The exact modulus of the generalized Kurdyka-Łojasiewicz property

by

Ziyuan Wang

B.Sc., Chongqing Jiaotong University, 2018

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE COLLEGE OF GRADUATE STUDIES

(Mathematics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Okanagan)

August 2020

The following individuals certify that they have read, and recommend to the College of Graduate Studies for acceptance, a thesis/dissertation entitled:

The exact modulus of the generalized Kurdyka-Łojasiewicz property

submitted by

Ziyuan Wang

in partial fulfilment of the requirements of the degree of

Master of Science

Dr. Shawn X. Wang, I. K. Barber School of Arts & Sciences
Supervisor

Dr. Heinz Bauschke, I. K. Barber School of Arts & Sciences
Supervisory Committee Member

Dr. John Braun, I. K. Barber School of Arts & Sciences
Supervisory Committee Member

Dr. Jahangir Hossain, School of Engineering
University Examiner

Abstract

This work aims to provide a self-contained analysis of the Kurdyka-Łojasiewicz (KL) property within the framework of nonsmooth analysis. Our work focuses on two aspects.

On one hand, we introduce the generalized KL property, a new concept that generalizes the classic KL property by employing nonsmooth desingularizing functions. Examples and calculus rules for this generalized notion are given. Our results are new and extend the classic KL property.

On the other hand, by introducing the exact modulus of the generalized KL property, we provide an answer to the open question: "*What is the optimal desingularizing function?*", which fills a gap in the current literature. The exact modulus is designed to be the smallest among all possible desingularizing functions. Examples are given to illustrate this pleasant property. We also provide ways to determine or at least estimate the exact modulus. In turn, we obtain explicit formulae for the optimal desingularizing function of locally convex continuously differentiable functions and polynomials on the line, which is usually considered to be challenging. Furthermore, by using the exact modulus, we find the sharpest upper bound for the trajectory of iterates generated the celebrated PALM algorithm.

Lay Summary

The Kurdyka-Łojasiewicz property implies that a given function can be "sharpened" around a point and plays a key role in the convergence analysis of many optimization algorithms. When verifying the KL property, one needs to find a desingularizing function to "sharpen" the given function. On one hand, we developed various explicit formulae for the desingularizing functions, which is usually considered to be challenging. On the other hand, by introducing a new concept called the exact modulus, we provided a way to find the optimal desingularizing function. In turn, a celebrated optimization algorithm is improved.

Table of Contents

List of Tables

List of Figures

Glossary of Notation and Symbols

Sets

$\mathbb{R}^n$	The n-dimensional Euclidean space	2
$\mathbb{R}^n \times \mathbb{R}^m$	The product space of $\mathbb{R}^n$ and $\mathbb{R}^m$	2
$\overline{\mathbb{R}}$	The extended real line $(-\infty, \infty]$	2
$\mathbb{N}$	The set of positive natural numbers $\{1, 2, 3, \ldots\}$	2
$\mathbb{B}(\bar{x}; r)$	Open ball centered at $\bar{x} \in \mathbb{R}^n$ with radius $r > 0$	2
$\mathbb{B}_{\mathbb{R}^n}$	The Euclidean closed unit ball in $\mathbb{R}^n$	2
$[x, y]$	The line segment from $x \in \mathbb{R}^n$ to $y \in \mathbb{R}^n$	2
$\operatorname{int} K$	Interior of a set $K \subseteq \mathbb{R}^n$	2
$\operatorname{cl} K$	Closure of a set $K \subseteq \mathbb{R}^n$	2

Functions and operators

$\operatorname{dist}(\cdot, K)$	Distance function of a set $K \subseteq \mathbb{R}^n$	2
$\delta_K(\cdot)$	Indicator function of a set $K \subseteq \mathbb{R}^n$	2
$\operatorname{dom} f$	Domain of $f : \mathbb{R}^n \to \overline{\mathbb{R}}$	2
$\mathcal{W}(\cdot)$	The Lambert $\mathcal{W}$ function	7
∇f	Gradient operator of $f : \mathbb{R}^n \to \overline{\mathbb{R}}$	9
$\hat{\partial} f$	Fréchet subgradient operator of $f : \mathbb{R}^n \to \overline{\mathbb{R}}$	15
∂f	Limiting subgradient operator of $f : \mathbb{R}^n \to \overline{\mathbb{R}}$	16
$\operatorname{prox}_\lambda^f$	Proximal operator of $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ with parameter λ	20

Abbreviations

KL	The Kurdyka-Łojasiewicz property	1
PALM	Proximal alternating linearized minimization algorithm	1
D. C.	Functions representable as differences of convex functions	34

Acknowledgments

To Shawn: It is a privilege to work with you. The last two years have been tough undoubtedly, but also the most incredible study experience I have ever had, which made up my mind of choosing a career in mathematics. I deeply appreciate your patient guidance, constant encouragement and all those questions that made me want to sweat, without which this thesis could not be completed. Thank you!

I would like to express my gratitude to Dr. Heinz Bauschke and Dr. John Braun for being my supervisory committee members, and Dr. Jahangir Hossain for being the university examiner. In particular, I deeply thank Dr. Heinz Bauschke for his many helpful comments and continuous support throughout the completion of this thesis.

Lots of thanks to Hui, Manish, Salha, Shambhavi, Walaa and Wei, my friends and colleagues in the COCANA lab, for all the discussions and jokes we had together. I also would like to thank Mrs. Hilda, who kindly lent me a room when I arrived Canada for the first time.

Many many thanks to all my friends, in particular Chris, Dongying, Grace, Ruichao, Yiping and You Rao, for our weekend hotpot parties and all the chats from cohomology to the Game of Thrones that keep me from becoming, I quote, a "lengkuwuqing de shuxue jiqi". Special thanks to You, who drove me to buy groceries every two weeks during the COVID-19 lockdown.

Finally, I would like to thank all members in Unit 5 for their continuous support that makes UBC such a wonderful place to study.

Dedication

To my parents and grandmother.

Chapter 1

What to expect

This thesis aims to provide a self-contained analysis of *the Kurdyka-Łojasiewicz property* (KL) and address several challenging questions arose in this area within the framework of nonsmooth analysis. The notion of the KL property is originated from *algebraic geometry*, but has thrived in optimization over the past decade. For instance, Bolte et al. [9] showed that the KL property guarantees that every sequence generated by the *proximal alternating linearized minimization* (PALM) algorithm enjoys finite length property. When verifying the KL property, one needs to find a *desingularizing function*, which is concave and continuously differentiable on the interior of its domain, to "sharpen" the given function. Most published articles emphasize desingularizing functions of the form $\varphi(t) = c \cdot t^{1-\theta}$ for some $c > 0$ and $\theta \in [0, 1)$.

Tools from real analysis and nonsmooth analysis are introduced in Chapter 2 and Chapter 3, respectively. We provide a literature review and an introduction to challenges and open questions about the KL property in Chapter 4.

Our main results are in Chapters 5–8. In Chapter 5, beginning with functions on the real line, we provide several explicit formulae for the desingularizing functions, which is usually considered to be challenging. We introduce the generalized KL property in Chapter 6, which generalizes the classic KL property by employing nonsmooth desingularizing functions. Concrete examples and various calculus rules of this generalized notion are given. In Chapter 7, we provide an answer to the open question: "*What is the optimal desingularizing function?*" by defining the exact modulus of the generalized KL property. We will show that the exact modulus takes various forms depending on the given function, and it is the smallest among all possible desingularizing functions. Examples are also given to illustrate this pleasant property. Furthermore, we show that there exists $f : \mathbb{R} \to \mathbb{R}$ who has the KL property but the desingularizing function cannot take the usual form $\varphi(t) = c \cdot t^{1-\theta}$. Finally, in Chapter 8, we revisit the celebrated PALM algorithm and investigate the algorithmic impact of the exact modulus.

Chapter 2

Elements in real analysis

Throughout this thesis,

$$\mathbb{R}^n \text{ is the real Euclidean space}$$

equipped with dot product $\langle x, y\rangle = x^T y$ and the Euclidean norm $\|x\| = \sqrt{\langle x, x\rangle}$.

$$\mathbb{R}^n \times \mathbb{R}^m \text{ is the product space}$$

equipped with norm $\|(x, y)\|_{\mathbb{R}^n\times\mathbb{R}^m} = \sqrt{\|x\|^2_{\mathbb{R}^n} + \|y\|^2_{\mathbb{R}^m}}$. Denote by $\mathbb{N}$ the set of positive natural numbers, i.e., $\mathbb{N} = \{1, 2, 3, \ldots\}$.

We denote by $\overline{\mathbb{R}}$ the extended real line $(-\infty, \infty]$. The effective domain of $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is

$$\operatorname{dom} f = \{x \in \mathbb{R}^n : f(x) < \infty\},$$

and f is *proper*, if $\operatorname{dom} f \neq \emptyset$. The function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be *Lipschitz smooth* with modulus $L > 0$, if $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$ for every $x, y \in \operatorname{int}\operatorname{dom} f$, which we regime to be nonempty.

We will use the following notations of line segments:

$$[x, y] = \{\lambda x + (1 - \lambda)y : \lambda \in [0, 1]\}, (x, y) = \{\lambda x + (1 - \lambda)y : \lambda \in (0, 1)\}.$$

The open ball centered at $\bar{x}$ with radius r is

$$\mathbb{B}(\bar{x}; r) = \{x \in \mathbb{R}^n : \|x - \bar{x}\| < r\},$$

and we use $\mathbb{B}_{\mathbb{R}^n}$ to denote the closed unit ball in $\mathbb{R}^n$. For a subset $K \subseteq \mathbb{R}^n$, the interior and closure of K are defined by

$$\operatorname{int} K = \{x \in \mathbb{R}^n : \mathbb{B}(x; \varepsilon) \subseteq K \text{ for some } \varepsilon > 0\}.$$

$$\operatorname{cl} K = \{x \in \mathbb{R}^n : \mathbb{B}(x; \varepsilon) \cap K \neq \emptyset, \forall \varepsilon > 0\},$$

respectively. The distance function of $K \subseteq \mathbb{R}^n$ is $\operatorname{dist}(\cdot, K) : \mathbb{R}^n \to \overline{\mathbb{R}}$,

$$x \mapsto \inf\{\|x - y\| : y \in K\},$$

where $\operatorname{dist}(x, K) \equiv \infty$ if $K = \emptyset$. The indicator function of K is given by

$$\delta_K(x) = \begin{cases} 0, & x \in K; \\ \infty, & x \notin K. \end{cases}$$

2.1 Auxiliary results

Fact 2.1. *Let $x, y \in \mathbb{R}^n$. Then the following inequalities hold:*

(i) $\|x \pm y\|^2 = \|x\|^2 + \|y\|^2 \pm 2\langle x, y\rangle$;

(ii) $|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$;

Proof. (i) $\|x+y\|^2 = \langle x+y, x+y\rangle = \langle x, x+y\rangle + \langle y, x+y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle = \|x\|^2 + \|y\|^2 + 2\langle x, y\rangle$. For the other case, replace y by $-y$ and invoke the linearity of inner product.

(ii) If $x = 0$ or $y = 0$, then the desired inequality holds immediately. For nonzero vectors $x, y \in \mathbb{R}^n$, invoking (i) yields

$$\begin{aligned} 0 \leq \left\| x - \frac{\langle x, y\rangle}{\|y\|^2} y \right\|^2 &= \|x\|^2 + \frac{|\langle x, y\rangle|^2}{\|y\|^2} - 2\frac{|\langle x, y\rangle|^2}{\|y\|^2} \\ &= \|x\|^2 - \frac{|\langle x, y\rangle|^2}{\|y\|^2}, \end{aligned}$$

which is equivalent to $|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$. □

Fact 2.2. *Let $\alpha, \beta \geq 0$. Then we have*

$$\max\{\alpha, \beta\} = \frac{1}{2}(\alpha + \beta) + \frac{1}{2}|\alpha - \beta| \leq \alpha + \beta.$$

Proof. Suppose first that $\alpha \geq \beta$. Then we have

$$\frac{1}{2}(\alpha + \beta) + \frac{1}{2}|\alpha - \beta| = \frac{1}{2}(\alpha + \beta) + \frac{1}{2}(\alpha - \beta) = \alpha = \max\{\alpha, \beta\}.$$

On the other hand, if $\alpha < \beta$ then one has

$$\frac{1}{2}(\alpha + \beta) + \frac{1}{2}|\alpha - \beta| = \frac{1}{2}(\alpha + \beta) + \frac{1}{2}(\beta - \alpha) = \beta = \max\{\alpha, \beta\}.$$

Hence one concludes that

$$\max\{\alpha, \beta\} = \frac{1}{2}(\alpha + \beta) + \frac{1}{2}|\alpha - \beta| \leq \frac{1}{2}(\alpha + \beta) + \frac{1}{2}\left(|\alpha| + |\beta|\right) = \alpha + \beta,$$

where the last equality holds because $\alpha, \beta \geq 0$. □

Definition 2.3. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}^m$ and let $(s_n)_{n\in\mathbb{N}}$ be a sequence of real numbers.

(i) The sequence $(x_n)_{n\in\mathbb{N}}$ is said to be *Cauchy*, if for every $\varepsilon > 0$, there exits $N > 0$ such that $\|x_n - x_m\| \leq \varepsilon$ whenever $n, m > N$.

(ii) We say that $\bar{x} \in \mathbb{R}^m$ is a *limit point* of the sequence $(x_n)_{n\in\mathbb{N}}$, if there exists a subsequence $(x_{n_k})_{k\in\mathbb{N}} \subseteq (x_n)_{n\in\mathbb{N}}$ such that $x_{n_k} \to \bar{x}$ as $k \to \infty$.

(iii) The *upper and lower limits* of $(s_n)_{n\in\mathbb{N}}$ are values in $[-\infty, \infty]$ defined by

$$\limsup_{n\to\infty} s_n = \lim_{N\to\infty} \sup\{s_n : n > N\} = \inf_{N>0} \sup_{n>N} s_n \tag{2.1}$$

and

$$\liminf_{n\to\infty} s_n = \lim_{N\to\infty} \inf\{s_n : n > N\} = \sup_{N>0} \inf_{n>N} s_n. \tag{2.2}$$

Fact 2.4. *Let $(s_n)_{n\in\mathbb{N}}$ and $(t_n)_{n\in\mathbb{N}}$ be sequences of real numbers, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}^m$. Then the following properties hold:*

(i) *$(x_n)_{n\in\mathbb{N}}$ is convergent if and only if it is Cauchy.*

(ii) *The upper and lower limits $(s_n)_{n\in\mathbb{N}}$ and $(t_n)_{n\in\mathbb{N}}$ obey $\limsup_{n\to\infty} -s_n = -\liminf_{n\to\infty} s_n$ and*

$$\limsup_{n\to\infty}(s_n + t_n) \leq \limsup_{n\to\infty} s_n + \limsup_{n\to\infty} t_n,$$
$$\liminf_{n\to\infty}(s_n + t_n) \geq \liminf_{n\to\infty} s_n + \liminf_{n\to\infty} t_n.$$

(iii) *Let $(s_n)_{n\in\mathbb{N}}$ be a bounded sequence and denote by S the set of limit points of $(s_n)_{n\in\mathbb{N}}$. Then $S \neq \emptyset$ and*

$$\limsup_{n\to\infty} s_n = \sup S,$$

while

$$\liminf_{n\to\infty} s_n = \inf S.$$

Proof. See [22]. □

Definition 2.5. The *lower limit* of a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ at $\bar{x}$ is defined by

$$\liminf_{x\to\bar{x}} f(x) = \lim_{\varepsilon\to 0} \inf_{x\in\mathbb{B}(\bar{x};\varepsilon)} f(x)$$

The function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *lower semicontinuous* (lsc) at $\bar{x}$ if

$$f(\bar{x}) \leq \liminf_{x\to\bar{x}} f(x),$$

and lsc on $\operatorname{dom} f$ if this holds for every $\bar{x} \in \operatorname{dom} f$.

Definition 2.6. Let $K \subseteq \mathbb{R}^n$ be a subset. We say that

(i) K is *open*, if $K = \operatorname{int} K$; K is *closed*, if $K = \operatorname{cl} K$.

(ii) K is *compact*, if whenever K is contained in the union of a family of open sets, it is also contained in the union of a finite subfamily from that family.

(iii) K is *connected*, if K is not representable as the union of nonempty closed sets $A, B \subseteq K$ satisfying $A \cap B = \emptyset$.

(iv) K is *convex*, if $\lambda x + (1-\lambda)y \in K$, for every $x, y \in K$ and $\lambda \in (0,1)$.

Fact 2.7. *Every continuous map $f : \mathbb{R}^n \to \mathbb{R}$ attains maximum and minimum on a compact set $K \subseteq \mathbb{R}^n$.*

Proof. See [22]. □

Fact 2.8. *On finite-dimensional vector space $\mathbb{R}^n$, any norm $\|\cdot\|$ is equivalent to another norm $\|\cdot\|_0$, that is, there exist $c_1, c_2 > 0$ such that*

$$c_1 \|x\| \leq \|x\|_0 \leq c_2 \|x\|, \forall x \in \mathbb{R}^n.$$

Proof. See [12, Theorem 2.4.5]. □

2.2 Special functions

In this section we present properties of several functions that will be used in the sequel.

Lemma 2.9. *Let $P : \mathbb{R} \to \mathbb{R}$ be a real polynomial with degree $n \geq 0$. Then*

$$\lim_{x\to\infty} \frac{P(x)}{\exp(x)} = 0.$$

Proof. For every $n \geq 0$, by using the L'Hopital's rule, we have $\lim_{x\to\infty} \frac{x^n}{\exp(x)} = \lim_{x\to\infty} \frac{n!}{\exp(x)} = 0$. Suppose that $P(x) = \sum_{k=0}^{n} a_k x^k$. Then one has

$$\lim_{x\to\infty} \frac{P(x)}{\exp(x)} = \sum_{k=0}^{n} \lim_{x\to\infty} \frac{a_k x^k}{\exp(x)} = 0,$$

which completes the proof. □

Proposition 2.10. *Define* $f : \mathbb{R} \to \mathbb{R}$,

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

Then $f \in C^\infty$, *i.e., the* n*-th order derivative* $f^{(n)}$ *exists and is continuous for every* $n \in \mathbb{N}$.

Proof. We will prove by induction that for every $n \in \mathbb{N}$,

$$f^{(n)}(x) = \begin{cases} \exp\left(-\frac{1}{x^2}\right) \left[P_n\left(\frac{1}{x}\right) \sin\left(\frac{1}{x}\right) + Q_n\left(\frac{1}{x}\right) \cos\left(\frac{1}{x}\right)\right], & x \neq 0; \\ 0, & x = 0, \end{cases} \tag{2.3}$$

where P_n and Q_n are polynomials. This further implies that $f^{(n)}$ is continuous at 0 and $f \in C^n$. Indeed, for $0 < |x| < 1$

$$0 \leq |f^{(n)}(x)| \leq \exp\left(-\frac{1}{x}\right) |P_n\left(\frac{1}{x}\right)| + \exp\left(-\frac{1}{x}\right) |Q_n\left(\frac{1}{x}\right)| \to 0, \ x \to 0,$$

where the limit is implied by Lemma 2.9.

Now we prove (2.3). It's easy to see that for nonzero x,

$$f^{(1)}(x) = \exp\left(-\frac{1}{x^2}\right) \left[\frac{2}{x^3} \sin(\frac{1}{x}) - \frac{1}{x^2} \cos(\frac{1}{x})\right]$$

and $f^{(1)}(0) = \lim_{x \to 0} \frac{f(x)}{x} = 0$. Now assume that $f^{(n-1)}(x)$ obeys

$$f^{(n-1)}(x) = \begin{cases} \exp\left(-\frac{1}{x^2}\right) \left[P_{n-1}\left(\frac{1}{x}\right) \sin\left(\frac{1}{x}\right) + Q_{n-1}\left(\frac{1}{x}\right) \cos\left(\frac{1}{x}\right)\right], & x \neq 0; \\ 0, & x = 0. \end{cases}$$

For nonzero x,

$$\begin{aligned} f^{(n)}(x) &= \exp\left(-\frac{1}{x^2}\right) \Big[\frac{P_{n-1}^{(1)}(1/x) + Q_{n-1}(1/x)}{x^2} \sin(1/x) \\ &\quad - \frac{P_{n-1}(1/x) + Q_{n-1}^{(1)}(1/x)}{x^2} \cos(1/x)\Big] \\ &\quad + \frac{2}{x^3} \exp\left(-\frac{1}{x^2}\right) \left[P_{n-1}(1/x) \sin(1/x) + Q_{n-1}(1/x) \cos(1/x)\right] \\ &= \exp\left(-\frac{1}{x^2}\right) [P_n(1/x) \sin(1/x) + Q_n(1/x) \cos(1/x)], \end{aligned}$$

where P_n and Q_n are polynomials given by

$$P_n(y) = y^2 Q_{n-1}(y) + y^2 P_{n-1}^{(1)}(y) + 2y^3 P_{n-1}(y),$$
$$Q_n(y) = -y^2 P_{n-1}(y) - y^2 Q_{n-1}^{(1)}(y) + 2y^3 Q_{n-1}(y).$$

By Lemma 2.9 for $0 < |x| < 1$,

$$0 \leq |\frac{f^{n-1}(x)}{x}| \leq \exp\left(-\frac{1}{x}\right)\left|\frac{1}{x}P_{n-1}(\frac{1}{x})\right| + \exp\left(-\frac{1}{x}\right)\left|\frac{1}{x}Q_{n-1}\left(\frac{1}{x}\right)\right| \to 0,$$

as $x \to 0$, which means $f^{(n)}(0) = 0$. □

Another function of interest is the Lambert $\mathcal{W}$ function, which is the two-valued inverse of $x \mapsto xe^x$. We refer to [5] for a comprehensive collection of properties of this function. Denote by $\mathcal{W}_0$ and $\mathcal{W}_{-1}$ the principal and other branches of the Lambert $\mathcal{W}$ function.

