this manuscript significantly.

Endnotes
[1] In the remainder of this paper, we shall simply refer this pointwise definition as "the concave KL property" for the sake of simplicity. However, we would like to remind readers that the concave KL property is originally introduced as a property about function values instead of points; see Bolte et al. [6, theorem 14].

[2] For simplicity, we shall omit adjectives "pointwise" and "setwise" whenever there is no ambiguity. We also remind readers that one can define similarly generalized concave KL property around function values, which, however, will not be treated in the paper.

[3] Note that such majorant may not exist beyond the convex or semialgebraic case.

References
[1] Attouch H, Bolte J, Redont P, Soubeyran A (2010) Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Math. Oper. Res.* 35(2):438–457.
[2] Banert S, Boţ RI (2019) A general double-proximal gradient algorithm for d.c. programming. *Math. Program.* 178(1):301–326.
[3] Bauschke HH, Combettes PL (2017) *Convex Analysis and Monotone Operator Theory in Hilbert Spaces* (Springer, Cham).
[4] Bolte J, Daniilidis A, Lewis A (2007) The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J. Optim.* 17(4):1205–1223.
[5] Bolte J, Sabach S, Teboulle M (2014) Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Programming* 146(1):459–494.
[6] Bolte J, Daniilidis A, Lewis A, Shiota M (2007) Clarke subgradients of stratifiable functions. *SIAM J. Optim.* 18(2):556–572.
[7] Bolte J, Daniilidis A, Ley O, Mazet L (2010) Characterizations of Łojasiewicz inequalities: Subgradient flows, talweg, convexity. *Trans. Amer. Math. Soc.* 362(6):3319–3363.
[8] Bolte J, Nguyen TP, Peypouquet J, Suter BW (2017) From error bounds to the complexity of first-order descent methods for convex functions. *Math. Programming* 165(2):471–507.
[9] Kurdyka K (1998) On gradients of functions definable in o-minimal structures. *Ann. Inst. Fourier (Grenoble)* 48(3):769–783.
[10] Li G, Pong TK (2018) Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods. *Foundations Comput. Math.* 18(5):1199–1232.
[11] Liu T, Pong TK, Takeda A (2019) A refined convergence analysis of pDCA$_e$ with applications to simultaneous sparse recovery and outlier detection. *Comput. Optim. Appl.* 73:69–100.
[12] Łojasiewicz S (1963) Une propriété topologique des sous-ensembles analytiques réels. *Les équations aux Dérivées Partielles* 117:87–89.
[13] Mordukhovich B (2006) *Variational Analysis and Generalized Differentiation I: Basic Theory* (Springer-Verlag, Berlin).
[14] Ochs P, Chen Y, Brox T, Pock T (2014) iPiano: Inertial proximal algorithm for nonconvex optimization. *SIAM J. Imaging Sci.* 7(2):1388–1419.
[15] Pugh CC (2015) *Real Mathematical Analysis* (Springer, Cham).
[16] Rockafellar RT (1970) *Convex Analysis* (Princeton University Press, Princeton, NJ).
[17] Rockafellar RT, Wets RJB (1998) *Variational Analysis* (Springer-Verlag, Berlin).
[18] Stromberg KR (2015) *An Introduction to Classical Real Analysis* (American Mathematical Society, Providence, RI).
[19] Wen B, Chen X, Pong TK (2018) A proximal difference-of-convex algorithm with extrapolation. *Comput. Optim. Appl.* 69:297–324.
[20] Won JH, Xu J, Lange K (2019) Projection onto Minkowski sums with application to constrained learning. *Proc. 36th Internat. Conf. Machine Learning*, vol. 97 (PMLR, Cambridge, MA), 3642–3651.
[21] Yu P, Li G, Pong TK (2019) Deducing Kurdyka-Łojasiewicz exponent via inf-projection. Preprint, submitted February 10, https://arxiv.org/abs/1902.03635.
[22] Yu P, Pong TK, Lu Z (2021) Convergence rate analysis of a sequential convex programming method with line search for a class of constrained difference-of-convex optimization problems. *SIAM J. Optim.* 31(3):2024–2054.