Fact 2.11. *The following statements hold:*

(i) For $y \geq -1/e$, one has

$$\mathcal{W}(y) = \begin{cases} \{\mathcal{W}_0(y), \mathcal{W}_{-1}(y)\}, & -\frac{1}{e} \leq y < 0; \\ \mathcal{W}_0(y), & y \geq 0. \end{cases}$$

(ii) For every $y \geq 0$, one has $y = xe^x \Leftrightarrow x = \mathcal{W}(y) = \mathcal{W}_0(y)$.

Proof. (i) Note that $(xe^x)' = (x+1)e^x$. Hence the function $x \mapsto xe^x$ is decreasing on $(-\infty, -1]$ and increasing on $[-1, \infty)$, with global minimizer $-1/e$. Therefore the equation $y = xe^x$ is solvable for x only when $y \geq -1/e$. On the other hand, since $y = xe^x \geq 0$ for $x \geq 0$ and the function $x \mapsto xe^x$ is increasing on $[0, \infty)$, one concludes that the inverse $\mathcal{W}(y)$ is single-valued for $y \geq 0$.

(ii) We learn from (i) that $\mathcal{W}(y)$ is single-valued for $y \geq 0$. Therefore by the definition $y = xe^x \Leftrightarrow x = \mathcal{W}(y) = \mathcal{W}_0(y)$. □

2.3 A bit of set-valued analysis

In this section we collect useful definitions and facts about set-valued mappings $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ that will be used in the sequel. For a systematic exposition of this subject, see Aubin and Frankowska's *Set-valued Analysis* [2].

Definition 2.12. Let $(S_n)_{n\in\mathbb{N}}$ be a sequence of sets in $\mathbb{R}^n$ and let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a set-valued mapping.

(i) The *Painlevé-Kuratowski outer limit* of $(S_n)_{n\in\mathbb{N}}$ is defined by

$$\operatorname*{Limsup}_{n\to\infty} S_n = \{x \in \mathbb{R}^n : \exists\, (S_{n_k})_{k\in\mathbb{N}} \subseteq (S_n)_{n\in\mathbb{N}}\,, \exists x_k \in S_{n_k}, x_k \to x \text{ as } k \to \infty\}.$$

Equivalently the outer limit of $(S_n)_{n\in\mathbb{N}}$ consists of all limit points of sequences $(x_n)_{n\in\mathbb{N}}$ with $x_n \in S_n$.

(ii) The *Painlevé-Kuratowski outer limit* of F at $\bar{x}$ is

$$\begin{aligned}\operatorname*{Limsup}_{x\to\bar{x}} F(x) &= \bigcup_{x_n\to\bar{x}} \operatorname*{Limsup}_{n\to\infty} F(x_n)\\ &= \{y \in \mathbb{R}^n : \exists x_n \to \bar{x}, \exists y_n \in F(x_n), y_n \to y\}.\end{aligned}$$

Moreover, we say that F is *outer semicontinuous* at $\bar{x}$, if

$$\operatorname{Limsup}_{x\to\bar{x}} F(x) \subseteq F(\bar{x})$$

or equivalently $\operatorname{Limsup}_{x\to\bar{x}} F(x) = F(\bar{x})$.

Example 2.13. Let $F(x) = \{r \in \mathbb{R} : r \geq f(x)\}$, where $f(x)$ is defined by

$$f(x) = \begin{cases} 0, & x = 0;\\ 1, & x \neq 0.\end{cases}$$

Then F is outer semicontinuous at 0.

Proof. For $x \neq 0$, $F(x) = [1, \infty)$ while $F(x) = [0, \infty)$ for $x = 0$. Then we have

$$\operatorname{Limsup}_{x\to 0} F(x) = [1, \infty) \cup [0, \infty) = [0, \infty),$$

which means F is outer semicontinuous at 0. □

Chapter 3

Elements in nonsmooth analysis

In this section, we discuss key objects and facts in nonsmooth analysis that will be used in the sequel. Recall that we say $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *(Fréchet) differentiable* at $\bar{x} \in \operatorname{int} \operatorname{dom} f$, if there exists a unique vector $\nabla f(\bar{x}) \in \mathbb{R}^n$ such that

$$\lim_{\substack{x \to \bar{x} \\ x \neq \bar{x}}} \frac{f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x} \rangle}{\|x - \bar{x}\|} = 0, \tag{3.1}$$

in which case $\nabla f(\bar{x})$ is called a *gradient*. Moreover, we say f is *(Fréchet) differentiable* on $\operatorname{int} \operatorname{dom} f$, which we regime to be nonempty, if (3.1) holds for every $\bar{x} \in \operatorname{int} \operatorname{dom} f$.

3.1 Convex functions and their variants

Definition 3.1. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}} = (-\infty, \infty]$ and suppose that $\operatorname{dom} f$ is a convex set. We say that

(i) f is *convex*, if for $\forall x, y \in \operatorname{dom} f$ and $\forall \lambda \in (0, 1)$,

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y). \tag{3.2}$$

(ii) f is *strictly convex*, if for $\forall x, y \in \operatorname{dom} f$ with $x \neq y$ and $\forall \lambda \in (0, 1)$,

$$f(\lambda x + (1 - \lambda) y) < \lambda f(x) + (1 - \lambda) f(y). \tag{3.3}$$

Definition 3.2. Let $f : \mathbb{R}^n \to [-\infty, \infty)$ and suppose that $\operatorname{dom} f$ is a convex set. We say that f is *concave*, if $-f$ is convex; f is *strictly concave*, if $-f$ is strictly convex.

Fact 3.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable, and let $C \subseteq \mathbb{R}^n$ be an open convex set. Then f is convex on C if and only if one of the following holds:*

(i) For every $x, y \in C$, $\langle x - y, \nabla f(x) - \nabla f(y) \rangle \geq 0$;

(ii) *For every* $x, y \in C$, $\langle x - y, \nabla f(y)\rangle + f(y) \leq f(x)$;

(iii) $\nabla^2 f(x)$ *is positive semi-definite for every* $x \in C$, *if* f *is* C^2.

Proof. See [4, Proposition 17.7]. □

Now we introduce some stronger notions of convex functions.

Definition 3.4. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and let $\operatorname{dom} f$ be a convex set.

(i) f is *uniformly convex* with modulus ϕ, if for every $x, y \in E$ and $\lambda \in (0, 1)$,

$$f(\lambda x + (1-\lambda)y) + \lambda(1-\lambda)\phi(\|x-y\|) \leq \lambda f(x) + (1-\lambda)f(y), \quad (3.4)$$

where $\phi : \mathbb{R}_+ \to [0, \infty]$ is an strictly increasing function that vanishes only at 0;

(ii) f is *β-strongly convex*, if it is uniformly convex with modulus $\phi = (\beta/2)|\cdot|^2$ for some $\beta > 0$.

Fact 3.5. *Let* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be proper and let* $\beta > 0$. *Then* f *is* β*-strongly convex if and only if* $f - (\beta/2)\|\cdot\|^2$ *is convex.*

Proof. See [4, Proposition 10.8]. □

The following concept captures the essence of how "convex" a function is.

Definition 3.6. [4] Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be proper and convex. The *exact modulus of convexity* of f is $\psi(t) : \mathbb{R}_+ \to [0, \infty]$

$$t \mapsto \inf_{\substack{x \in \operatorname{dom} f, y \in \operatorname{dom} f \\ \|x-y\| = t, \lambda \in (0,1)}} \frac{\lambda f(x) + (1-\lambda)f(y) - f(\lambda x + (1-\lambda)y)}{\lambda(1-\lambda)}.$$

It is intrinsically difficult to compute the exact modulus of convexity. However, one can estimate it by the following fact.

Fact 3.7. *Let* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be proper and convex, with exact modulus of convexity* ψ, *and set* $\Gamma : \mathbb{R} \to [-\infty, \infty]$,

$$t \mapsto \inf\left\{\frac{f(x) + f(y)}{2} - f\left(\frac{x+y}{2}\right) : x, y \in \operatorname{dom} f, t = \|x - y\|\right\}.$$

Then $2\Gamma \leq \psi \leq 4\Gamma$.

Proof. See [4, Propostion 10.14]. □

Example 3.8. Let $f_1(x) = x^2$ and let $f_2(x) = x^4$. Let $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ be an strictly increasing function that vanishes only at 0 and obey $\phi(t) \leq \frac{t^4}{8}$ for all $t \geq 0$. Then the following assertions hold:

(i) f_1 is 2-strongly convex;

(ii) f_2 is uniformly convex with modulus $\phi(t)$, but f_2 is not strongly convex;

Proof. (i) This is trivially true by Fact 3.5.

(ii) We start with estimating the exact modulus of convexity. For a fixed $t \geq 0$,

$$\begin{aligned}
\Gamma(t) &= \inf\left\{\frac{f(x)+f(y)}{2} - f\left(\frac{x}{2}+\frac{y}{2}\right) : t = |x-y|, x, y \in \mathbb{R}\right\} \\
&= \inf\left\{\frac{x^4+y^4}{2} - \left(\frac{x+y}{2}\right)^4 : y = x-t \text{ or } y = x+t, x \in \mathbb{R}\right\} \\
&= \inf\left\{\frac{x^4+y^4}{2} - \left(\frac{x+y}{2}\right)^4 : y = x-t, x \in \mathbb{R}\right\} \\
&= \inf\left\{\frac{3t^2}{2}x^2 - \frac{3t^3}{2}x + \frac{7}{16}t^4 : x \in \mathbb{R}\right\} = \frac{t^4}{16},
\end{aligned}$$

where the third equality holds because of symmetry. Hence invoking Fact 3.7 yields that the exact modulus of convexity $\psi(t)$ satisfies

$$\frac{t^4}{8} \leq \psi(t) \leq \frac{t^4}{4}.$$

Above estimation further implies that for every $x, y \in \mathbb{R}$ and $\lambda \in (0,1)$

$$\phi(|x-y|) \leq \psi(|x-y|) \leq \frac{\lambda f(x) + (1-\lambda)f(y) - f(\lambda x + (1-\lambda)y)}{\lambda(1-\lambda)}.$$

Therefore f_2 is uniformly convex with modulus $\phi(t)$. By Fact 3.5, we have f_2 is β-strongly convex for $\beta > 0$ if and only if

$$\begin{aligned}
f_2(x) - \frac{\beta}{2}x^2 \text{ is convex} &\Leftrightarrow (f_2 - \frac{\beta}{2}|\cdot|^2)^{(2)}(x) \geq 0, \forall x \in \mathbb{R} \\
&\Leftrightarrow 12x^2 - \beta \geq 0, \forall x \in \mathbb{R} \\
&\Rightarrow \beta \leq 0,
\end{aligned}$$

which is absurd. □

We will use frequently the following notion:

Definition 3.9. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, let $x \in \operatorname{dom} f$, and let $d \in \mathbb{R}^n$. The *directional derivative* of f at x in the direction d is

$$f'(x; d) = \lim_{\alpha \to 0^+} \frac{f(x + \alpha d) - f(x)}{\alpha}.$$

Remark 3.10. On the real line, the right and left derivatives are given by

$$f'_+(x) = f'(x; 1) = \lim_{\alpha \to 0^+} \frac{f(x + \alpha) - f(x)}{\alpha}$$

and

$$f'_-(x) = -f'(x; -1) = \lim_{\alpha \to 0^-} \frac{f(x + \alpha) - f(x)}{\alpha}.$$

Fact 3.11. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex, let $x \in \operatorname{dom} f$ and let $y \in \mathbb{R}^n$. Then the following hold:*

(i) $f'(x; y)$ exists and

$$f'(x; y) = \inf_{\alpha > 0} \frac{f(x + \alpha y) - f(x)}{\alpha}.$$

(ii) $f'(x; y - x) + f(x) \leq f(y)$.

(iii) $f'(x; \cdot)$ is sublinear.

Proof. See [4, Proposition 17.2]. □

Fact 3.12. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz smooth with modulus $L > 0$. Then for every $x, y \in \mathbb{R}^n$,*

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

Proof. See [4, Theorem 18.15(iii)]. □

3.2 Convex functions on the real line

We collect in this section some fine properties of convex functions on the real line.

Fact 3.13. *Let $f : \mathbb{R} \to \overline{\mathbb{R}}$ be a closed proper convex function. Then f'_+ and f'_- are increasing functions on $\mathbb{R}$, finite on the interior of* $\operatorname{dom} f$*, such that*

$$f'_-(x) \leq f'_+(x).$$

Moreover, for every $x, y \in \operatorname{int} \operatorname{dom} f$, one has

$$f(y) - f(x) = \int_x^y f'_-(t)dt = \int_x^y f'_+(t)dt.$$

Proof. See [20, Theorem 24.1, Corollary 24.2.1]. □

The next technical lemma concerning the finiteness of an improper integral will play a key role in our analysis.

Lemma 3.14. *Let $m \in \mathbb{N}$ obey $m \geq 2$. For each $i \in \{1, \ldots, m\}$, let $h_i : (0, \infty) \to \mathbb{R}_+$ be such that $\lim_{s \to 0^+} h(s) = \infty$. Suppose that for each i the function $\varphi_i : [0, \infty) \to \mathbb{R}_+$ given by $\varphi_i(t) = \int_0^t h_i(s)ds$ for $t \in (0, \infty)$ and $\varphi_i(0) = 0$, is finite and right-continuous at 0. Then the function $\varphi : [0, \infty) \to \mathbb{R}_+$ defined by*

$$\varphi(t) = \int_0^t \max_{1 \leq i \leq m} h_i(s)ds, \forall t \in (0, \infty)$$

and $\varphi(0) = 0$, is finite and right-continuous at 0.

Proof. Note that φ is an improper integral. Hence we have for $t > 0$

$$\varphi(t) = \lim_{u \to 0^+} \int_u^t \max_{1 \leq i \leq m} h_i(s)ds.$$

We prove the desired statement by induction. Recall from Fact 2.2 that $\max\{\alpha, \beta\} \leq \alpha + \beta$ for $\alpha, \beta \geq 0$. Let $m = 2$. Then one has

$$\max\{h_1(s), h_2(s)\} \leq h_1(s) + h_2(s).$$

Hence we have for $t > 0$

$$\begin{aligned}
\varphi(t) &\leq \lim_{u \to 0^+} \int_u^t h_1(s) + h_2(s)ds \\
&= \varphi_1(t) + \varphi_2(t) - \lim_{u \to 0^+} [\varphi_1(u) + \varphi_2(u)] \\
&= \varphi_1(t) + \varphi_2(t) < \infty,
\end{aligned}$$

where the last equality is implied by the right-continuity of φ_1 and φ_2 at 0. Taking $t \to 0^+$, one gets $\lim_{t\to 0^+} \varphi(t) = 0$.

Define $H(s) = \max_{1\le i\le m-1} h_i(s)$ and suppose that $\psi(t) = \int_0^t H(s)ds$ is finite and right-continuous at 0 with $\psi(0) = 0$. Then one has for $t > 0$

$$\begin{aligned}\varphi(t) &= \lim_{u\to 0^+} \int_u^t \max\{H(s), h_m(s)\}ds \le \lim_{u\to 0^+} \int_u^t H(s) + h_m(s)ds \\ &= \psi(t) + \varphi_m(t) - \lim_{u\to 0^+} [\psi(u) + \varphi_m(u)] \\ &= \psi(t) + \varphi_m(t) < \infty.\end{aligned}$$

Consequently, we have

$$0 \le \liminf_{t\to 0^+} \varphi(t) \le \limsup_{t\to 0^+} \big(\psi(t) + \varphi_m(t)\big) = 0,$$

where the last equality is implied by the right-continuity of ψ and φ_m at 0, which means that $\lim_{t\to 0^+} \varphi(t) = 0$. □

Fact 3.15. *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 0. 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. Let $(t_n)_{n\in\mathbb{N}}$ be such that $t_n \to 0^+$ and define $h_n(s) = h(s) \cdot \mathcal{X}_{[0,t_n]}(s)$ for each n, where $\mathcal{X}_{[0,t_n]}(s) = 1$ if $s \in [0, t_n]$ and $\mathcal{X}_{[0,t_n]}(s) = 0$ otherwise. Then one has $h_n(s) \to \mathcal{X}_{\{0\}}(s)$ pointwise almost everywhere on $[0, \infty)$ with $|h_n(s)| \le h(s)$ for all n. Hence by the Lebesgue dominated convergence theorem, one concludes that $\varphi(t_n) = \int_0^{t_n} h(s)ds = \int h_n(s)ds \to \int \mathcal{X}_{\{0\}}(s)ds = 0$, which means that $\varphi(t)$ is right-continuous at 0.

Let $\lambda \in (0, 1)$ and set $t_\lambda = (1 - \lambda)t_0 + \lambda t_1$. Then $\lambda = (t_\lambda - t_0)/(t_1 - t_0)$ and $1 - \lambda = (t_1 - t_\lambda)/(t_1 - t_0)$. Note that $h(s)$ is decreasing. Hence one has

$$\begin{aligned}\varphi(t_\lambda) - \varphi(t_0) &= \int_{t_0}^{t_\lambda} h(s)ds \ge (t_\lambda - t_0)h(t_\lambda), \\ \varphi(t_1) - \varphi(t_\lambda) &= \int_{t_\lambda}^{t_1} h(s)ds \le (t_1 - t_\lambda)h(t_\lambda).\end{aligned}$$

Consequently, we have

$$\begin{aligned}
&(1-\lambda)\left[\varphi(t_\lambda)-\varphi(t_0)\right]+\lambda\left[\varphi(t_\lambda)-\varphi(t_1)\right]\\
&\geq\left[(1-\lambda)(t_\lambda-t_0)-\lambda(t_1-t_\lambda)\right]h(t_\lambda)\\
&=\left[\frac{(t_1-t_\lambda)(t_\lambda-t_0)}{t_1-t_0}-\frac{(t_\lambda-t_0)(t_1-t_\lambda)}{t_1-t_0}\right]h(t_\lambda)=0,
\end{aligned}$$

which means that $\varphi(t_\lambda)\geq(1-\lambda)\varphi(t_0)+\lambda\varphi(t_1)$. Fix $\lambda\in(0,1)$ and $t_1\in(0,\eta)$. Taking $t_0\to 0^+$ yields that

$$\varphi(\lambda t_1)=\lim_{t_0\to 0^+}\varphi((1-\lambda)t_0+\lambda t_1)\geq\lambda\varphi(t_1)+(1-\lambda)\lim_{t_0\to 0^+}\varphi(t_0)=\lambda\varphi(t_1),$$

where the first equality holds because φ is increasing and $\varphi((1-\lambda)t_0+\lambda t_1)\geq\varphi(\lambda t_1)$ for every $t_0>0$. Therefore, one concludes that φ is concave on $[0,\eta)$.

Fix $t\in(0,\eta)$. Then for every $t'<t$, one has

$$\varphi(t)-\varphi(t')=\int_{t'}^{t}h(s)ds\geq(t-t')h(t)\Rightarrow\frac{\varphi(t')-\varphi(t)}{t'-t}\geq h(t).$$

Taking $t'\to t^-$, one gets $\varphi'_-(t)\geq h(t)$.

If in addition h is continuous, then by applying the fundamental theorem of calculus, one concludes that φ is C^1 on $(0,\eta)$. □

3.3 Subgradients beyond convex analysis

We discuss in this section well-known facts about several *subgradients* in nonconvex setting, and give concrete examples to satisfy these properties. We will use frequently the notation

$$x_n\xrightarrow{f}\bar{x}\Leftrightarrow x_n\to\bar{x}\text{ with }f(x_n)\to f(\bar{x}).$$

Definition 3.16. 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 $x\in\operatorname{dom}f$,

$$f(x)\geq f(\bar{x})+\langle v,x-\bar{x}\rangle+o(\|x-\bar{x}\|).\tag{3.5}$$

Moreover, (3.5) is equivalent to

$$\liminf_{x\to\bar{x}}\frac{f(x)-f(\bar{x})-\langle v,x-\bar{x}\rangle}{\|x-\bar{x}\|}\geq 0,\tag{3.6}$$

by asserting $o(\|x-\bar{x}\|)=\min\{0,f(x)-f(\bar{x})-\langle v,x-\bar{x}\rangle)\}$.

(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

$$v \in \{u \in \mathbb{R}^n : \exists x_n \xrightarrow{f} \bar{x}, \exists u_n \in \hat{\partial} f(x_n), u_n \to u\}, \tag{3.7}$$

i.e., $\partial f(\bar{x}) = \operatorname{Limsup}_{x \xrightarrow{f} \bar{x}} \hat{\partial} f(x)$. The notation $\operatorname{dom} \partial f$ denotes the set of all limiting-subdifferentiable points, i.e., $\operatorname{dom} \partial f = \{x \in \mathbb{R}^n : \partial f(x) \neq \emptyset\}$.

Fact 3.17. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and $\bar{x} \in \operatorname{dom} f$. Then the following statements hold:*

(i) The subgradients sets $\hat{\partial} f(\bar{x})$ and $\partial f(\bar{x})$ are always closed, with $\hat{\partial} f(\bar{x})$ is convex and $\hat{\partial} f(\bar{x}) \subseteq \partial f(\bar{x})$;

(ii) If f is convex and $\bar{x} \in \operatorname{int} \operatorname{dom} f$, then

$$f'(x; d) = \max\{\langle g, d\rangle : g \in \partial f(x)\},$$

$$\hat{\partial} f(\bar{x}) = \partial f(\bar{x}) = \{v \in \mathbb{R}^n : f(x) \geq f(\bar{x}) + \langle v, x - \bar{x}\rangle, \forall x \in \operatorname{dom} f\}.$$

In particular, for f defined on the line, $\partial f(x) = [f'_-(x), f'_+(x)] \cap \mathbb{R}$.

Proof. See [21, Theorem 8.6, Proposition 8.12] and [4, Proposition 17.16, Theorem 17.18]. □

Fact 3.18. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and $\bar{x} \in \operatorname{dom} f$. Then the following statements are true:*

(i) If f is differentiable at $\bar{x}$, then $\hat{\partial} f(\bar{x}) = \{\nabla f(\bar{x})\}$, and therefore $\nabla f(\bar{x}) \in \partial f(\bar{x})$;

(ii) If f is smooth on a neighborhood of $\bar{x}$, then $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$;

(iii) Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and $x \in \operatorname{dom} g$. If f is smooth on a neighborhood of $\bar{x}$, then $\hat{\partial}(f+g)(\bar{x}) = \hat{\partial} g(\bar{x}) + \nabla f(\bar{x})$ and $\partial(f+g)(\bar{x}) = \partial g(\bar{x}) + \nabla f(\bar{x})$;

(iv) Let $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ be proper, and let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be smooth. Let $\Psi(x, y) = f(x) + g(y) + F(x, y)$. Then for $(\bar{x}, \bar{y}) \in \operatorname{dom} f \times \operatorname{dom} g$, one has

$$\partial \Psi(\bar{x}, \bar{y}) = [\partial f(\bar{x}) + \nabla_x F(\bar{x}, \bar{y})] \times [\partial g(\bar{y}) + \nabla_y F(\bar{x}, \bar{y})], \tag{3.8}$$

where $\nabla_x F(x, y)$ and $\nabla_y F(x, y)$ denote the partial gradient of $F(x, y)$ on $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, respectively.

Proof. (i) According to (3.1), one has $f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x}\rangle = o(\|x - \bar{x}\|)$, which means that $\nabla f(\bar{x}) \in \hat{\partial} f(\bar{x})$. Let $v \in \hat{\partial} f(\bar{x})$. Then one has for every $x \in \operatorname{dom} f$,

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

Let $d = v - \nabla f(\bar{x})$ and let $x = \bar{x} + rd$ for $r \in \mathbb{R}$. Above inequality yields

$$r\langle \nabla f(\bar{x}) - v, v - \nabla f(\bar{x})\rangle \geq o(\|rd\|) \Rightarrow \|v - \nabla f(\bar{x})\| \leq -\frac{o(\|rd\|)}{\|rd\|}, \forall r \neq 0.$$

Taking $r \to 0$ we get $\|v - \nabla f(\bar{x})\| \leq 0$. Hence $v = \nabla f(\bar{x})$ and $\hat{\partial} f(\bar{x}) = \{\nabla f(\bar{x})\}$. Invoking Fact 3.17 gives $\nabla f(\bar{x}) \in \partial f(\bar{x})$.

(ii) Let $r > 0$ and suppose that f is smooth at x for every $x \in \mathbb{B}_r$. By (i), one has $\hat{\partial} f(x) = \nabla f(x)$. Smoothness implies that f and ∇f are continuous around $\bar{x}$. Hence $\partial f(\bar{x}) = \operatorname{Limsup}_{x \xrightarrow{f} \bar{x}} \hat{\partial} f(x) = \operatorname{Limsup}_{x \to \bar{x}} \nabla f(x) = \nabla f(\bar{x})$.

(iii) "$\hat{\partial} g(\bar{x}) + \nabla f(\bar{x}) \subseteq \hat{\partial}(f + g)(\bar{x})$": Let $v \in \hat{\partial} g(\bar{x})$. Then for $x \in \operatorname{dom} g \cap \operatorname{dom} f$,

$$g(x) + f(x) \geq g(\bar{x}) + f(\bar{x}) + \langle v + \nabla f(\bar{x}), x - \bar{x}\rangle + o(\|x - \bar{x}\|),$$

meaning that $v + \nabla f(\bar{x}) \in \hat{\partial}(f + g)(\bar{x})$.

"$\hat{\partial}(f+g)(\bar{x}) \subseteq \hat{\partial} g(\bar{x}) + \nabla f(\bar{x})$": Observe that $g(x) = (f+g)(x) + (-f)(x)$. Then applying above inclusion to $x \mapsto (f + g)(x) + (-f)(x)$ yields

$$\hat{\partial}(f + g)(\bar{x}) - \nabla f(\bar{x}) \subseteq \hat{\partial} f(\bar{x}) \Leftrightarrow \hat{\partial}(f + g)(\bar{x}) \subseteq \hat{\partial} g(\bar{x}) + \nabla f(\bar{x}).$$

Next we show $\partial(f + g)(\bar{x}) = \partial g(\bar{x}) + \nabla f(\bar{x})$. Let $v \in \partial g(\bar{x})$. Then there exist $x_n \xrightarrow{g} \bar{x}$ and $v_n \in \hat{\partial} g(x_n)$ such that $v_n \to v$. It follows that $v_n + \nabla f(x_n) \in \hat{\partial}(g+f)(x_n)$, and continuity of f implies the limit $v + \nabla f(\bar{x}) \in \partial(g + f)(\bar{x})$, which means $\partial g(\bar{x}) + \nabla f(\bar{x}) \subseteq \partial(g + f)(\bar{x})$.

Now take $u \in \partial(g + f)(\bar{x})$, then there exist $y_n \xrightarrow{f+g} \bar{x}$ and $u_n \in \hat{\partial}(f + g)(y_n)$ such that $u_n \to u$. Clearly $u_n - \nabla f(y_n) \in \hat{\partial} g(y_n)$ and therefore $u - \nabla f(\bar{x}) \in \partial g(\bar{x})$, which completes the proof.

(iv) Applying [21, Propostion 10.5] to the function $(x, y) \mapsto f(x) + g(y)$ and (iii) proves the statement immediately. □

Fact 3.19. *Suppose that $f(x) = g(F(x))$ for a proper lsc function $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ and a smooth map $F : \mathbb{R}^n \to \mathbb{R}^m$. Let $\bar{x}$ be a point where f is finite and the Jacobian $\nabla F(\bar{x})$ is surjective. Then*

$$\hat{\partial} f(\bar{x}) = \nabla F(\bar{x})^* \hat{\partial} g(F(\bar{x})) \text{ and } \partial f(\bar{x}) = \nabla F(\bar{x})^* \partial g(F(\bar{x})). \tag{3.9}$$

Proof. See [21, Exercise 10.7]. □

Fact 3.20. *Suppose that* $f(x) = \min_{1\le i\le m} f_i(x)$*, where* $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is lsc and proper function. Let* $x \in \bigcap \operatorname{dom} \partial f_i$ *and* $I(x) = \{i : f(x) = f_i(x)\}$*. Then*

$$\partial f(x) \subseteq \bigcup_{i\in I(x)} \partial f_i(x).$$

Proof. See [17, Proposition 4.9]. □

Fact 3.21. *Suppose that* $f(x) = \sum_{i=1}^n f_i(x_i)$*, where* $f_i : \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$ *is proper and lsc for each* $i \in \{1, \dots, n\}$*. Then we have*

$$\partial f(x) = \prod_{i=1}^n \partial f_i(x).$$

Proof. See [21, Proposition 10.5]. □

Now we compute the Fréchet and limiting subgradients of some functions on the line.

Example 3.22. Let $f_1(x) = |x|$, $f_2(x) = -|x|$ and let f_3 and f_4 be defined by

$$f_3(x) = \begin{cases} 0, & x = 0; \\ 1, & x \neq 0. \end{cases} \quad f_4(x) = \begin{cases} x^2 \sin(1/x), & x \neq 0; \\ 0, & x = 0. \end{cases}$$

Let $f_5(x) = -|x|^{1/2}$. Then the following statements hold:

(i) $\hat{\partial} f_1(0) = \partial f_1(0) = [-1, 1]$, which has Fact 3.17.

(ii) $\hat{\partial} f_2(0) = \emptyset, \partial f_2(0) = \{-1, 1\}$, meaning that for nonconvex functions $\hat{\partial} f(\bar{x})$ and $\partial f(\bar{x})$ are not necessarily identical.

(iii) $\hat{\partial} f_3(0) = \partial f_3(0) = \mathbb{R}$, which shows $\hat{\partial} f(\bar{x})$ and $\partial f(\bar{x})$ can be identical for nonconvex functions.

(iv) $\hat{\partial} f_4(0) = \{0\}, \partial f_4(0) = [-1, 1]$, which verifies Fact 3.18(i) and shows that Fact 3.18(ii) fails without smoothness.

(v) $\hat{\partial} f_5(0) = \partial f_5(0) = \emptyset$. This means that $\hat{\partial} f(\bar{x})$ and $\partial f(\bar{x})$ may be empty.

Proof. (i) For $\hat{\partial} f_1(0)$, one has

$$\begin{aligned} v \in \hat{\partial} f_1(0) &\Leftrightarrow \liminf_{x\to 0} \frac{|x| - vx}{|x|} = 1 - |v| \geq 0 \\ &\Leftrightarrow |v| \leq 1. \end{aligned}$$

For $x \neq 0$, Fact 3.18 implies $\hat{\partial} f_1(x) = -1$ if $x < 0$, and $\hat{\partial} f_1(x) = 1$ if $x > 0$. Hence $\partial f_1(0) = \operatorname{Limsup}_{x \xrightarrow{f} 0} \hat{\partial} f(x) = \operatorname{Limsup}_{x\to 0} \hat{\partial} f(x) = [-1, 1]$.

(ii) Let $v \in \mathbb{R}$ and suppose that $v \in \hat{\partial} f_2(0)$. Then we have

$$v \in \hat{\partial} f_2(0) \Leftrightarrow \liminf_{x\to 0} \frac{-|x| - vx}{|x|} = -1 - |v| \geq 0,$$

which is absurd. Hence $\hat{\partial} f_2(0) = \emptyset$. Applying Fact 3.18 (i), one gets $\hat{\partial} f_2(x) = 1$ for $x < 0$ and $\hat{\partial} f_2(x) = -1$ for $x > 0$. f_2 is continuous, hence we have $\partial f_2(0) = \operatorname{Limsup}_{x\to 0} \hat{\partial} f_2(x) = \{-1, 1\}$.

(iii) First we analyze $\hat{\partial} f_3(0)$. Since for every $v \in \mathbb{R}$ we have

$$\liminf_{x\to 0} \frac{1 - vx}{|x|} = \infty > 0,$$

$\hat{\partial} f_3(0) = \mathbb{R}$. Applying Fact 3.17 gives $\hat{\partial} f_3(0) \subseteq \partial f_3(0) \subseteq \mathbb{R}$, $\partial f_3(0) = \mathbb{R}$.

(iv) For $v \in \mathbb{R}$, one has $v \in \hat{\partial} f_4(0)$ if and only if

$$\liminf_{x\to 0} \frac{x^2 \sin(1/x) - vx}{|x|} = -|v| \geq 0 \Leftrightarrow v = 0.$$

Hence $\hat{\partial} f_4(0) = \{0\}$. Invoking Fact 3.18 (i) we know that for $x \neq 0$, $\hat{\partial} f_4(x) = (x^2 \sin(1/x))' = 2x \sin(1/x) - \cos(1/x)$. Then $\partial f_4(0) = \limsup_{x \xrightarrow{f} 0} \hat{\partial} f_4(x) = \operatorname{Limsup}_{x\to 0}\{2x \sin(1/x) - \cos(1/x)\} = [-1, 1]$.

(v) Let $v \in \mathbb{R}$. Then $v \in \hat{\partial} f_5(0)$ if and only if

$$\liminf_{x\to 0} \frac{-|x|^{1/2} - vx}{|x|} = \liminf_{x\to 0} -|x|^{-1/2} - v \cdot \operatorname{sgn}(x) = -\infty,$$

meaning that $\hat{\partial} f_5(0) = \emptyset$. For nonzero x, $\hat{\partial} f_5(x) = f_5'(x) = -\operatorname{sgn}(x)/(2|x|^{1/2})$. Therefore $\partial f_5(0) = \emptyset$, as $|\hat{\partial} f_5(x)| = |f_5'(x)| = 1/(2|x|^{1/2}) \to \infty$ as $x \to 0$. $\square$

Recall for a norm $\|\cdot\|$ on $\mathbb{R}^n$, the corresponding dual norm is given by

$$\|\cdot\|_* = \sup\{\langle \cdot, u\rangle : \|u\| = 1\}.$$

Example 3.23. Let $\|\cdot\|$ be a norm and let $\|\cdot\|_*$ be its dual norm. Let $f(x) = \|x\|$. Then $\partial f(0) = \mathbb{B}_{\|\cdot\|_*}(0;1) = \{g \in \mathbb{R}^n : \|g\|_* \leq 1\}$.

Proof. Let $g \in \partial f(0)$. Then by Fact 3.17 one has $f(0) + \langle g, x\rangle \leq f(x) \Leftrightarrow \langle g, x\rangle \leq \|x\|, \forall x \in \mathbb{R}^n \Leftrightarrow \langle g, \frac{x}{\|x\|}\rangle \leq 1, \forall x \neq 0 \Leftrightarrow \|g\|_* \leq 1$. □

Fact 3.24. *For a proper function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, the set-valued mapping $x \mapsto \partial f(x)$ is outer semicontinuous at $\bar{x} \in \operatorname{dom} f$ with respect to f-attentive convergence $x \xrightarrow{f} \bar{x}$.*

Proof. See [21, Propostion 8.7] □

Fact 3.25. *If a proper function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ has a local minimum at $\bar{x}$, then*

$$0 \in \partial f(\bar{x}).$$

If in addition f is convex, then $0 \in \partial f(\bar{x}) \Leftrightarrow \bar{x} \in \operatorname{argmin} f$.

Proof. Suppose that $f(x) \geq f(\bar{x})$ for $x \in \mathbb{B}(\bar{x};\varepsilon)$. Then one has

$$\liminf_{x\to\bar{x}} \frac{f(x) - f(\bar{x}) - \langle 0, x - \bar{x}\rangle}{\|x - \bar{x}\|} = \liminf_{x\to\bar{x}} \frac{f(x) - f(\bar{x})}{\|x - \bar{x}\|} \geq 0,$$

which means $0 \in \hat{\partial} f(\bar{x}) \subseteq \partial f(\bar{x})$. If in addition f is convex, then by Fact 3.17(ii) one concludes that

$$f(x) \geq f(\bar{x}), \forall x \in \operatorname{dom} f \Rightarrow 0 \in \partial f(\bar{x}),$$

which completes the proof. □

3.4 Proximal mapping

For consistency, we follow the definition of *proximal mapping* in [9].

Definition 3.26. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc and let λ be a positive real. The *proximal mapping* $\operatorname{prox}_\lambda^f$ is defined by

$$\operatorname{prox}_\lambda^f(x) = \operatorname{argmin}_{y\in\mathbb{R}^n} \left\{ f(y) + \tfrac{\lambda}{2} \|x - y\|^2 \right\}.$$

Recall that we say $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *coercive*, if $\lim_{\|x\|\to\infty} f(x) = \infty$.

Fact 3.27. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, lsc and coercive. Then f attains its minimum.*

Proof. See [21, Theorem 1.9]. □

Fact 3.28. *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 $v \in \mathbb{R}^n$ we have*

$$\operatorname{prox}_\lambda^f\left(x - \tfrac{1}{\lambda}v\right) = \operatorname{argmin}_y \left\{\langle y - x, v\rangle + \tfrac{\lambda}{2}\|x - y\|^2 + f(y)\right\}, \forall x \in \mathbb{R}^n.$$

Proof. Fix $x \in \mathbb{R}^n$. It is easy to see that $y \mapsto f(y) + \frac{\lambda}{2}\|x - y\|^2$ is proper and lsc. For every $y \in \operatorname{dom} f$, one has

$$\begin{aligned} f(y) + \frac{\lambda}{2}\|x - y\|^2 &\geq \inf_{\mathbb{R}^n} f + \frac{\lambda}{2}\|x - y\|^2 = \inf_{\mathbb{R}^n} f + \frac{\lambda}{2}(\|x\|^2 + \|y\|^2 - 2\langle x, y\rangle) \\ &\geq \inf_{\mathbb{R}^n} f + \frac{\lambda}{2}\|x\|^2 + \frac{\lambda}{2}\|y\|\left(\|y\| - 2\|x\|\right) \to \infty, \text{ as } \|y\| \to \infty. \end{aligned}$$

This means $y \mapsto f(y) + \frac{\lambda}{2}\|x - y\|^2$ is coercive. Therefore by applying Fact 3.27 the proximal mapping $\operatorname{prox}_\lambda^f(x)$ is nonempty. On the other hand, we have

$$\begin{aligned} \operatorname{prox}_\lambda^f\left(x - \tfrac{1}{\lambda}v\right) &= \operatorname{argmin}_y \left\{f(y) + \tfrac{\lambda}{2}\left\|x - \tfrac{1}{\lambda}v - y\right\|^2\right\} \\ &= \operatorname{argmin}_y \left\{f(y) + \tfrac{\lambda}{2}\|x - y\|^2 + \langle y - x, v\rangle + \tfrac{\lambda}{2}\left\|\tfrac{1}{\lambda}v\right\|^2\right\} \\ &= \operatorname{argmin}_y \left\{\langle y - x, v\rangle + \tfrac{\lambda}{2}\|x - y\|^2 + f(y)\right\}, \end{aligned}$$

which completes the proof. □

Example 3.29. Let $f(x) = |x|$ and let $\lambda > 0$. Then for $x \in \mathbb{R}$, $\operatorname{prox}_\lambda^f(x) = [|x| - 1/\lambda]_+ \cdot \operatorname{sgn}(x)$, where $[u]_+ = \max\{0, u\}$.

Proof. See [4, Example 24.20]. □

Chapter 4

An introduction to the Kurdyka-Łojasiewicz property

This chapter is devoted to introducing the Kurdyka-Łojasiewicz property, a concept at the heart of many proximal algorithms. We will use the notation $[0 < f < r] = \{x : 0 < f(x) < r\}$. For $\eta \in (0, \infty]$, denote by $\mathcal{K}_\eta$ the class of functions $\varphi : [0, \eta) \to \mathbb{R}_+$ satisfying

(i) $\varphi : [0, \eta) \to \mathbb{R}_+$ is concave and continuous with $\varphi(0) = 0$;

(ii) φ is C^1 on $(0, \eta)$;

(iii) $\varphi'(t) > 0$ for all $t \in (0, \eta)$.

The following definition is formulated by Attouch et al. [1, Definition 3.1] and known as the standard definition of the KL property:

Definition 4.1. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc. We say f has the *Kurdyka-Łojasiewicz property* (KL) at $\bar{x} \in \operatorname{dom} \partial f$, if there exist neighborhood $U \ni \bar{x}$, $\eta \in (0, \infty]$ and a function $\varphi \in \mathcal{K}_\eta$ such that for all $x \in U \cap [0 < f - f(\bar{x}) < \eta]$,

$$\varphi'(f(x) - f(\bar{x})) \cdot \operatorname{dist}(0, \partial f(x)) \geq 1. \tag{4.1}$$

The function φ is called a *desingularizing function* of f at $\bar{x}$. We say f is a *KL function* if it has the KL property at every $\bar{x} \in \operatorname{dom} \partial f$.

Remark 4.2. (i) Throughout this thesis, we say f satisfies the KL property at $\bar{x}$ with respect to U, η and $\varphi(t)$, if (4.1) holds for $x \in U \cap [0 < f - f(\bar{x}) < \eta]$. The KL property at $\bar{x}$ can be interpreted as whether the function f is "amenable to sharpness". To be specific, we say f is *sharp* around $\bar{x}$ if for some neighborhood $U \ni \bar{x}$, we have $\operatorname{dist}(0, \partial f(x)) \geq 1$ on $U \backslash \{\bar{x}\}$. Then KL property at $\bar{x}$ means that f can be "sharpened" around $\bar{x}$ by reparameterizing its value with $\varphi \in \mathcal{K}_\eta$, see Example 5.1.

(ii) Verifying the KL property of f at $\bar{x}$ can be viewed as a process of finding a desingularizing function $\varphi \in \mathcal{K}_\eta$ such that (4.1) holds for given neighborhood $U \ni \bar{x}$ and $\eta > 0$. Most published results emphasize desingularizing functions of the form

$$\varphi(t) = m \cdot t^{1-\theta}$$

for $m > 0$ and *KL exponent* $\theta \in [0, 1)$. Suppose that $m = 1/[c(1-\theta)]$ for some $c > 0$. Then one sees easily that (4.1) reduces to (4.2).

(iii) An interesting feature of the KL property is that desingularizing functions for given $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, $U \ni \bar{x}$ and $\eta > 0$ are numerous. This feature will be made clear as we present examples and facts of KL functions.

4.1 The origin and recent developments

In this section we provide a brief historical review of the KL property. We refer to [8] for a comprehensive survey of this object.

Originated from *algebraic geometry*, the KL property is named after Stanislaw Łojasiewicz and Krzysztof Kurdyka, whose work laid the foundation of this object. Recall that an infinitely differentiable function f is *real-analytic*, if f is representable by a power series with real coefficients, i.e., $f(x) = \sum_{n=1}^{\infty} a_n (x - x_0)^n$. In 1963, Łojasiewicz [15] proved that for a real-analytic function f and a critical point $\bar{x}$, there exist $\theta \in [0, 1)$ and $c > 0$ such that

$$\frac{|f(x) - f(\bar{x})|^\theta}{\|\nabla f(x)\|} \leq c,$$

for x around $\bar{x}$. The above inequality was then extended by Kurdyka [13] to the framework of C^1 functions.

After the pioneering work of Łojasiewicz and Kurdyka, researchers turned to obtain a nonsmooth version of the KL inequality. Recall the following definition:

Definition 4.3. [6] (i) A subset A of $\mathbb{R}^n$ is called *semianalytic*, if each point of $\mathbb{R}^n$ admits a neighborhood V for which $A \cap V$ assumes the form:

$$\bigcup_{i=1}^{p} \bigcap_{j=1}^{q} \{x \in V : f_{ij}(x) = 0, g_{ij}(x) > 0\},$$

where functions $f_{ij}, g_{ij} : V \to \mathbb{R}$ are real-analytic functions for all $1 \leq i \leq p$, $1 \leq j \leq q$.

(ii) The set A is called *subanalytic* if each point of $\mathbb{R}^n$ admits a neighborhood V such that

$$A \cap V = \{x \in \mathbb{R}^n : (x, y) \in B\}$$

where B is a bounded semianalytic subset in $\mathbb{R}^n \times \mathbb{R}^m$ for some $m \geq 1$.

(iii) A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called *subanalytic* if its graph is a subanalytic subset of $\mathbb{R}^n \times \mathbb{R}$.

Bolte et al. [6, Theorem 3.1] proved that for a subanalytic function f with closed domain on which f is continuous, there exists $\theta \in [0, 1)$ such that

$$|f(x) - f(\bar{x})|^{-\theta} \cdot \operatorname{dist}(0, \partial f(x)) \geq c, \tag{4.2}$$

for some $c > 0$ and x around critical point $\bar{x}$, and studied the applications of (4.2) in subgradient dynamical system. To the best of our knowledge, the work by Bolte et al. [6] is the first article that brought the KL property to the optimization community.

The standard definition of the KL property (Definition 4.1) is introduced by Attouch et al. [1, Definition 3.1] as a generalization of (4.2). Indeed, (4.1) reduces to (4.2) if the desingularizing function $\varphi(t) \in \mathcal{K}_\eta$ assumes the form

$$\varphi(t) = \frac{t^{1-\theta}}{c(1-\theta)}.$$

The scalar $\theta \in [0, 1)$ is known as the *KL exponent*. It is worth mentioning that although the class $\mathcal{K}_\eta$ consists of functions with various forms, most literature emphasizes on the form $\varphi(t) = t^{1-\theta}/c(1-\theta)$.

The KL property have significant impacts on optimization. We refer to Section 4.3 for a brief introduction and Chapter 8 for a discussion in detail.

4.2 Classic results

Now we collect some well-known facts about the KL property.

4.2.1 The KL property at non-stationary points

Fact 4.4. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc. Let $\bar{x} \in \operatorname{dom} \partial f$ be a non-stationary point and let $\theta \in [0, 1)$. Then there exists $\varepsilon \in (0, 1]$ such that f has the KL property at $\bar{x}$ with respect to $U = \mathbb{B}(\bar{x}; \varepsilon)$, $\eta = \varepsilon$ and $\varphi_\theta(t) = \frac{t^{1-\theta}}{\varepsilon(1-\theta)}$.*

Proof. We show first that there exists $\varepsilon \in (0, 1]$ such that

$$\operatorname{dist}(0, \partial f(x)) \geq \varepsilon$$

whenever $\|x - \bar{x}\| < \varepsilon$ and $|f(x) - f(\bar{x})| < \varepsilon$. If not, then we can take $x_n \to \bar{x}$ with $f(x_n) \to f(\bar{x})$ and $v_n \in \partial f(x_n)$ such that

$$\|v_n\| \leq \frac{1}{n}, \forall n \in \mathbb{N}.$$

Taking $n \to \infty$, one gets by Fact 3.24 that $0 \in \operatorname{Limsup}_{x \xrightarrow{f} \bar{x}} \partial f(x) \subseteq \partial f(\bar{x})$, which contradicts to the assumption that $\bar{x}$ is a non-stationary point. For all $x \in \mathbb{B}(\bar{x}; \varepsilon) \cap [0 < f - f(\bar{x}) < \varepsilon]$ we have

$$\frac{1}{\varepsilon}(f(x) - f(\bar{x}))^{-\theta} \operatorname{dist}(0, \partial f(x)) \geq (f(x) - f(\bar{x}))^{-\theta} \geq \varepsilon^{-\theta} \geq 1,$$

where the last inequality holds because $\varepsilon \in (0, 1]$. □

Remark 4.5. (i) It is noted in [1, Remark 3.2(b)] that proper lsc functions satisfy the KL property at non-stationary points. This fact is then proved in detail by Li and Pong [14]. We provide a proof here for self-containedness.

(ii) According to Fact 4.4, we only need to consider stationary points when verifying the KL property.

Corollary 4.6. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc. If $0 \notin \partial f(x)$ for every $x \in \operatorname{dom} \partial f$ then f is a KL function.*

Proof. Apply Fact 4.4. □

Example 4.7. The following functions are KL:

(i) The Sigmoid function $f(x) = 1/(1 + e^{-x})$.

(ii) The log-barrier function

$$g(x) = \begin{cases} -\ln(a - x), & x < a; \\ \infty, & \text{otherwise}, \end{cases}$$

where $a > 0$.

Proof. Note that $f'(x) = e^x/(1+e^x)^2 > 0$ for $x \in \mathbb{R}$ and $g'(x) = 1/(a-x) > 0$ for $x < a$. Hence Corollary 4.6 assures that f and g are KL. □

4.2.2 Semialgebraic examples

Another celebrated result states that semialgebraic functions are KL.

Definition 4.8. (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 $f : \mathbb{R}^n \to (-\infty, \infty]$ is called *semialgebraic*, if its graph

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

is semialgebraic.

Fact 4.9. *[9] Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a proper and lsc function and let $\bar{x} \in \operatorname{dom} \partial f$. If f is semialgebraic, then it satisfies the KL property at $\bar{x}$ with $\varphi(t) = c \cdot t^{1-\theta}$ for some $c > 0$ and $\theta \in [0, 1)$.*

Remark 4.10. (i) Fact 4.9 only asserts the existence of certain $\theta \in [0, 1)$ and $c > 0$ such that $\varphi(t) = c \cdot t^{1-\theta} \in \mathcal{K}_\eta$ and (4.1) holds.

(ii) Note that semialgebraic functions are semianalytic with the neighborhood $V \ni x$ being $\mathbb{R}^n$ for every $x \in \mathbb{R}^n$. Indeed, for a polynomial with degree n, its k-th derivative vanishes for $k > n$, which means the Taylor expansion converges. Hence polynomials are real-analytic, which implies that semialgebraic functions are semianalytic.

(iii) It has been shown that a more general class of functions satisfies the KL property, namely proper and lsc functions definable on an o-minimal structure [7]. For example, semialgebraic functions are definable on their respective structure [7, Remark 5(i)]. To avoid further complication, we will not elaborate this result.

Example 4.11. Many functions arising from optimization problems are semialgebraic. Here are some examples borrowed from [9]:

(i) Real polynomial functions.

(ii) Indicator functions of semialgebraic sets.

(iii) The sparsity measure $\|\cdot\|_0$ defined by

$$\|x\|_0 = \text{Number of non-zero coordinates of } x.$$

Proof. (i) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a polynomial and let $F : \mathbb{R}^{n+1} \to \mathbb{R}$ be defined by $F(x, y) = f(x) - y$. Then F is a polynomial and

$$\operatorname{gph} f = \{(x, y) \in \mathbb{R}^{n+1} : f(x) = y\} = \{(x, y) \in \mathbb{R}^{n+1} : F(x, y) = 0\},$$

which implies that $\operatorname{gph} f$ is semialgebraic.

(ii) Let $A \subseteq \mathbb{R}^n$ be a semialgebraic set and suppose that for polynomials $g_{ij}, h_{ij} : \mathbb{R}^n \to \mathbb{R}$ one has

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

Then we have

$$\begin{aligned}
\operatorname{gph} \delta_A &= \{(x, y) \in \mathbb{R}^{n+1} : \delta_A(x) = y\} = \{(x, y) \in \mathbb{R}^{n+1} : x \in A, y = 0\} \\
&= \bigcup_{j=1}^{p} \bigcap_{i=1}^{q} \{(x, y) \in \mathbb{R}^{n+1} : g_{ij}(x) = 0, h_{ij}(x) < 0, y = 0\} \\
&= \bigcup_{j=1}^{p} \bigcap_{i=1}^{q} \{(x, y) \in \mathbb{R}^{n+1} : G_{ij}(x, y) = 0, H_{ij}(x, y) < 0, F(x, y) = 0\},
\end{aligned}$$

where $G_{ij}(x, y) = g_{ij}(x), H_{ij}(x, y) = h_{ij}(x)$ and $F(x, y) = y$ are polynomials on $\mathbb{R}^n \times \mathbb{R}$.

(iii) Let $n \in \mathbb{N}$ and let $I \subseteq \{1, \ldots, n\}$. Define $f_I, g_I : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ by

$$f_I(x, y) = \prod_{i=1, i \notin I}^{n} x_i \text{ and } g_I(x, y) = y - n + |I|,$$

where x_i is the i-th coordinate of $x \in \mathbb{R}^n$ and $|I|$ denotes the cardinality of set I. Define for $i \in \{1, \ldots, n\}$

$$J_i^I = \begin{cases} 0, & i \in I; \\ \mathbb{R} \backslash \{0\}, & i \notin I. \end{cases}$$

and $h_i(x, y) = x_i$. Then for vector $x \in \mathbb{R}^n$ with $\|x\|_0 = n - |I|$, we have

$$(x, \|x\|_0) \in \left(\prod_{i=1}^{n} J_i^I \right) \times \{n - |I|\}.$$

Observe the following:

$$\left(\prod_{i=1}^{n} J_i^I\right) \times \{n - |I|\} = \{(x, n - |I|) : x_i = 0 \text{ for } i \in I, x_i \neq 0 \text{ for } i \notin I\}$$
$$= \{(x, y) : h_i(x, y) = 0 \text{ for } i \in I, f_I(x, y) \neq 0, g_I(x, y) = 0\}$$
$$= \{(x, y) : h_i(x, y) = 0 \text{ for } i \in I, f_I(x, y) > 0, g_I(x, y) = 0\}$$
$$\cup \{(x, y) : h_i(x, y) = 0 \text{ for } i \in I, f_I(x, y) < 0, g_I(x, y) = 0\},$$

Hence the graph of $\|\cdot\|_0$

$$\operatorname{gph} \|\cdot\|_0 = \bigcup_{I \subseteq \{1,\ldots,n\}} \left[\left(\prod_{i=1}^{n} J_i^I\right) \times \{n - |I|\}\right]$$

is semialgebraic. □

Example 4.12. The exponential function $f(x) = e^x$ is not semialgebraic.

Proof. We will prove the statement by contradiction. Suppose that $\operatorname{gph} f$ was semialgebraic. Then

$$\operatorname{gph} f = \{(x, e^x) : x \in \mathbb{R}\} = \bigcup_{j=1}^{p} \bigcap_{i=1}^{q} \{(x, y) \in \mathbb{R}^2 : g_{ij}(x, y) = 0, h_{ij}(x, y) < 0\}$$

for some polynomials $g_{ij}, h_{ij} : \mathbb{R}^2 \to \mathbb{R}$.

We claim that to each $j \in \{1, \ldots, p\}$ corresponds at least one $i \in \{1, \ldots, q\}$ such that $G_j(x, y) = g_{ij}(x, y)$ is a non-zero polynomial. Otherwise there would be some j_0 such that $g_{ij_0}(x, y) \equiv 0$ for every $i \in \{1, \ldots, q\}$, which further implies that

$$O = \bigcap_{i=1}^{q} \{(x, y) \in \mathbb{R}^2 : g_{ij_0}(x, y) = 0, h_{ij_0}(x, y) < 0\}$$
$$= \bigcap_{i=1}^{q} \{(x, y) \in \mathbb{R}^2 : h_{ij_0}(x, y) < 0\} \subseteq \operatorname{gph} f.$$

Note that $[h_{ij_0} < 0]$ is open for every $i \in \{1, \ldots, q\}$. Then the curve $\operatorname{gph} f$ contains an open set O, which is absurd.

Define $G(x, y) = \prod_{j=1}^{p} G_j(x, y)$. Then $G(x, y)$ vanishes on $\operatorname{gph} f$. Suppose that $G(x, y)$ has degree $n \in \mathbb{N} \cup \{0\}$. If $n = 0$, then $G(x, y) \equiv c$ for some constant $c \neq 0$ and one gets a contradiction immediately because $G(x, y)$ has

no roots. Moreover, if $G(x, y) = G(x)$ then we have $G(x) \equiv 0$, contradicting to the fact that $G(x)$ is a non-zero polynomial. Now we consider the case where $n \geq 1$ and G has two variables. In this case, rearranging if necessary, we assume that G has the following form:

$$G(x, y) = \sum_{k=0}^{n} h_k(x) y_k - h(x),$$

for nonzero polynomials h_k, $k = 1, \ldots, n$, and a possibly zero polynomial $h(x)$. Then $G(x, e^x) = 0$ implies that

$$h(x) = \sum_{k=0}^{n} h_k(x) e^{kx}, \forall x \in \mathbb{R}.$$

Suppose that h has degree $m \leq n$. Then by taking the $m+1$-th order derivative, one gets for some nonzero polynomials $\tilde{h}_k$

$$0 = \sum_{k=0}^{n} \tilde{h}_k(x) e^{kx}, \forall x \in \mathbb{R},$$

which contradicts to the fact that polynomial has finitely many roots and $e^x > 0$. Note that $\tilde{h}_k, k = 1, \ldots, m$ are polynomials because they are linear combinations of $h_k^{(j)}(x), k = 1, \ldots, n, j = 0, \ldots, m+1$. □

4.2.3 Convex examples

The KL property is also related to convexity through a specific feature of convex functions.

Fact 4.13. *Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a proper, lsc and convex function. Let $\bar{x} \in \operatorname{argmin} f \neq \emptyset$. Assume that f satisfies the following growth condition: There exist neighborhood $U \ni \bar{x}$, $\eta > 0$, $c > 0$ and $r \geq 1$ such that for $x \in U \cap [\min f < f < \min f + \eta]$*

$$f(x) \geq f(\bar{x}) + c \cdot \operatorname{dist}(x, \operatorname{argmin} f)^r. \tag{4.3}$$

Then f has the KL property at $\bar{x}$ for $\varphi(t) = rc^{-1/r} t^{1/r}$ on the set $x \in U \cap [\min f < f < \min f + \eta]$.

Proof. Let $x \in \operatorname{dom} \partial f$, $x^* \in \operatorname{argmin} f$ and let $v \in \partial f(x)$. Then one has by convexity and Fact 3.17 that

$$f(x) - \min f = f(x) - f(x^*) \leq \langle v, x - x^* \rangle \leq \|v\| \cdot \|x^* - x\|.$$

Taking the infimum over all $v \in \partial f(x)$ and $x^* \in \operatorname{argmin} f$, one has

$$\operatorname{dist}(0, \partial f(x)) \geq \frac{f(x) - \min f}{\operatorname{dist}(x, \operatorname{argmin} f)}. \tag{4.4}$$

Note that $\varphi'(t) = c^{-1/r} t^{1/r-1}$. Hence we have for $x \in U \cap [\min f < f < \min f + \eta]$

$$\begin{aligned}\varphi'(f(x) - f(\bar{x})) \cdot \operatorname{dist}(0, \partial f(x)) &= c^{-1/r}(f(x) - \min f)^{1/r-1} \cdot \operatorname{dist}(0, \partial f(x)) \\ &\geq c^{-1/r} \frac{(f(x) - \min f)^{1/r}}{\operatorname{dist}(x, \operatorname{argmin} f)} \geq 1,\end{aligned}$$

where the first inequality follows from (4.4) and the last one is implied by (4.3). □

Remark 4.14. Fact 4.13 first appeared in [6, Remark 3.5] as a modification for [6, Theorem 3.3], but no proof is given. We provided a proof here for the sake of self-containedness.

Example 4.15. Let $f_1(x) = |x|^p$, $f_2(x) = -\ln(1 - |x|^p)$ and $f_3(x) = \tan(|x|^p)$ for $p \geq 1$. For each $i \in \{1, 2, 3\}$, $f_i(x)$ satisfies the KL property at $\bar{x} = 0$ with respect to $U = \operatorname{dom} f_i$, $\eta = \infty$ and $\varphi(t) = p \cdot t^{1/p}$.

Proof. Note that $(-\ln(1-t))' = 1/(1-t) \geq 1$ for $0 \leq t < 1$ and $(\tan(t))' = 1/\cos^2(t) \geq 1$ for $t \in [0, \pi/2)$. Hence for $i \in \{1, 2, 3\}$ we have $f_i(x) \geq |x|^p$ on $\operatorname{dom} f_i$. Furthermore one has for $x \in \mathbb{R} \cap [0 < f_i - f_i(\bar{x}) < \infty]$

$$f_i(x) \geq f_i(\bar{x}) + |x|^p = \operatorname{dist}(x, \operatorname{argmin} f_i)^p,$$

where the last equality holds because f_i has unique minimizer 0 for each $i \in \{1, \ldots, 3\}$. Hence by applying Fact 4.13 we conclude that f_i has the KL property with respect to $U = \mathbb{R}$, $\eta = \infty$ and $\varphi(t) = rc^{-1/r}t^{1/r}$ with $c = 1$ and $r = p$. □

What follows is a special case of Fact 4.13.

Fact 4.16. *[1] Assume that $f : \mathbb{R}^n \to (-\infty, \infty]$ is uniformly convex with modulus $\phi = c|\cdot|^r$. Then f has the KL property on* $\operatorname{dom} f$ *with $U = \operatorname{dom} f$, $\eta = \infty$ and $\varphi(t) = rc^{-1/r}t^{1/r}$.*

Proof. By (3.4) for every $x, y \in \operatorname{dom} f$ and $\lambda \in (0, 1)$

$$\begin{aligned}\phi(\|x - y\|) &\leq \frac{\lambda f(x) + (1-\lambda) f(y) - f(\lambda x + (1-\lambda) y)}{\lambda(1-\lambda)} \\ &= \frac{f(x) - f(y)}{1-\lambda} + \frac{f(y) - f(y + \lambda(x-y))}{\lambda(1-\lambda)}.\end{aligned}$$

Taking $\lambda \to 0^+$ we get $\phi(\|x-y\|) \le f(x)-f(y)-f'(y;x-y)$ and furthermore by Fact 3.11

$$f(x) \ge f(y) + \phi(\|x-y\|) + \langle g, x-y\rangle$$

for every $g \in \partial f(y)$. In particular for $\bar{x} \in \operatorname{argmin} f$ and $0 \in \partial f(\bar{x})$ we have

$$f(x) \ge f(\bar{x}) + \phi(\|x-\bar{x}\|) = f(\bar{x}) + c\cdot\|x-\bar{x}\|^r \ge f(x) + c\cdot \operatorname{dist}(x, \operatorname{argmin} f)^r,$$

Hence applying Fact 4.13 completes the proof. □

Example 4.17. Consider $f(x) = x^4$. Define $\phi(t) = c\cdot t^4$ for $c > 0$. If $c \le 1/8$, then f satisfies the KL property at 0 with respect to $U = \mathbb{R}$, $\eta = \infty$ and $\varphi(t) = 4c^{-1/4}t^{1/4}$.

Proof. By Example 3.8(ii), $f(x)$ is uniformly convex with modulus $\phi(t)$ if $\phi(t) \le t^4/8$ on $[0,\infty)$. Hence applying Fact 4.16 completes the proof. □

However, convex functions are not necessarily KL. Bolte et al. claimed in [9, Page 492] the function f given below fails to satisfy Definition 4.1.

Fact 4.18. *There exists a C^2 convex function $f : \mathbb{R}^2 \to \mathbb{R}$ with $\min f = 0$ whose set of minimizers is the unit disk. Moreover, for every $\eta > 0$ and continuous strictly increasing function $\varphi : [0,\eta] \to \mathbb{R}$ that is C^1 on $(0,\eta)$ and $\varphi(0) = 0$, we have*

$$\inf\{\|\nabla(\varphi\circ f)(x)\| : x \in [0 < f < \eta]\} = 0. \tag{4.5}$$

Proof. See [8, Theorem 36]. □

Remark 4.19. It is not straightforward to see how the function f given above fails to satisfy Definition 4.1. However no proof has been given yet.

In contrast, the proof becomes easy if we adopt the following definition of the KL property in [8]: Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper function with $\inf f > -\infty$. We say that f has the KL property if there exist $\eta > \inf f$ and a continuous strictly increasing function $\varphi : [\inf f, \eta) \to \mathbb{R}$ that is C^1 on $(\inf f, \eta)$ and $\varphi(\inf f) = 0$ such that for $x \in [\inf f < f < \eta]$

$$\operatorname{dist}(0, \partial(\varphi\circ f)(x)) \ge 1. \tag{4.6}$$

Indeed, if (4.6) was true for some $\eta_0 > 0$ and φ_0 then we would have for every $x \in [0 < f < \eta_0]$

$$\operatorname{dist}(0, \partial(\varphi_0\circ f)(x)) = \|\nabla(\varphi_0\circ f)(x)\| \ge 1,$$

which further implies $\inf\{\|\nabla(\varphi_0\circ f)(x)\| : x \in [0 < f < \eta_0]\} \ge 1$ and contradicts to (4.5).

4.2.4 Morse functions

Recall a C^2 function $f : \mathbb{R}^n \to \mathbb{R}$ is a *Morse function*, if the Hessian $\nabla^2 f(\bar{x})$ at each critical point $\bar{x}$ is an invertible square matrix.

Attouch et al. noted in [1, Page 14] that Morse functions satisfy the KL property at $\bar{x}$ with $\phi(t) = c\sqrt{t}$ for some $c > 0$. However no detailed proof was given. A proof is provided below for the sake of completeness.

Theorem 4.20. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^2 function and let $\bar{x}$ be a critical point at which the Hessian $\nabla^2 f(\bar{x})$ is invertible. Then there exist $U \ni \bar{x}$ and $c > 0$ such that f has the KL property at $\bar{x}$ with respect to some neighborhood $U \ni \bar{x}$, $\eta > 0$ and $\varphi(t) = c \cdot t^{1/2}$.*

Proof. The Hessian $\nabla^2 f(\bar{x})$ is assumed to be invertible, which means that $\nabla^2 f(\bar{x})$ is injective and $\nabla^2 f(\bar{x}) \cdot u \neq 0$ for every nonzero $u \in \mathbb{R}^n$. Define the following quantities:

$$c_1 = \sup_{\|u\|=1} \left\|\nabla^2 f(\bar{x}) \cdot u\right\| \text{ and } c_2 = \inf_{\|u\|=1} \left\|\nabla^2 f(\bar{x}) \cdot u\right\|.$$

Then c_1 and c_2 are positive reals by applying Fact 2.7 to the continuous map $u \mapsto \left\|\nabla^2 f(\bar{x}) \cdot u\right\|$. The Taylor expansion of f at $\bar{x}$ gives

$$f(x) - f(\bar{x}) = \langle x - \bar{x}, \nabla^2 f(\bar{x})(x - \bar{x})\rangle + o(\|x - \bar{x}\|^2),$$

which further implies that

$$\frac{|f(x) - f(\bar{x})|}{\|x - \bar{x}\|^2} \leq \left\|\nabla^2 f(\bar{x}) \cdot \frac{x - \bar{x}}{\|x - \bar{x}\|}\right\| + \frac{|o(\|x - \bar{x}\|^2)|}{\|x - \bar{x}\|^2} \leq c_1 + \frac{|o(\|x - \bar{x}\|^2)|}{\|x - \bar{x}\|^2}.$$

We claim that there is $\varepsilon_1 > 0$ such that for every x with $\|x - \bar{x}\| < \varepsilon_1$

$$\frac{|f(x) - f(\bar{x})|}{\|x - \bar{x}\|^2} \leq 2c_1. \tag{4.7}$$

Otherwise one could take $x_n \to \bar{x}$ with

$$2c_1 < c_1 + \frac{|o(\|x_n - \bar{x}\|^2)|}{\|x_n - \bar{x}\|^2}, \forall n \in \mathbb{N} \Rightarrow 2c_1 \leq c_1,$$

which is absurd as $c_1 > 0$.

Now we apply the Taylor expansion to $\nabla f(\bar{x})$, which yields

$$\nabla f(x) = \nabla^2 f(\bar{x}) \cdot (x - \bar{x}) + o(\|x - \bar{x}\|). \tag{4.8}$$

Suppose for every $\varepsilon > 0$ there exists $x \in \mathbb{B}(\bar{x}; \varepsilon)$ such that

$$\frac{\|\nabla f(x)\|}{\|x - \bar{x}\|} < \frac{c_2}{2}.$$

Then we can take a sequence $x_n \to \bar{x}$ such that for every $n \in \mathbb{N}$

$$\frac{c_2}{2} > \frac{\|\nabla f(x_n)\|}{\|x_n - \bar{x}\|} \geq \frac{\left\|\nabla^2 f(\bar{x}) \cdot (x_n - \bar{x})\right\|}{\|x_n - \bar{x}\|} - \left\| \frac{o(\|x_n - \bar{x}\|)}{\|x_n - \bar{x}\|} \right\| \geq c_2 - \left\| \frac{o(\|x_n - \bar{x}\|)}{\|x_n - \bar{x}\|} \right\|,$$

where the second inequality is implied by triangle inequality. Taking $n \to \infty$, we get $c_2/2 \geq c_2$ which is absurd because $c_2 > 0$. Hence for some $\varepsilon_2 > 0$ and all x with $\|x - \bar{x}\| < \varepsilon_2$

$$\frac{\|\nabla f(x)\|}{\|x - \bar{x}\|} \geq \frac{c_2}{2}. \tag{4.9}$$

Let $\eta > 0$ and let $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$. Then for every $x \in \mathbb{R}^n$ with $\|x - \bar{x}\| < \varepsilon$ and $0 < f(x) - f(\bar{x}) < \eta$,

$$2c_2^{-1}\sqrt{2c_1}|f(x) - f(\bar{x})|^{-\frac{1}{2}} \cdot \|\nabla f(x)\| \geq 2c_2^{-1} \|x - \bar{x}\|^{-1} \|\nabla f(x)\| \geq 1,$$

where the first inequality follows from (4.7) and the second one follows from (4.9). Hence one concludes that f has the KL property at $\bar{x}$ with respect to $U = \mathbb{B}(\bar{x}; \varepsilon)$, $\eta > 0$ and $\varphi(t) = 4c_2^{-1}\sqrt{2c_1} \cdot t^{1/2}$. □

Example 4.21. The negative *Boltzmann-Shannon* entropy defined by

$$f(x) = \begin{cases} \sum_{i=1}^n x_i \ln(x_i) - x_i, & x \in \mathbb{R}^n_{++}; \\ 0, & x \in \mathbb{R}^n_+ \backslash \mathbb{R}^n_{++}; \\ \infty, & \text{otherwise} \end{cases}$$

has the KL property at $\bar{x} = (1, \ldots, 1)$ with respect to some neighborhood $U \ni \bar{x}$, $\eta = \infty$ and $\varphi(t) = 4\sqrt{2t}$.

Proof. Note that $\nabla^2 f(\bar{x}) = \mathrm{Id}$. Hence we have

$$c_1 = \sup_{\|u\|=1} \left\|\nabla^2 f(\bar{x}) \cdot u\right\| = \sup_{\|u\|=1} \|u\| = 1,$$

$$c_2 = \inf_{\|u\|=1} \left\|\nabla^2 f(\bar{x}) \cdot u\right\| = \inf_{\|u\|=1} \|u\| = 1.$$

Applying Theorem 4.20 completes the proof. □

4.3 The KL exponent and its impact on optimization

Recall that we say $\theta \in [0,1)$ is a *KL exponent* for f at $\bar{x}$, if there exist $c > 0$, a neighborhood $U \ni \bar{x}$ and $\eta \in (0,\infty]$ such that for $x \in U \cap [0 < f - f(\bar{x}) < \eta]$, one has

$$|f(x) - f(\bar{x})|^{-\theta} \cdot \operatorname{dist}(0, \partial f(x)) \geq c,$$

which means that f has the KL property with desingularizing function $\varphi(t) = t^{1-\theta}/\left(c(1-\theta)\right)$.

The KL property has been shown to play a key role in many proximal algorithms. The work by Bolte et al. [9] is one of the cornerstones. Bolte et al. introduced the PALM algorithm to solve a nonsmooth and nonconvex problem. They showed that the consecutive gap of a sequence generated by the PALM algorithm is summable under the KL property. Their technique has been employed by many researchers. For instance, Banert and Bot [3] recently designed a proximal gradient algorithm for solving D. C. programming problem. The convergence of their algorithm is guaranteed by the KL property. The very recent work by Yu et al. [23] also used the KL property to ensure the convergence of their proposed algorithm. A prototypical result of convergence rate can be summarized from theses studies:

Proposition 4.22. *[14] Consider a certain algorithm of interest and objective function Ψ. Suppose that Ψ satisfies the KL property with an exponent $\theta \in [0,1)$, and that the sequence $(z_k)_{k\in\mathbb{N}}$ generated by a algorithm is bounded. Then the following assertions hold:*

(i) If $\theta = 0$, $(z_k)_{k\in\mathbb{N}}$ converges in finitely many steps.

(ii) If $\theta \in (0, \frac{1}{2}]$, $(z_k)_{k\in\mathbb{N}}$ converges locally linearly.

(iii) If $\theta \in (\frac{1}{2}, 1)$, $(z_k)_{k\in\mathbb{N}}$ converges locally sublinearly.

Remark 4.23. Here we only provide a prototypical result because no algorithm has been introduced yet. See Fact 8.10 for a specific version of Proposition 4.22. Similar results can also be found in [3, Lemma 4] and [23, Theorem 3.8].

Despite the above prototypical result shows a promising estimation of the convergence rate, in practice one still needs to find an explicit formula or at least an estimation for the KL exponent of the objective function. However, the KL exponent is usually hard to determine or estimate, as noted in [16,

Page 63, Section 2.1]. For instance, we learn from Fact 4.9 and Example 4.11 that polynomials are KL functions and admit desingularizing functions of the form $\varphi(t) = c \cdot t^{1-\theta}$. However, only very recently did Bolte et al. [10] obtain an explicit formula of the KL exponent for piecewise polynomial convex functions on $\mathbb{R}^n$. Even for nonconvex polynomials on the line, the explicit formula of KL exponent is still unknown. Li and Pong [14] also obtained the KL exponent for several structured optimization models, by deducing the KL exponent from a specific structure called Luo-Tseng error bound and developing various calculus rules for KL exponents.

4.4 Challenges and open questions

To summarize, KL functions are ubiquitous and have powerful applications in optimization. However, the following aspects are challenging or not well-understood in the current literature:

- The classic characterizations of the KL property, such as the semialgebraic and subanalytical arguments, are far away from the usual context of nonsmooth analysis and emphasize on desingularizing functions of the form $\varphi(t) = c \cdot t^{1-\theta}$. On the other hand, it is usually challenging to determine the KL exponent θ as we discussed in the previous section, while in practice it is important to have this information.

- When verifying the KL property of f at $\bar{x}$ with respect to given $U \ni \bar{x}$ and $\eta \in (0, \infty]$, one needs to find a desingularizing function to "sharpen" f. As Fact 4.4 and Example 4.17 suggest, for a given function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, desingularzing functions $\varphi \in \mathcal{K}_\eta$ of the usual form $\varphi(t) = c \cdot t^{1-\theta}$ such that KL property holds at $\bar{x} \in \operatorname{dom} \partial f$ are numerous. On the other hand, desingularizing functions can have forms other than the prominent one. For instance, the function $x \mapsto \sin(x)$ admits $\varphi(t) = \arcsin(t-1) + \pi/2$ at $3\pi/2$, see Example 5.6. Such observation leads to a natural question:

  What is the optimal desingularizing function with respect to U and η?

  However, to the best of our knowledge, there is no discussion about the aforementioned question in the current literature.

We attack the above limitations in the rest of this thesis:

- In Chapter 5, we develop several characterizations of the KL property using tools from nonsmooth analysis with explicit formulas for desingularizing functions. It is worth noting that our formulas do not assume the form $\varphi(t) = c \cdot t^{1-\theta}$.

- Chapter 6 is devoted to extending the classic KL property. We introduce the generalized KL property, an extension of the classic KL property given in Definition 4.1. We provide several examples to demonstrate the difference between this generalized notion and Definition 4.1. Various calculus rules are also developed.

- In Chapter 7, we define *the exact modulus of the generalized KL property* in order to address the question: *What is the optimal desingularizing function?* We will show that the exact modulus is indeed the optimal desingularizing function and provide several examples to illustrate this pleasant property. It is worth noting that the exact modulus is sometimes hard to be computed directly. Therefore we also provide a methodology for estimating the exact modulus.

- In Chapter 8, we revisit the celebrated *proximal alternating linearized minimization* (PALM) method and investigate the algorithmic impacts of the exact modulus of the generalized KL property on the PALM method. We will show that the exact modulus helps us to obtain *the sharpest upper bound* of the consecutive iterates gap of a sequence generated by the PALM algorithm, then study the interplay between the convergence rate of the PALM algorithm and the exact modulus of the generalized KL property.

Chapter 5

Characterizations of the KL property: A view from nonsmooth analysis

In this chapter, we obtain several sufficient conditions for the KL property, by using tools from nonsmooth analysis. Our results are new and provide explicit formula for the desingularizing function.

Section 5.1 is devoted to providing a self-contained analysis for examples and non-examples of the KL property by using the definition. This makes contrast to Chapter 4, where several examples are presented by using well-established results.

In Section 5.2, we provide a characterization of the KL property for locally convex C^1 functions on the line, then revisit the one-dimensional version of the well-known fact that polynomials are KL.

A sufficient condition of the KL property for nonsmooth convex functions on $\mathbb{R}^n$ is given in Section 5.3.

5.1 Examples of KL and non-KL functions

Example 5.1. (The first set of examples) The following statements hold:

(i) Let $f_1(x) = |x|$. Then f_1 has the KL property at 0 with respect to $U_1 = \mathbb{R}$, $\eta_1 = \infty$ and $\varphi_1(t) = t$.

(ii) Let $f_2(x) = x^2$. Then f_2 has the KL property at 0 with respect to $U_2 = \mathbb{R}$, $\eta_2 = \infty$ and $\varphi_2(t) = t^{1/2}$.

(iii) Let $f_3(x) = 1 - e^{-|x|}$. Then f_3 has the KL property at 0 with respect to $U_3 = \mathbb{R}$, $\eta_3 = 1 - e^{-1}$ and $\varphi_3(t) = e \cdot t$.

Proof. (i) According to Example 3.22 we have for $x \neq 0$

$$\partial f_1(x) = \begin{cases} 1, & x > 0; \\ -1, & x < 0. \end{cases} \Rightarrow \operatorname{dist}(0, \partial f_1(x)) = 1.$$

Hence for $x \in \mathbb{R} \cap [0 < f_1 < \infty] = \mathbb{R} \backslash \{0\}$,

$$\varphi_1'(f_1(x)) \cdot \operatorname{dist}(0, \partial f_1(x)) = \operatorname{dist}(0, \partial f_1(x)) = 1.$$

(ii) By applying Fact 3.18(ii) we have $\operatorname{dist}(0, \partial f_2(x)) = |f_2'(x)| = |2x|$. Then for $x \in \mathbb{R} \cap [0 < f_2 < \infty] = \mathbb{R} \backslash \{0\}$

$$\varphi_2'(f_2(x)) \cdot \operatorname{dist}(0, \partial f_2(x)) = \frac{1}{2\sqrt{x^2}} \cdot |2x| = \frac{1}{2|x|} \cdot |2x| = 1.$$

(iii) By using Fact 3.18(ii) we have for $x \neq 0$

$$\partial f_3(x) = \operatorname{sgn}(x) \cdot e^{-|x|} \Rightarrow \operatorname{dist}(0, \partial f_3(x)) = e^{-|x|}.$$

For $x \in \mathbb{R} \cap [0 < f_3 < 1 - e^{-1}] = (-1, 1) \backslash \{0\}$

$$\varphi_3'(f_3(x)) \cdot \operatorname{dist}(0, \partial f_3(x)) = e \cdot e^{-|x|} \geq e \cdot e^{-1} = 1,$$

which completes the proof. □

Remark 5.2. As noted in Remark 4.2(i), the KL property of f at $\bar{x}$ means that f can be sharpened around $\bar{x}$ through reparameterization. The function f_1 is already sharp around 0 because $\operatorname{dist}(0, \partial f_1(x)) = 1$ for nonzero x. Therefore reparameterization is not necessary and one simply needs to take $\varphi_1(t) = t$. In contrast, f_2 and f_3 are originally not sharp around 0, but can be made so by reparameterizing the function values with φ_2 and φ_3.

Example 5.3. (**Non-KL function**) Consider the following function

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

Then f fails to satisfy the KL property at 0.

Proof. Note that $f(x)$ is C^∞ according to Proposition 2.10. Then Fact 3.18(ii) implies that

$$\partial f(x) = f'(x) = \frac{2\sin(1/x)}{x^3 e^{1/x^2}} - \frac{\cos(1/x)}{x^2 e^{1/x^2}}$$

for $x \neq 0$. Let $x_n = (-1)^n/(2n\pi)$ and $y_n = (-1)^n/(\frac{\pi}{2} + 2n\pi)$. Then $x_n \to 0$ and $y_n \to 0$ as $n \to \infty$. Moreover, one has $f'(x_n) = -1/(x_n^2 e^{1/x_n^2}) < 0$ for each $n \in \mathbb{N}$ and $f'(y_n) = 2/(|y_n|^3 e^{1/y_n^2}) > 0$. Therefore there exists sequence $(z_n)_{n\in\mathbb{N}}$ such that z_n lies between x_n and y_n and

$$f'(z_n) = 0,$$

for every $n \in \mathbb{N}$. This means for every $U \ni 0$ and $\eta > 0$, one has $z_n \in U \cap [0 < f - f(0) < \eta]$ for n sufficiently large such that $\operatorname{dist}(0, \partial f(z_n)) = |f'(z_n)| = 0$. □

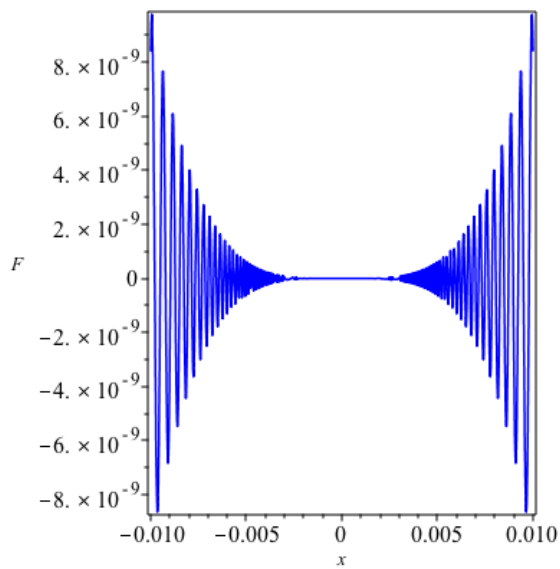

Figure 5.1: A Lipschitz smooth function that fails to be KL, see Proposition 5.4.

Proposition 5.4. *Let $f : \mathbb{R} \to \mathbb{R}$ be given by*

$$f(x) = \begin{cases} x^2 \sin(1/x), & x \neq 0; \\ 0, & x = 0. \end{cases}$$

Then $F(x) = \int_0^x f(t)dt$ is Lipschitz smooth and fails to satisfy the KL property at the critical point 0.

Proof. We claim that $f(x)$ is Lipschitz with modulus 3. Indeed, we have $f'(0) = \lim_{x\to 0} f(x)/x = 0$ and for nonzero x

$$|f'(x)| = |2x \cdot \sin(1/x) - \cos(1/x)| \leq |2x| \cdot |\sin(1/x)| + |\cos(1/x)| \leq 2 + 1 = 3.$$

Then the mean value theorem implies that for every $x, y \in \mathbb{R}$, there is some ξ between them, such that

$$|f(x) - f(y)| \leq |f'(\xi)||x - y| \leq 3|x - y|.$$

Note that F is well-defined as f is continuous. Taking the sequence $x_n = (-1)^n/[(2n+1)\pi]$, one has $x_n \to 0$ as $n \to \infty$ and $F'(x_n) = f(x_n) = 0$. Hence for every neighborhood $U \ni 0$ and $\eta > 0$, there exists n sufficiently large such that $x_n \in U \cap [0 < F(x) - F(0) < \eta]$ and $\operatorname{dist}(0, \partial F(x_n)) = 0$. □

Example 5.5. (**Opposite of KL function is not necessarily KL**) Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} -|\sin(\frac{1}{x})e^{-\frac{1}{x^2}}|, & x \neq 0; \\ 0, & x = 0. \end{cases}$$

Then f has the KL property at 0 whereas $-f$ does not.

Proof. Note that f has the KL property at 0 because $U \cap [0 < f - f(0) < \eta] = \emptyset$ for every $U \ni 0$ and $\eta > 0$. However, $-f$ fails to satisfy KL at 0 according to Example 5.3. □

5.2 The KL property on the real line

Our results so far emphasize on desingularizing functions of the form $\varphi(t) = c \cdot t^{1-\theta}$ for some $c > 0$ and $\theta \in [0, 1)$. The next example shows that φ can have other form.

Example 5.6. Let $f(x) = \sin x$. Then f has the KL property at $\bar{x} = 3\pi/2$ with respect to $U = (\pi, 2\pi)$, $\eta = 1$ and $\varphi(t) = \arcsin(t - 1) + \pi/2$.

Proof. For $x \in (\pi, 2\pi) \cap [0 < f - f(\bar{x}) < 1] = (\pi, 2\pi) \backslash \{3\pi/2\}$,

$$\varphi'(f(x) - f(\bar{x})) \cdot \operatorname{dist}(0, \partial f(x)) = \frac{|\cos x|}{\sqrt{1 - (\sin x + 1 - 1)^2}} = \frac{|\cos x|}{\sqrt{1 - \sin^2 x}} = 1,$$

as claimed. □

Remark 5.7. Note that by applying Theorem 4.20 we can show that $x \mapsto \sin(x)$ also admits desingularizing function $\varphi_1(t) = c \cdot t^{1/2}$ for some $c > 0$ at $\bar{x} = 3\pi/2$.

We will use frequently the following notions: For $f : \mathbb{R} \to \overline{\mathbb{R}}$, $\bar{x} \in \operatorname{dom} f$ and $(a, b) \subseteq \operatorname{dom} f$ with $-\infty \leq a < b \leq \infty$ and $\bar{x} \in (a, b)$, define $f_{a,\bar{x}} : (a - \bar{x}, 0] \to \mathbb{R}$ and $f_{b,\bar{x}} : [0, b - \bar{x}) \to \mathbb{R}$ by

$$f_{a,\bar{x}}(x) = f(x + \bar{x}) - f(\bar{x}), \forall x \in (a - \bar{x}, 0], \tag{5.1}$$

$$f_{b,\bar{x}}(x) = f(x + \bar{x}) - f(\bar{x}), \forall x \in [0, b - \bar{x}). \tag{5.2}$$

For $z \in \{a, b\}$, we adopt the convention $(f_{z,\bar{x}}^{-1})' \equiv 0$ when the inverse of $f_{z,\bar{x}}$ does not exist. Set $\eta_{a,b,\bar{x}} = \min\{f(a) - f(\bar{x}), f(b) - f(\bar{x})\}$. It is also convenient to set $f(x) = \infty$ if $x \in \{\infty, -\infty\}$. These conventions help us to write a closed-form formula for the desingularizing function $\varphi(t)$.

Proposition 5.8. *Let $f : \mathbb{R} \to \overline{\mathbb{R}}$ be convex and C^1 on the interval $(a, b) \subseteq \operatorname{int} \operatorname{dom} f$, where $-\infty \leq a < b \leq \infty$. Let $\bar{x}$ be a minimizer. Set $\eta = \eta_{a,b,\bar{x}}$ and define for $s \in (0, \eta)$*

$$h(s) = \max\{(-f_1^{-1})'(s)), (f_2^{-1})'(s)\},$$

where $f_1 = f_{a,\bar{x}}$ and $f_2 = f_{b,\bar{x}}$. Furthermore, define $\varphi : [0, \eta) \to \mathbb{R}_+$ by

$$\varphi(t) = \int_0^t h(s)ds, \forall t \in (0, \eta)$$

and $\varphi(0) = 0$. Then f has the KL property at $\bar{x}$ with respect to $U = (a, b)$, η and φ.

Proof. Note that there are four types of minimizer. Hence we need to consider the KL property at $\bar{x}$ in four cases:

Case 1: We begin with the case where $f(\bar{x}) < f(x)$ on $(a, b)\backslash\bar{x}$. The function f is assumed to be convex, then $f_1'(x)$ is increasing by using Fact 3.13. Note that $\bar{x}$ is the unique minimizer of f on (a, b). Hence we have $f_1'(x) < 0$ on $(a - \bar{x}, 0)$. Consequently, one concludes that f_1 is strictly decreasing and invertible. Next by applying the chain rule to the identity

$$f_1^{-1}(f(x) - f(\bar{x})) = f_1^{-1}(f_1(x - \bar{x})) = x - \bar{x}, \forall x \in (a, \bar{x}),$$

one gets

$$(f_1^{-1})'(f(x) - f(\bar{x})) \cdot f'(x) = (-f_1^{-1})'(f(x) - f(\bar{x})) \cdot |f'(x)| = 1, \tag{5.3}$$

where the left equality holds because $f'(x) < 0$ for $x \in (a, \bar{x})$. A similar argument shows that $f_2'(x) > 0$ on $(0, b - \bar{x})$ and hence f_2 is invertible. By using the chain rule we have

$$(f_2^{-1})'(f(x) - f(\bar{x})) \cdot f'(x) = (f_2^{-1})'(f(x) - f(\bar{x})) \cdot |f'(x)| = 1. \tag{5.4}$$

Now we work towards showing that $\varphi(t)$ is well-defined. By using the formula for the derivative of the inverse function, we have for $s \in (0, \eta)$

$$(-f_1^{-1})'(s) = \frac{1}{(-f_1')(f_1^{-1}(s))} > 0 \text{ and } (f_2^{-1})'(s) = \frac{1}{f_2'(f_2^{-1}(s))} > 0.$$

Observe that

$$\lim_{s\to 0^+}(-f_1')(f_1^{-1}(s)) = \lim_{s\to 0^+}(-f')(f_1^{-1}(s) + \bar{x}) = -f'(\bar{x}) = 0.$$

Therefore we have $\lim_{s\to 0^+}(-f_1^{-1})'(s) = \infty$, and consequently $\lim_{s\to 0^+} h(s) = \infty$. This means $\varphi(t)$ is indeed an improper integral. Applying Fact 2.2 yields

$$h(s) \leq (-f_1^{-1})'(s) + (f_2^{-1})'(s),$$

It then follows that for fixed $t \in (0, \eta)$, we have

$$\begin{aligned}
\varphi(t) &= \lim_{u\to 0^+}\int_u^t h(s)ds \leq \lim_{u\to 0^+}\int_u^t (-f_1^{-1})'(s) + (f_2^{-1})'(s)ds \\
&\leq \left[(-f_1^{-1})(t) + f_2^{-1}(t)\right] - \lim_{u\to 0^+}\left[(-f_1^{-1})(u) + f_2^{-1}(u)\right] \\
&= (-f_1^{-1})(t) + f_2^{-1}(t) < \infty,
\end{aligned}$$

meaning that $\varphi(t)$ is well-defined. Note that the last equality is implied by the right-continuity of f_i^{-1} at 0 and the fact that $f_i^{-1}(0) = 0$ for each i.

Next we show $\varphi \in \mathcal{K}_\eta$. Note that $-f_1'(s)$ is decreasing and f_1^{-1} is strictly decreasing. Hence we have that $(-f_1^{-1})'(s) = 1/(-f_1')(f_1^{-1}(s))$ is a decreasing function. A similar argument shows that $(f_2^{-1})(s)$ is also decreasing. On the other hand, $h(s)$ is continuous because f_1 and f_2 are C^1 on the interior of their respective domains. Therefore, one concludes $h(s)$ is a continuous decreasing function. Applying Fact 3.15 shows that $\varphi \in \mathcal{K}_\eta$.

Now we are ready to satisfy the KL property at $\bar{x}$. For every $x \in (a, b) \cap [0 < f - f(\bar{x}) < \eta] = (a, b)\backslash\{\bar{x}\}$,

$$\begin{aligned}
&\varphi'(f(x) - f(\bar{x})) \cdot \operatorname{dist}(0, \partial f(x)) = h\left(f(x) - f(\bar{x})\right) \cdot |f'(x)| \\
&= \max\{(-f_1^{-1})'(f(x) - f(\bar{x})), (f_2^{-1})'(f(x) - f(\bar{x}))\} \cdot |f'(x)| \\
&\geq 1,
\end{aligned}$$

where the last inequality is implied by (5.3) and (5.4).

Case 2: Now consider the case where $f(x) = f(\bar{x})$ for $x \in (a, c]$ and $f(x) > f(\bar{x})$ for $x \in (c, b)$ for some $c \in (a, b)$. Let us satisfy the KL property

at $\bar{x} = c$ first. Note that $f_1(x) = f(x + \bar{x}) - f(\bar{x}) = 0$ for $x \in (a - \bar{x}, 0]$. Hence by our convention $(f_1^{-1})' \equiv 0$, which means $h(s) = (f_2^{-1})'(s)$. For $x \in (a, b) \cap [0 < f - f(\bar{x}) < \eta] = (c, b)$,

$$\varphi'(f(x) - f(\bar{x})) \cdot \mathrm{dist}(0, \partial f(x)) = (f_2^{-1})'(f(x) - f(\bar{x})) \cdot |f'(x)| = 1,$$

where the last equality is implied by (5.4). As for $\bar{x} \in (a, c)$, by taking $U = (a, c)$, one has $U \cap [0 < f - f(\bar{x}) < \eta] = \emptyset$, which means there is nothing to prove.

Case 3: Consider the symmetry case where $f(x) > f(\bar{x})$ on (a, c) and $f(x) = f(\bar{x})$ on $[c, b)$ for some $c \in (a, b)$. Let $\bar{x} = c$. Then f_1 is invertible and similarly by our convention the function $h(s) = (-f_1^{-1})'(s)$. For $x \in (a, b) \cap [0 < f - f(\bar{x}) < \eta] = (a, c)$

$$\varphi'(f(x) - f(\bar{x})) \cdot \mathrm{dist}(0, \partial f(x)) = (-f_1^{-1})'(f(x) - f(\bar{x}))|f'(x)| = 1,$$

where the last equality follows from (5.3). If $\bar{x} \in (c, b)$, we have $U \cap [0 < f - f(\bar{x}) < \eta] = \emptyset$ by taking $U = (a, b)$.

Case 4: Finally we consider the case where $f(x) = f(\bar{x})$ on some subinterval $[c, d] \subseteq (a, b)$ and $f(\bar{x}) < f(x)$ on $(a, b) \backslash [c, d]$. Let $\bar{x} = c$. Then $f_2(x) = 0$ on $[0, d - c]$ and by our convention $h(s) = (-f_1^{-1})'(s)$. Hence by invoking (5.3) one has for $x \in (a, d) \cap [0 < f - f(\bar{x}) < \eta] = (a, c)$

$$\varphi'(f(x) - f(\bar{x})) \cdot \mathrm{dist}(0, \partial f(x)) = (-f - 1^{-1})'(f(x) - f(\bar{x}) \cdot |f'(x)| = 1.$$

Let $\bar{x} = d$. Then $h(s)$ becomes $(f_2^{-1})'(s)$ similarly. Invoking (5.4) yields that for $x \in [c, b) \cap [0 < f - f(\bar{x}) < \eta] = (d, b)$

$$\varphi'(f(x) - f(\bar{x})) \cdot \mathrm{dist}(0, \partial f(x)) = (f_2^{-1})'(f(x) - f(\bar{x})) \cdot |f'(x)| = 1.$$

As for the case where $\bar{x} \in (c, d)$, let $U = (c, d)$. Then $U \cap [0 < f - f(\bar{x}) < \eta] = \emptyset$ hence there is nothing to prove. □

Remark 5.9. (i) When proving the KL property at endpoints of $[c, d]$, where $[c, d]$ is the set of minimizers in case 4, we can construct a desingularizing function $\bar{\varphi}$ that applies to both c and d. Set $m = \inf\{f(x) : x \in (a, b)\}$. Let $f_3(x) = f(x + c) - m$ for $x \in (a - c, 0]$ and $f_4(x) = f(x + d) - m$ for $x \in [0, b - d)$. Let $\bar{\varphi}$ be defined by

$$\bar{\varphi}(y) = \int_0^y \max\{(-f_3^{-1})'(t), (f_4^{-1})'(t)\} dt,$$

for $y > 0$ and $\bar{\varphi}(0) = 0$. Then for $x \in (a, d) \cap [0 < f(x) - f(c) < f(a)] = (a, c)$,

$$\bar{\varphi}'(f(x)-f(\bar{x}))\cdot\mathrm{dist}(0, \partial f(x)) = \max\{(-f_3^{-1})'(f(x)), (f_4^{-1})'(f(x))\}\cdot|f'(x)| \geq 1,$$

which means f has the KL property at c with $\bar{\varphi}(t)$. Similarly, for $x \in (c, b) \cap [0 < f(x) - f(d) < f(b)] = (d, b)$,

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

which means f has the KL property at d with $\bar{\varphi}(t)$. However, we will learn from Chapter 6 that $\bar{\varphi}$ is not the optimal desingularizing function.

(ii) Our first attempt to construct $\varphi(t)$ is to take some $\eta > 0$ sufficiently small so that $h(s) = \max\{(-f_1^{-1})(s), f_2^{-1}(s)\}$ becomes either $(-f_1^{-1})(s)$ or $f_2^{-1}(s)$ on the interval $[0, \eta)$. This leads to a question of independent interests: Let f and g be two smooth strictly increasing convex functions defined on $[0, \infty)$ with $f(0) = g(0)$.

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

The answer is no. See the following proposition.

Proposition 5.10. *There exist strictly increasing convex C^2 functions $f, g : \mathbb{R}_+ \to \mathbb{R}_+$ with $f(0) = g(0) = 0$ and $\inf\{x > 0 : f(x) = g(x)\} = 0$. To be precise, let $h : \mathbb{R} \to \mathbb{R}$ be given by*

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

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 $g(x) - f(x) = h(x)$.

Proof. Note that $h(x) \in C^\infty$ by Proposition 2.10. We now show that h is representable as 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, the functions $g_1(x)$ and $f_1(x)$ are strictly

increasing C^1 functions with $f_1'(x) = h_-''(x)$ and $g_1'(x) = h_+''(x)$. Since $h''(x) = h_+''(x) - h_-''(x)$, one has

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

Applying again the fundamental theorem of calculus, one has f and g are strictly increasing C^2 functions with $f'(x) = f_1'(x)$ and $g'(x) = g_1'(x)$. Fact 3.3 implies that f and g are both convex. Furthermore, we have

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

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

Example 5.11. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} \exp(-x^{-2}), & x \neq 0; \\ 0, & x = 0. \end{cases}$$

Then f has the KL property at $\bar{x} = 0$ with respect to $U = (-\sqrt{2/3}, \sqrt{2/3})$, $\eta = \exp(-3/2)$ and $\varphi(t)$ defined by

$$\varphi(t) = \begin{cases} \sqrt{-1/\ln t}, & 0 < t < \eta; \\ 0, & t = 0. \end{cases}$$

Proof. We begin with showing that f is locally convex around 0. Elementary calculus gives

$$f''(x) = \left(\frac{4}{x^6} - \frac{6}{x^4}\right) e^{-\frac{1}{x^2}} \geq 0 \Leftrightarrow \frac{4}{x^6} - \frac{6}{x^4} \geq 0 \Leftrightarrow |x| \leq \sqrt{\frac{2}{3}}.$$

Therefore by applying Fact 3.3, we conclude that f is convex on $[-\sqrt{2/3}, \sqrt{2/3}]$. Define $f_1(x) = f(x)$ for $x \in (-\sqrt{2/3}, 0]$ and $f_2(x) = f(x)$ for $x \in [0, \sqrt{2/3})$. Then we have $h(s) = \max\{(-f_1^{-1})'(s), (f_2^{-1})'(s)\} = (f_2^{-1})'(s)$ because f is even. For $0 < y \leq \exp(-3/2)$, $y = \exp(-x^{-2}) \Leftrightarrow |x| = \sqrt{-1/\ln y}$. Hence $f_2^{-1}(y) = \sqrt{-1/\ln y}$.

By applying Proposition 5.8, we have

$$\varphi(t) = \int_0^t (f_2^{-1})'(s)ds = f_2^{-1}(t) = \sqrt{-\frac{1}{\ln(t)}}, \forall t > 0,$$

which completes the proof. □

Theorem 5.12. *Every convex C^1 function $f : \mathbb{R} \to \mathbb{R}$ is KL. In particular, if* $\operatorname{argmin} f \neq \emptyset$ *then f has the KL property at $\bar{x} \in \operatorname{argmin} f$ with respect to $U = \mathbb{R}$, $\eta = \infty$ and the function $\varphi(t)$ given in Proposition 5.8.*

Proof. By Fact 4.4 it suffices to satisfy the KL property on $\operatorname{argmin} f$. For $\bar{x} \in \operatorname{argmin} f$, applying Proposition 5.8 shows that f has the KL property at $\bar{x}$ with respect to $U = \mathbb{R}$, $\eta = \infty$ and $\varphi(t)$. □

Remark 5.13. When $\operatorname{dom} f = (a, b) \subset \mathbb{R}$ for $-\infty < a < b < \infty$, a careful examination of the proof in Proposition 5.8 shows that f has the KL property at $\bar{x} \in \operatorname{argmin} f$ with respect to $U = (a, b)$, $\eta = \infty$ and $\varphi(t)$.

Example 5.14. The following assertions hold:

(i) The function $f(x) = -\ln(1 - x^2)$ has the KL property at $\bar{x} = 0$ with respect to $U_1 = (-1, 1)$, $\eta_1 = \infty$ and $\varphi_1(t) = \sqrt{1 - e^{-t}}$.

(ii) The function $g(x) = \tan(x^2)$ has the KL property at $\bar{x} = 0$ with respect to $U_2 = (-\sqrt{\pi/2}, \sqrt{\pi/2})$, $\eta_2 = \infty$ and $\varphi_2(t) = \sqrt{\arctan t}$.

Proof. (i) Define $f_1(x) = f(x)$ for $x \leq 0$ and $f_2(x) = f(x)$ for $x \geq 0$. Note that f is even and $s = f(x) \Leftrightarrow |x| = \sqrt{1 - \exp(-s)}$ for $s \in (0, \eta_1)$. Hence we have $(-f_1^{-1})(s) = f_2^{-1}(s) = \sqrt{1 - \exp(-s)}$ and

$$\varphi_1(t) = \int_0^t \max\{(-f_1^{-1})'(s), (f_2^{-1})'(s)\} ds = \sqrt{1 - \exp(-t)}.$$

By Theorem 5.12 we conclude that f has the KL property at 0 with respect to $U_1 = (-1, 1)$, $\eta_1 = \infty$ and $\varphi_1(t)$.

(ii) Define $g_1(x) = g(x)$ for $x \leq 0$ and $g_2(x) = g(x)$ for $x \geq 0$. Because g is even and $s = g(x) \Leftrightarrow |x| = \sqrt{\arctan(s)}$ for $s \in (0, \eta_2)$, one concludes that $(-g_1^{-1})(s) = g_2^{-1}(s) = \sqrt{\arctan(s)}$ and

$$\varphi_2(t) = \int_0^t \max\{(-g_1^{-1})'(s), (g_2^{-1})'(s)\} ds = \int_0^t (-g_1^{-1})'(s) ds = \sqrt{\arctan(t)}.$$

Then invoking Theorem 5.12 yields that g has the KL property at 0 with respect to $U = (-\sqrt{\pi/2}, \sqrt{\pi/2})$, $\eta_2 = \infty$ and $\varphi_2(t) = \sqrt{\arctan(t)}$. □

Recall that for a polynomial $f : \mathbb{R} \to \overline{\mathbb{R}}$ with degree $n \geq 3$, there are four cases for a stationary point $\bar{x}$: For some $\varepsilon > 0$

	$(\bar{x}-\varepsilon,\bar{x})$	$(\bar{x},\bar{x}+\varepsilon)$
Local minimizer	$f'(x)<0$	$f'(x)>0$
Local maximizer	$f'(x)>0$	$f'(x)<0$
Inflection point	$f'(x)>0$	$f'(x)>0$
Inflection point	$f'(x)<0$	$f'(x)<0$

Table 5.1: Four types of stationary points for polynomials with degree $n \geq 3$.

The following lemma shows that polynomials are locally convex/concave around stationary points.

Lemma 5.15. *Let $f : \mathbb{R} \to \mathbb{R}$ be a polynomial and let $\bar{x}$ be a stationary point of f. Then the following statements hold:*

(i) If $\bar{x}$ is a local minimizer then there exists $(a,b) \ni \bar{x}$ such that f is convex on (a,b).

(ii) If $\bar{x}$ is a local maximizer, then there exists $(a,b) \ni \bar{x}$ such that f is concave on (a,b).

(iii) If $\bar{x}$ is an inflection point then there exists $(a,b) \ni \bar{x}$ such that f is either convex on $(a,\bar{x})$ or convex on $(\bar{x},b)$.

Proof. The statement is trivially true for polynomials with degree $n \leq 2$. Suppose that f has degree $n \geq 3$ and $\bar{x}$ is a stationary point. We will proceed the proof by considering four cases, see Table 5.1.

(i) The point $\bar{x}$ is a local minimizer means that there exists $\varepsilon_1 > 0$ such that $f'(x) < 0$ on $(\bar{x}-\varepsilon_1,\bar{x})$ and $f'(x) > 0$ on $(\bar{x},\bar{x}+\varepsilon_1)$. Note that $f''(\bar{x}) \geq 0$. If not, we would have $f'(x) < 0$ for every x sufficiently close to $\bar{x}$ with $x > \bar{x}$, which contradicts to the assumption that $\bar{x}$ is a local minimizer.

Next we show that there exists $0 < \varepsilon < \varepsilon_1$ such that $f''(x) > 0$ on $(\bar{x}-\varepsilon,\bar{x}+\varepsilon)$. Suppose the contrary was true. Then for every $0 < \varepsilon < \varepsilon_1$ there is some $y_\varepsilon \in (\bar{x}-\varepsilon,\bar{x}+\varepsilon)$ such that $f''(y_\varepsilon) \leq 0$. Hence by intermediate value theorem there exists z_ε between $\bar{x}$ and y_ε such that $f''(z_\varepsilon) = 0$, which is absurd because the polynomial f'' has at most $n-2$ roots.

Hence we conclude that $f''(x) > 0$ on $(\bar{x}-\varepsilon,\bar{x}+\varepsilon)$ for $\varepsilon > 0$. Set $a = \bar{x}-\varepsilon$ and $b = \bar{x}+\varepsilon$. Invoking Fact 3.3 completes the proof.

(ii) Note that $\bar{x}$ is a local minimizer of $-f$ if it is a local maximizer of f. Then applying (i) to $-f$ completes the proof.

(iii) We need to consider two cases when $\bar{x}$ is an inflection point. Let us begin with the one where $f'(x)$ is positive on $(\bar{x}-\varepsilon,\bar{x}+\varepsilon)$ for some $\varepsilon > 0$ sufficiently small.

Shrinking $\varepsilon > 0$ if necessary, we assume that $f''(x)$ has no root on $(\bar{x}, \bar{x}+\varepsilon)$. This implies that $f''(x)$ is either positive or negative on $(\bar{x}, b)$. If $f''(x)$ was negative, we would have $f'(x) < 0$ on $(\bar{x}, \bar{x}+\varepsilon)$, which contradicts to our assumption. Hence we conclude that $f(x)$ is convex on $(\bar{x}, b)$.

Now suppose that $f'(x)$ is negative on $(\bar{x}-\varepsilon, \bar{x}+\varepsilon)$. The similar argument as above shows that $f''(x) > 0$ on $(\bar{x}-\varepsilon, \bar{x})$, which means f is convex on$(\bar{x}-\varepsilon, \bar{x})$. Setting $a = \bar{x}-\varepsilon$ and $b = \bar{x}+\varepsilon$ completes the proof. □

Recall from Fact 4.9 and Example 4.11 that polynomials on $\mathbb{R}^n$ are KL. With Proposition 5.8 at hand, we are now ready to revisit the one-dimensional version of this well-known result.

Theorem 5.16. *Polynomials on the real line are KL. In particular, for stationary point $\bar{x}$ and some interval $(a, b) \ni \bar{x}$, where $-\infty \leq a < b \leq \infty$, the following assertions hold:*

(i) If $\bar{x}$ is a local minimizer then f has the KL property at $\bar{x}$ with respect to $U = (a, b)$, $\eta_1 = \eta_{a,b,\bar{x}}$ and $\varphi_1(t)$ given by

$$\varphi_1(t) = \int_0^t h(s)ds, \forall t \in (0, \eta)$$

and $\varphi_1(t) = 0$, where $h(s) = \max\{(-f_1^{-1})'(s), (f_2^{-1})'(s)\}$, $f_1 = f_{a,\bar{x}}$ and $f_2 = f_{b,\bar{x}}$.

(ii) If $\bar{x}$ is an inflection point such that $f(x) < f(\bar{x})$ on $(a, \bar{x})$ and $f(x) > f(\bar{x})$ on $(\bar{x}, b)$, then f has the KL property at $\bar{x}$ with respect to $U = (a, b)$, $\eta_2 = f(b) - f(\bar{x})$ and $\varphi_2(t) = f_{b,\bar{x}}^{-1}(t)$.

(iii) If $\bar{x}$ is an inflection point such that $f(x) > f(\bar{x})$ on $(a, \bar{x})$ and $f(x) < f(\bar{x})$ on $(\bar{x}, b)$, then f has the KL property at $\bar{x}$ with respect to $U = (a, b)$, $\eta_3 = f(a) - f(\bar{x})$ and $\varphi_3(t) = -f_{a,\bar{x}}^{-1}(t)$.

(iv) If $\bar{x}$ is a local maximizer, then f has the KL property at $\bar{x}$ with respect to $U = (a, b)$, every $\eta_4 > 0$ and $\varphi_4 \in \mathcal{K}_{\eta_4}$.

Proof. (i) According to Lemma 5.15, there exists a interval $(a, b) \ni \bar{x}$ on which f is convex. Then invoking Proposition 5.8 completes the proof.

(ii) By Lemma 5.15, there exists $b > \bar{x}$ such that f is convex on $(\bar{x}, b)$. Then it is easy to see that $\varphi_2 \in \mathcal{K}_{\eta_2}$. Note that for $x \in (\bar{x}, b)$

$$f_{b,\bar{x}}^{-1}(f(x) - f(\bar{x})) = x - \bar{x} \Rightarrow (f_{b,\bar{x}}^{-1})'(f(x) - f(\bar{x})) \cdot f'(x) = 1.$$

Hence for $x \in (a,b) \cap [0 < f - f(\bar{x}) < \eta_2] = (\bar{x}, b)$,

$$\varphi_2'(f(x) - f(\bar{x})) \cdot \operatorname{dist}(0, \partial f(x)) = (f_{b,\bar{x}}^{-1})'(f(x) - f(\bar{x})) \cdot f'(x) = 1.$$

(iii) Apply an argument similar to (ii).

(iv) For every $\eta_4 > 0$, one has $(a,b) \cap [0 < f - f(\bar{x}) < \eta_4] = \emptyset$. Hence there is nothing to prove. □

Remark 5.17. Compared to the well-known semialgebraic argument, our proof is new and provides an explicit formula for the desingularizing function φ. Our result is also consistent with [10, Corollary 9], where Bolte et al. discovered that piecewise polynomial convex functions on $\mathbb{R}^n$ with degree m has KL exponent $\theta = 1 - 1/(1 + (m-1)^n)$ and $\varphi(t) = c \cdot t^{1-\theta}$ for some $c > 0$. For instance let $f(x) = x^4$. Then [10, Corollary 9] implies f admits KL exponent $\theta = 3/4$ and $\varphi(t) = t^{1/4}$ at 0, whereas our result shows the same. However, ours applies to even nonconvex polynomials, for example $x \mapsto x^3$.

5.3 A sufficient condition for the KL property on $\mathbb{R}^n$

Proposition 5.18. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, lsc and convex, and let $\bar{x} \in \operatorname{argmin} f$. If there exists $\varepsilon > 0$ such that $\varepsilon \cdot \mathbb{B}_{\mathbb{R}^n} \subseteq \partial f(\bar{x})$, then f has the KL property at $\bar{x}$ with respect to $U = \mathbb{R}^n$, $\eta = \infty$ and $\varphi(t) = \varepsilon^{-1} t$.*

Proof. The function f is assumed to be convex, then by Fact 3.17 for every $u \in \mathbb{R}^n$ with $\|u\| = 1$,

$$f'(\bar{x}; u) = \max\{\langle u, g\rangle : g \in \partial f(\bar{x})\} \geq \varepsilon \langle u, u\rangle = \varepsilon. \tag{5.5}$$

Invoking Fact 3.11 and (5.5) gives

$$\begin{aligned} f(x) &\geq f(\bar{x}) + f'(\bar{x}; x - \bar{x}) = f(\bar{x}) + \|x - \bar{x}\| \cdot f'\left(\bar{x}; \frac{x - \bar{x}}{\|x - \bar{x}\|}\right) \\ &\geq f(\bar{x}) + \varepsilon \|x - \bar{x}\|. \end{aligned}$$

Equivalently for every $x \in \mathbb{R}^n$ with $x \neq \bar{x}$,

$$f(x) - f(\bar{x}) \geq \varepsilon \cdot \|x - \bar{x}\|.$$

This further implies that for $g \in \partial f(x)$, we have

$$\varepsilon \leq \frac{f(x) - f(\bar{x})}{\|x - \bar{x}\|} \leq \frac{\langle g, x - \bar{x}\rangle}{\|x - \bar{x}\|} \leq \|g\|.$$

Hence by taking the infimum over $g \in \partial f(x)$ one concludes that for every $x \neq \bar{x}$

$$\operatorname{dist}(0, \partial f(x)) \geq \varepsilon.$$

This means for every $x \in \mathbb{R}^n \cap [0 < f(x) - f(\bar{x}) < \eta] \subseteq \mathbb{R}^n \backslash \{\bar{x}\}$,

$$\varepsilon^{-1} \cdot \operatorname{dist}(0, \partial f(x)) \geq 1,$$

which completes the proof. □

Remark 5.19. One can also proceed in the proof with an alternative approach involving Fact 4.13. Note that taking the infimum over $\bar{x} \in \operatorname{argmin} f$ of the inequality $f(x) - f(\bar{x}) \geq \varepsilon \|x - \bar{x}\|$ yields

$$f(x) - \min f \geq \varepsilon \operatorname{dist}(x, \operatorname{argmin} f).$$

Then applying Fact 4.13 completes the proof.

Corollary 5.20. *Every norm $f(x) = \|x\|$ has the KL property at 0 with respect to $U = \mathbb{R}^n$, $\eta > 0$ and $\varphi(t) = c^{-1} \cdot t$ for some $c > 0$.*

Proof. By Fact 2.8 and Example 3.23 there exists some $c > 0$ such that $c \cdot \mathbb{B}_{\mathbb{R}^n} \subseteq \partial f(0) = \mathbb{B}_{\|\cdot\|_*}[0; 1]$, where $\mathbb{B}_{\mathbb{R}^n}$ denotes the Euclidean unit ball and $\mathbb{B}_{\|\cdot\|_*}[0; 1]$ is the dual unit ball of norm $\|\cdot\|$. Then applying Proposition 5.18 completes the proof. □

Chapter 6

The generalized KL property

In this chapter, we introduce *the generalized KL property*, a new concept that generalizes the classic KL property by employing nonsmooth desingularizing functions, and provide several calculus rules for this generalized notion. Our results are new and extend the classic KL property.

Let $\eta \in (0, \infty]$. 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 concave and strictly increasing on $[0, \eta)$.

The left derivative of φ at $t \in (0, \eta)$ is given by

$$\varphi'_-(t) = \lim_{s \to t^-} \frac{\varphi(s) - \varphi(t)}{s - t} = \lim_{\alpha \to 0^+} \frac{\varphi(t) - \varphi(t - \alpha)}{\alpha}. \tag{6.1}$$

6.1 Definition and examples

Let us begin with a lemma concerning the properties of $\varphi \in \Phi_\eta$.

Lemma 6.1. *Let $\eta \in (0, \infty]$ and let $\varphi \in \Phi_\eta$. Then the following properties hold:*

(i) The function $t \mapsto \varphi'_-(t)$ is decreasing and $\varphi'_-(t) > 0$ for $t \in (0, \eta)$. If in addition $\varphi(t)$ is strictly concave, then $t \mapsto \varphi'_-(t)$ is strictly decreasing.

(ii) For $0 \le s < t < \eta$, $\varphi'_-(t) \le \frac{\varphi(t) - \varphi(s)}{t - s}$.

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

Proof. (i) Applying Fact 3.13 to the convex function $-\varphi$ yields that $(-\varphi)'_-$ is increasing, which means $\varphi'_- = -(-\varphi)'_-$ is decreasing. Fix $t \in (0, \eta)$. Then for every $0 < s < t$,

$$\frac{\varphi(s) - \varphi(t)}{s - t} > 0 \Rightarrow \varphi'_-(t) = \lim_{s \to t^-} \frac{\varphi(s) - \varphi(t)}{s - t} \ge 0.$$

Suppose that $\varphi'_-(t_0) = 0$ for some $t_0 \in (0, \eta)$. Then by the monotonicity of $\varphi'_-(t)$, we would have by Fact 3.13 that

$$\varphi(t) - \varphi(t_0) = \int_{t_0}^{t} \varphi'_-(s)ds \leq (t - t_0)\varphi'_-(t_0) = 0,$$

which contradicts to the assumption that φ is strictly increasing. Hence we conclude that $\varphi'_-(t) > 0$ on $(0, \eta)$.

Assume in addition that φ is strictly concave and there exists an interval $[t_0, t_1]$ on which φ'_- is a constant $c > 0$. Set $t_\lambda = (1-\lambda)t_0 + \lambda t_1$ for $\lambda \in (0, 1)$. Then we have

$$\varphi(t_\lambda) - \varphi(t_0) = \int_{t_0}^{t_\lambda} \varphi'_-(s)ds = \lambda(t_1 - t_0)c,$$

$$\varphi(t_1) - \varphi(t_\lambda) = \int_{t_\lambda}^{t_1} \varphi'_-(s)ds = (1 - \lambda)(t_1 - t_0)c,$$

which implies that

$$\varphi(t_\lambda) - (1 - \lambda)\varphi(t_0) - \lambda\varphi(t_1) = (1 - \lambda)\left[\varphi(t_\lambda) - \varphi(t_0)\right] - \lambda\left[\varphi(t_1) - \varphi(t_\lambda)\right] = 0,$$

contradicting to the strictly concavity. Hence we conclude that φ'_- is strictly decreasing.

(ii) By using (6.1) and Fact 3.11 one has

$$\varphi'_-(t) = \lim_{\alpha \to 0^+} \frac{-\varphi(t - \alpha) - (-\varphi(t))}{\alpha} = \inf_{\alpha > 0} \frac{-\varphi(t - \alpha) - (-\varphi(t))}{\alpha} \leq \frac{-\varphi(s) + \varphi(t)}{t - s},$$

where the second equality holds because $-\varphi$ is convex and the last inequality follows from choosing $\alpha = t - s > 0$.

(iii) Take $u \in (0, t)$. Then invoking Fact 3.13 yields $\varphi(t) - \varphi(u) = \int_u^t \varphi'_-(s)ds$. Recall that φ is right-continuous at 0 with $\varphi(0) = 0$. Hence we have

$$\varphi(t) = \lim_{u \to 0^+} \left[\varphi(t) - \varphi(u)\right] = \lim_{u \to 0^+} \int_u^t \varphi'_-(s)ds < \infty.$$

Let $(u_n)_{n \in \mathbb{N}}$ be a decreasing sequence such that $u_1 < t$ and $u_n \to 0+$. Then $\varphi(t) = \lim_{n \to \infty} \int_{u_n}^t \varphi'_-(s)ds$. Define for each $n \in \mathbb{N}$ the function $f_n : (0, t] \to \mathbb{R}$ by

$$f_n(s) = \begin{cases} \varphi'_-(s), & s \in (u_n, t]; \\ 0, & s \in (0, u_n). \end{cases}$$

Recall that $\varphi'_-(s) > 0$. Then the sequence $(f_n)_{n\in\mathbb{N}}$ satisfies: (a)$f_n \le f_{n+1}$ for every $n \in \mathbb{N}$; (b)$f_n(s) \to \varphi'_-(s)$ pointwise for $s \in (0,t)$; (c)The integral $\int_0^t f_n(s)ds = \int_{u_n}^t \varphi'_-(s)ds = \varphi(t) - \varphi(u_n) < \infty$ for every $n \in \mathbb{N}$. By applying the monotone convergence theorem, we conclude that

$$\begin{aligned}\varphi(t) &= \lim_{n\to\infty} \int_{u_n}^t \varphi'_-(s)ds = \lim_{n\to\infty} \int_0^t f_n(s)ds \\ &= \int_0^t \lim_{n\to\infty} f_n(s)ds = \int_0^t \varphi'_-(s)ds,\end{aligned}$$

which completes the proof. □

Definition 6.2. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc. Let $\bar{x} \in \operatorname{dom} \partial f$.

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

$$\varphi'_-(f(x) - f(\bar{x})) \cdot \operatorname{dist}(0, \partial f(x)) \ge 1. \tag{6.2}$$

(ii) We say that f is a *generalized KL function*, if f has the generalized KL property at every $x \in \operatorname{dom} \partial f$.

Remark 6.3. (i) In contrast to Definition 4.1, the generalized KL property allows us to take nonsmooth desingularizing function. The benefits of such extension can be found in Example 7.6.

(ii) Note that the KL property implies the generalized KL property because $\mathcal{K}_\eta \subseteq \Phi_\eta$.

Example 6.4. The following statements hold:

(i) **(A convex example)** 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 f has the generalized KL property at 0 with $U = \mathbb{R}$, $\eta = \infty$ and $\varphi_1(t)$ defined by

$$\varphi_1(t) = \begin{cases} \sqrt{2t}, & 0 \le t < \rho^2/2; \\ t/(2\rho) + 3\rho/4, & t \ge \rho^2/2. \end{cases}$$

(ii)**(A nonconvex example)** Consider the following function

$$g(x) = \begin{cases} 1 - e^{-|x|}, & |x| \le 1; \\ (1 - e^{-1})|x|, & |x| > 1. \end{cases}$$

Then g has the generalized KL property at 0 with $U = \mathbb{R}$, $\eta = \infty$ and φ_2 given by

$$\varphi_2(t) = \begin{cases} e \cdot t, & t < 1 - e^{-1}; \\ \frac{e}{e-1} \cdot t + e - 2, & t \geq 1 - e^{-1}. \end{cases}$$

Proof. (i) For $x \in \mathbb{R}\backslash\{0\}$, one has

$$\partial f(x) = \begin{cases} x, & 0 < |x| < \rho; \\ \operatorname{sgn}(x) \cdot [\rho, 2\rho], & x = |\rho|; \\ \operatorname{sgn}(x) \cdot 2\rho, & |x| > \rho \end{cases}$$

and hence

$$\operatorname{dist}(0, \partial f(x)) = \begin{cases} |x|, & 0 < |x| \leq \rho; \\ 2\rho, & |x| > \rho. \end{cases}$$

On the other hand, we have for $t > 0$

$$(\varphi_1)'_-(t) = \begin{cases} 1/\sqrt{2t}, & t \leq \rho^2/2; \\ 1/2\rho, & t > \rho^2/2. \end{cases}$$

Then for $x \in \mathbb{R} \cap [0 < f \leq \rho^2/2] = [-\rho, \rho]\backslash\{0\}$,

$$(\varphi_1)'_-(f(x)) \cdot \operatorname{dist}(0, \partial f(x)) = \frac{|x|}{\sqrt{|x|^2}} = 1,$$

and for $x \in \mathbb{R} \cap [\rho^2/2 < f < \infty] = \mathbb{R}\backslash[-\rho, \rho]$,

$$(\varphi_1)'_-(f(x)) \cdot \operatorname{dist}(0, \partial f(x)) = \frac{2\rho}{2\rho} = 1,$$

which completes the proof.

(ii) Let us begin with computing $\partial g(1)$. For $v \in \mathbb{R}$, $v \in \hat{\partial} g(1)$ is equivalent to

$$\liminf_{x \to 1+} \frac{(1 - e^{-1})x - (1 - e^{-1}) - v(x - 1)}{x - 1} = 1 - e^{-1} - v \geq 0$$
$$\text{and } \liminf_{x \to 1-} \frac{1 - e^{-x} - (1 - e^{-1}) - v(x - 1)}{1 - x} = -e^{-1} + v \geq 0.$$

Hence $\hat{\partial} g(1) = [e^{-1}, 1 - e^{-1}]$. Note that $\hat{\partial} g(x) = g'(x)$ on $\mathbb{R}\backslash\{0, \pm 1\}$ and g is continuous. Hence we have

$$\lim_{x \xrightarrow{g} 1^+} \hat{\partial} g(x) = \lim_{x \to 1^+} g'(x) = 1 - e^{-1}, \lim_{x \xrightarrow{g} 1^-} \hat{\partial} g(x) = \lim_{x \to 1^-} g'(x) = e^{-1}.$$

It then follows that $\partial g(1) = [e^{-1}, 1 - e^{-1}]$. A similar argument shows that $\partial g(-1) = [e^{-1} - 1, -e^{-1}]$.

By using Fact 3.18 we conclude that for nonzero x,

$$\partial g(x) = \begin{cases} \operatorname{sgn}(x) \cdot (1 - e^{-1}), & |x| > 1; \\ \operatorname{sgn}(x) \cdot [e^{-1}, 1 - e^{-1}], & |x| = 1; \\ \operatorname{sgn}(x) \cdot e^{-|x|}, & 0 < |x| < 1, \end{cases}$$

which means

$$\operatorname{dist}(0, \partial g(x)) = \begin{cases} 1 - e^{-1}, & |x| > 1; \\ e^{-|x|}, & 0 < |x| \leq 1. \end{cases}$$

On the other hand, one has

$$(\varphi_2)'_-(t) = \begin{cases} e, & t \leq 1 - e^{-1}; \\ \frac{e}{e-1}, & t > 1 - e^{-1}. \end{cases}$$

Hence for every $x \in \mathbb{R} \cap [0 < g \leq 1 - e^{-1}] = [-1, 1]\backslash\{0\}$,

$$(\varphi_2)'_-(g(x)) \cdot \operatorname{dist}(0, \partial g(x)) = e \cdot e^{-|x|} \geq e \cdot e^{-1} = 1,$$

and for $x \in \mathbb{R} \cap [1 - e^{-1} < g < \infty] = \mathbb{R}\backslash[-1, 1]$

$$(\varphi_2)'_-(g(x)) \cdot \operatorname{dist}(0, \partial g(x)) = \frac{e}{e-1} \cdot \frac{e-1}{e} = 1,$$

which completes the proof. □

6.2 Calculus rules

Theorem 6.5. *Suppose that $f(x) = \min_{1 \leq i \leq m} f_i(x)$ is continuous on $\operatorname{dom} \partial f$, where $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper and lsc function for each $i \in \{1, \ldots, m\}$. Let $\bar{x} \in \operatorname{dom} \partial f$ and suppose that $\bar{x} \in \bigcap_{i \in I(\bar{x})} \operatorname{dom} \partial f_i$, where $I(\bar{x}) = \{i : f_i(\bar{x}) =$*

$f(\bar{x})\}$. Assume further that for every $i \in I(\bar{x})$, the function f_i has the generalized KL property at $\bar{x}$ with respect to $U_i = \mathbb{B}(\bar{x}; \varepsilon_i)$ for some $\varepsilon_i > 0$, $\eta_i > 0$ and $\varphi_i \in \Phi_{\eta_i}$. Set $\eta = \min_{1 \le i \le m} \eta_i$ and define $\varphi : [0, \eta) \to \overline{\mathbb{R}}$ by

$$\varphi(t) = \int_0^t \max_{i \in I(\bar{x})} (\varphi_i)'_-(s) ds.$$

Then there exists $\varepsilon > 0$ such that f has the generalized KL property at $\bar{x}$ with respect to $U = \mathbb{B}(\bar{x}; \varepsilon)$, η and φ.

Proof. Let us show first that $\varphi(t)$ is well-defined and belongs to Φ_η. By assumption, the function $\varphi_i(t)$ is finite and right-continuous at 0 for each i. Moreover, we have $\varphi_i(t) = \int_0^t (\varphi_i)'_-(s) ds$ by using Fact 3.13. Then Lemma 3.14 guarantees that $\varphi(t)$ is well-defined. Furthermore, invoking Fact 3.15 yields that the function $\varphi \in \Phi_\eta$ and for $t \in (0, \eta)$

$$\varphi'_-(t) \ge \max_{i \in I(\bar{x})} \{(\varphi_i)'_-(t)\}, \forall i \in I(\bar{x}). \tag{6.3}$$

Now we work towards the existence of some $\varepsilon_0 > 0$ such that

$$I(x) \subseteq I(\bar{x}), \tag{6.4}$$

whenever $x \in \mathbb{B}(\bar{x}; \varepsilon_0)$. We claim that there exists $\varepsilon_0 > 0$ such that

$$\min_{i \notin I(\bar{x})} f_i(x) > f(x), \forall x \in \mathbb{B}(\bar{x}; \varepsilon_0),$$

from which (6.4) readily follows. To see (6.4), taking $i_0 \in I(x)$ and supposing that $i_0 \notin I(\bar{x})$, our claim implies that $f_{i_0}(x) \ge \min_{i \notin I(\bar{x})} f_i(x) > f(x) = f_{i_0}(x)$ whenever $x \in \mathbb{B}(\bar{x}; \varepsilon_0)$, which is absurd. Next we satisfy the aforementioned claim. Suppose to the contrary that for every $\varepsilon > 0$ there exists some $x \in \mathbb{B}(\bar{x}; \varepsilon)$ such that $\min_{i \notin I(\bar{x})} f_i(x) \le f(x)$. Then there exists a sequence $x_n \to \bar{x}$ with

$$\min_{i \notin I(\bar{x})} f_i(x_n) \le f(x_n), \forall n \in \mathbb{N}.$$

Taking $n \to \infty$, one has by the lower semi-continuity of $x \mapsto \min_{i \notin I(\bar{x})} f_i(x)$ and the continuity of $f(x)$ that

$$\min_{i \notin I(\bar{x})} f_i(\bar{x}) \le \liminf_{n \to \infty} \left(\min_{i \notin I(\bar{x})} f_i(x_n) \right) \le \liminf_{n \to \infty} f(x_n) = f(\bar{x}),$$

which is absurd because $\min_{i \notin I(\bar{x})} f_i(\bar{x}) > f(\bar{x})$.

On the other hand, by assumption one has for every $i \in I(\bar{x})$ and $x \in \mathbb{B}(\bar{x};\varepsilon_i) \cap [0 < f_i - f_i(\bar{x}) < \eta_i]$

$$\operatorname{dist}(0, \partial f_i(x)) \geq \frac{1}{(\varphi_i)'_-(f_i(x) - f_i(\bar{x}))}. \tag{6.5}$$

Moreover, invoking Fact 3.20, we have for every $x \in \operatorname{dom} \partial f$

$$\partial f(x) \subseteq \bigcup_{i \in I(x)} \partial f_i(x).$$

Take $u \in \partial f(x)$ with $\|u\| = \operatorname{dist}(0, \partial f(x))$. Then the inclusion above implies that $u \in \partial f_{i_0}(x)$ for some $i_0 \in I(x)$, and consequently

$$\operatorname{dist}(0, \partial f(x)) = \|u\| \geq \operatorname{dist}(0, \partial f_{i_0}(x)) \geq \min_{i \in I(x)} \operatorname{dist}(0, \partial f_i(x)). \tag{6.6}$$

Set $\varepsilon = \min_{0 \leq i \leq m} \varepsilon_i$ and $\eta = \min_{1 \leq i \leq m} \eta_i$. Take $x \in \mathbb{B}(\bar{x};\varepsilon) \cap [0 < f - f(\bar{x}) < \eta]$. Then we have $x \in \mathbb{B}(\bar{x};\varepsilon_i)$ and $0 < f(x) - f(\bar{x}) = f_i(x) - f_i(\bar{x}) < \eta_i$ for every $i \in I(x) \subseteq I(\bar{x})$.

Now we work towards showing the generalized KL property. For $x \in \mathbb{B}(\bar{x};\varepsilon) \cap [0 < f - f(\bar{x}) < \eta]$, we have

$$\begin{aligned}
\operatorname{dist}(0, \partial f(x)) &\geq \min_{i \in I(x)} \operatorname{dist}(0, \partial f_i(x)) \geq \min_{i \in I(x)} \frac{1}{(\varphi_i)'_-(f(x) - f(\bar{x}))} \\
&\geq \min_{i \in I(\bar{x})} \frac{1}{(\varphi_i)'_-(f(x) - f(\bar{x}))} = \frac{1}{\max_{i \in I(\bar{x})} (\varphi_i)'_-(f(x) - f(\bar{x}))} \\
&\geq \frac{1}{\varphi'_-(f(x) - f(\bar{x}))},
\end{aligned}$$

where the first inequality is implied by (6.6); the second one holds because of (6.5) and the fact that $x \in \mathbb{B}(\bar{x};\varepsilon_i) \cap [0 < f_i - f_i(\bar{x}) < \eta_i]$ for $i \in I(x) \subseteq I(\bar{x})$; the third one is implied by (6.4) and the last inequality holds because of (6.3). □

Remark 6.6. (i) Set $\eta_1 = \min\{1, \eta\}$. Suppose that for every i the function $\varphi_i(t) = t^{1-\theta_i}/(1-\theta_i)$, where $\theta_i \in [0, 1)$, and define $\theta = \max_{i \in I(\bar{x})} \theta_i$. Then the function $\varphi(t)$ reduces to

$$\varphi(t) = \int_0^t \max_{i \in I(\bar{x})} s^{-\theta_i} ds = \int_0^t s^{-\theta} ds = \frac{t^{1-\theta}}{1-\theta}, \forall t \in (0, \eta_1),$$

where the second equality holds because $\max_{i \in I(\bar{x})} s^{-\theta_i} = s^{-\theta}$ for $s \in (0, 1)$. In this way, we recovered a result by Li and Pong [14, Theorem 3.1], where

they proved that $f(x) = \min_{1\le i\le m} f_i(x)$ admits KL exponent θ at $\bar{x}$, where the function f_i has the KL property at $\bar{x}$ with KL exponent θ_i for each $i \in I(\bar{x})$.

(ii) Let us provide an example where the minimum of two lsc functions is continuous. Consider $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 1 - x^2, & x \neq 0; \\ 0, & x = 0. \end{cases} \text{ and } g(x) = \begin{cases} x^2/2, & x \neq 1; \\ 0, & x = 1. \end{cases}$$

Then the function $h(x) = \min\{f(x), g(x)\}$ satisfies

$$h(x) = \begin{cases} x^2/2, & |x| \le \sqrt{2/3}; \\ 1 - x^2, & |x| > \sqrt{2/3}, \end{cases}$$

which is a continuous function. However, it is worth noting that the minimum of lsc functions is usually not continuous. Therefore a more restrictive yet easier to be verified condition for Theorem 6.5 is that f_i is continuous for every i, in which case the continuity of f becomes automatic and our conclusion follows similarly.

Theorem 6.7. *Let $n_i \in \mathbb{N}, i = 1, \ldots, m,$, and let $n = \sum_{i=1}^m n_i$. For each i, let $f_i : \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$ be a proper and lsc function that is continuous on $\operatorname{dom} \partial f_i$. Furthermore, suppose that each f_i has the generalized KL property at $\bar{x}_i \in \operatorname{dom} \partial f_i$ with respect to $U_i = \mathbb{B}(\bar{x}_i; \varepsilon_i)$ for $\varepsilon_i > 0$, $\eta_i > 0$ and $\varphi_i \in \Phi_{\eta_i}$. Let $\varepsilon = \min \varepsilon_i$ and let η be a real number with $\eta < m \cdot \min \eta_i$. Define $\varphi : [0, \eta] \to \mathbb{R}$ by*

$$\varphi(t) = \int_0^t \max_{1\le i\le m} \left[(\varphi_i)'_- \left(\frac{s}{m} \right) \right] ds, \forall t \in (0, \eta],$$

and $\varphi(0) = 0$.

If the function $t \mapsto (\varphi_i)'_-(t)$ is invertible for each $i \in \{1, \ldots, m\}$, then the separable sum $f(x) = \sum_{i=1}^m f_i(x_i)$ has the generalized KL property at $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_m)$ with respect to $U = \mathbb{B}(\bar{x}; \varepsilon)$, η and φ.

Proof. Note that for each i, the function $\varphi_i \in \Phi_{\eta_i}$ is finite and right-continuous at 0. Hence by invoking Lemma 3.14 and Fact 3.15, one concludes that φ is finite and belongs to Φ_η with

$$\varphi'_-(t) \ge \max_{1\le i\le m} (\varphi_i)'_- \left(\frac{t}{m} \right) \ge (\varphi_i)'_- \left(\frac{t}{m} \right), \forall 1 \le i \le m. \tag{6.7}$$

Define $\phi_i(t) = (\varphi_i)'_-(t)$ for each $i \in \{1, \dots, m\}$. By Lemma 6.1, the function ϕ_i is decreasing for each i. By the assumption that f_i has the generalized KL property at $\bar{x}_i$, we have for $x_i \in U_i \cap [0 < f_i - f_i(\bar{x}_i) < \eta_i]$

$$\phi_i(f_i(x_i) - f_i(\bar{x}_i)) \geq \frac{1}{\operatorname{dist}(0, \partial f_i(x_i))}, \tag{6.8}$$

and furthermore, since ϕ_i^{-1} is decreasing, we have

$$f_i(x_i) - f_i(\bar{x}_i) = \phi_i^{-1}\left(\phi_i(f_i(x_i) - f_i(\bar{x}_i))\right) \leq \phi_i^{-1}\left(\frac{1}{\operatorname{dist}(0, \partial f_i(x_i))}\right).$$

Note that $1/\operatorname{dist}(0, \partial f_i(x_i)) \in \operatorname{dom} \phi_i^{-1}$ for each $i \in \{1, \dots, m\}$ because of (6.8). Shrinking ε_i if necessary, we assume $f_i(x_i) < f_i(\bar{x}_i) + \eta_i$ whenever $x_i \in U_i$. Therefore for every $x_i \in U_i$, we conclude that

$$f_i(x_i) - f_i(\bar{x}_i) \leq \phi_i^{-1}\left(\frac{1}{\operatorname{dist}(0, \partial f_i(x_i))}\right). \tag{6.9}$$

Note that the above inequality holds trivially when $f_i(x_i) \leq f_i(\bar{x}_i)$, because the right hand side is always positive.

Take $x = (x_1, \dots, x_m) \in U \cap [0 < f - f(\bar{x}) < \eta]$ and denote by $i^* = i(x)$ the index such that

$$\phi_{i^*}^{-1}\left(\frac{1}{\operatorname{dist}(0, \partial f_{i^*}(x_{i^*}))}\right) \geq \phi_i^{-1}\left(\frac{1}{\operatorname{dist}(0, \partial f_i(x_i))}\right), \forall i \in \{1, \dots, m\}.$$

Note that $\|x - \bar{x}\|^2 = \sum_{i=1}^m \|x_i - \bar{x}_i\|^2$. Then $\|x_i - \bar{x}_i\| \leq \|x - \bar{x}\| < \varepsilon \leq \varepsilon_i$ for each i. For simplicity, set $r = r(x) = 1/\operatorname{dist}(0, \partial f_{i^*}(x_{i^*}))$, where the value r depends on x because the index i^* does. Summing (6.9) from $i = 1$ to m yields

$$f(x) - f(\bar{x}) = \sum_{i=1}^m [f_i(x_i) - f_i(\bar{x}_i)] \leq \sum_{i=1}^m \phi_i^{-1}\left(\frac{1}{\operatorname{dist}(0, \partial f_i(x_i))}\right) \leq m \cdot \phi_{i^*}^{-1}(r). \tag{6.10}$$

Let $u \in \partial f(x)$ be such that $\|u\| = \operatorname{dist}(0, \partial f(x))$. Recall from Fact 3.21 that

$$\partial f(x) = \prod_{i=1}^m \partial f_i(x_i).$$

Then there exists $u_i \in \partial f_i(x_i)$ for each i such that $u = (u_1, \dots, u_m)$ and consequently for every i, one has,

$$\operatorname{dist}(0, \partial f(x)) = \|u\| \geq \|u_i\| \geq \operatorname{dist}(0, \partial f_i(x_i)).$$

In particular, one has

$$r \cdot \operatorname{dist}(0, \partial f(x)) \geq 1. \tag{6.11}$$

Now we prove the generalized KL property at $\bar{x}$. Take $x \in \mathbb{B}(\bar{x}; \varepsilon) \cap [0 < f - f(\bar{x}) < \eta]$. Note that the range of $\phi_{i^*}^{-1}$ satisfies $\operatorname{ran} \phi_{i^*}^{-1} = \operatorname{dom} \phi_{i^*} = \operatorname{dom}(\varphi_{i^*})'_- = (0, \eta_{i^*})$. Therefore $m \cdot \phi_{i^*}^{-1}(r) < m \cdot \eta_{i^*}$, where the value r depends on x. On the other hand, recall that η is defined to be a real number satisfying $\eta < m \cdot \min_i \eta_i \leq m \cdot \eta_{i^*}$. Therefore we need to consider the following two cases:

Case 1: If $m \cdot \phi_{i^*}^{-1}(r) < \eta$, then one has the following from (6.10) and the fact that φ'_- is decreasing

$$\varphi'_-(f(x) - f(\bar{x})) \geq \varphi'_- \left(m \cdot \phi_{i^*}^{-1}(r)\right). \tag{6.12}$$

Hence we have

$$\begin{aligned}
&\varphi'_-(f(x) - f(\bar{x})) \cdot \operatorname{dist}(0, \partial f(x)) \geq \varphi'_- \left(m \cdot \phi_{i^*}^{-1}(r)\right) \cdot \operatorname{dist}(0, \partial f(x)) \\
&\geq (\varphi_{i^*})'_- \left(\phi_{i^*}^{-1}(r)\right) \cdot \operatorname{dist}(0, \partial f(x)) = \phi_{i^*} \left(\phi_{i^*}^{-1}(r)\right) \cdot \operatorname{dist}(0, \partial f(x)) \\
&= r \cdot \operatorname{dist}(0, \partial f(x)) \geq 1.
\end{aligned}$$

where the second inequality is implied by (6.7) and the last one holds because of (6.11).

Case 2: Now we consider the case where $m \cdot \phi_{i^*}^{-1}(r) \geq \eta$. Note that $\eta < m \cdot \min \eta_i \leq m \cdot \eta_{i^*}$. Then we have $\eta/m \in \operatorname{dom} \phi_{i^*}$ and $\phi_{i^*}(\eta/m) < \infty$. On the other hand, the assumption $m \cdot \phi_{i^*}^{-1}(r) \geq \eta$ implies that $r \leq \phi_{i^*}(\frac{\eta}{m})$. Altogether, one concludes that

$$\begin{aligned}
&\varphi'_-(f(x) - f(\bar{x})) \cdot \operatorname{dist}(0, \partial f(x)) \geq \varphi'_-(\eta) \cdot \operatorname{dist}(0, \partial f(x)) \\
&\geq (\varphi_{i^*})'_- \left(\frac{\eta}{m}\right) \cdot \operatorname{dist}(0, \partial f(x)) \geq r \cdot \operatorname{dist}(0, \partial f(x)) \geq 1,
\end{aligned}$$

where the second inequality is implied by (6.7) and the last one holds because of (6.11). □

Remark 6.8. (i) A sufficient condition for $(\varphi_i)'_-(t)$ being invertible is that f_i admits a strictly concave desingularizing function $(\varphi_i)'_-(t)$, by using Lemma 6.1(i). It is also worth noting that one can take $\eta = \infty$ if $\eta_i = \infty$ for every i, in which case $m \cdot \phi_{i^*}^{-1}(r) < \eta$ is trivially true. Then by applying a similar argument in Case 1, one shows that f has the generalized KL property at $\bar{x}$ with respect to $U = \mathbb{B}(\bar{x}; \varepsilon)$, η and $\varphi(t)$ given by

$$\varphi(t) = \int_0^t \max_{1 \leq i \leq m} \left[(\varphi_i)'_- \left(\frac{s}{m}\right)\right] ds, \forall t \in (0, \eta),$$

and $\varphi(0) = 0$.

(ii) Setting $\eta < m$ and $\varphi_i(t) = t^{1-\theta_i}/(1-\theta_i)$, where $\theta_i \in [0,1)$ for every i, we have for $t \in (0,\eta] \subseteq (0,m)$

$$\varphi(t) = \int_0^t \max_{1\le i\le m} \left(\frac{s}{m}\right)^{-\theta_i} ds = \int_0^t \left(\frac{s}{m}\right)^{-\theta} ds = \frac{m^\theta}{1-\theta} \cdot t^{1-\theta},$$

where $\theta = \max_{1\le i\le m} \theta_i$, from which a result by Li and Pong [14, Theorem 3.3] is recovered, where the second equality holds because $\max_{1\le i\le m} t^{-\theta_i} = t^{-\theta}$ for $t \in (0,1)$ and $s/m < 1$.

Example 6.9. Let $n \in \mathbb{N}$. The following statements hold:

(i) The function $f(x) = \sum_{i=1}^n |x_i|^2$ has the generalized KL property at $a = (a_1, \dots, a_n)$ with respect to $U = \mathbb{R}^n$, $\eta = \infty$ and $\varphi(t) = \sqrt{n \cdot t}$.

(ii) The function $g(x) = \sum_{i=1}^n -\ln(1-|x_i|^2)$ has the generalized KL property at 0 with respect to $U = \mathbb{R}^n$, $\eta = \infty$ and $\varphi(t) = n\sqrt{1-e^{-t/n}}$.

(iii) The function $h(x) = \sum_{i=1}^n \tan(|x_i|^2)$ has the generalize KL property at 0 with respect to $U = \mathbb{R}^n$, $\eta = \infty$ and $\varphi(t) = n\sqrt{\arctan(t/n)}$.

Proof. (i) Define $f_i(x_i) = |x_i|^2$. Note that f_i is a even and convex function with $y = f_i(x) \Leftrightarrow |x_i| = \sqrt{y}$ for every i. Then by Theorem 5.12, the function f_i has the generalized KL property at 0 with respect to $U_i = \mathbb{R}$, $\eta_i = \infty$ and $\varphi_i(t) = \sqrt{t}$. Note that $\varphi_i'(t) = t^{-1/2}/2$ is invertible. Hence we have for $t \in (0,\eta)$

$$\varphi(t) = \int_0^t \max_{1\le i\le n} \varphi_i'\left(\frac{s}{n}\right) ds = \frac{1}{2}\int_0^t \left(\frac{s}{n}\right)^{-\frac{1}{2}} ds = \sqrt{n \cdot t}.$$

Applying Theorem 6.7 proves that f has the generalized KL property at $\bar{x}$ with $U = \mathbb{R}^n$, $\eta = \infty$ and $\varphi(t)$, where $\eta = \infty$ because $\eta_i = \infty$ for each i, see Remark 6.8(i).

(ii) Define $g_i(x_i) = -\ln(1-|x_i|^2)$ for each i. We learn from Example 5.14(i) that for each $i \in \{1,\dots,n\}$, the function g_i has the generalized KL property at 0 with respect to $U_i = \mathbb{R}$, $\eta_i = \infty$ and $\varphi_i(t) = \sqrt{1-e^{-t}}$. Note that $\varphi''(t) = e^{-t}(e^{-t}-2)/\left(4(1-e^{-t})^{3/2}\right) < 0$ for $t > 0$. Hence $\varphi_i' = e^{-t}/(2\sqrt{1-e^{-t}})$ is invertible, and Theorem 6.7 implies that g has the generalized KL property at 0 with $U = \mathbb{R}^n$, $\eta = \infty$ and

$$\varphi(t) = \int_0^t \max_{1\le i\le n} \varphi_i'(\frac{s}{n}) ds = \frac{1}{2}\int_0^t \frac{e^{-s/n}}{\sqrt{1-e^{-s/n}}} ds = n\sqrt{1-e^{-t/n}}.$$

Note that we take $\eta = \infty$ according to Remark 6.8(i).

(iii) Define $h_i(x_i) = \tan(|x_i|^2)$ for each i. Then Example 5.14(ii) shows that for each i the function h_i has the generalized KL property at 0 with respect to $U_i = \mathbb{R}$, $\eta_i = \infty$ and $\varphi_i(t) = \sqrt{\arctan(t)}$ whose derivative $\varphi_i'(t) = (t^2+1)^{-1}/\left(2\sqrt{\arctan(t)}\right)$ is invertible. Therefore according to Theorem 6.7 and Remark 6.8(i), one shows that h has the generalized KL property with respect to $U = \mathbb{R}^n$, $\eta = \infty$ and $\varphi(t)$ given by

$$\varphi(t) = \int_0^t \max_{1\le i\le n} \frac{(s^2/n^2+1)^{-1}}{2\sqrt{\arctan(s/n)}} ds = n\sqrt{\arctan(t/n)},$$

which completes the proof. □

The following technical lemma will help us to obtain a composition rule for the generalized KL property.

Lemma 6.10. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a smooth map and let $\bar{x} \in \mathbb{R}^n$. If the Jacobian $\nabla F(\bar{x})$ has rank m, then there exist $\alpha > 0$ and $\varepsilon > 0$ such that for $x \in \mathbb{B}(\bar{x};\varepsilon)$,*

$$\|y\| \le \alpha \left\|(\nabla F(x))^*(y)\right\|, \forall y \in \mathbb{R}^m\backslash\{0\}. \tag{6.13}$$

Proof. By the assumption, $\nabla F(\bar{x}) : \mathbb{R}^n \to \mathbb{R}^m$ is surjective continuous linear map. Then by using the open mapping theorem, there exists $\alpha > 0$ such that

$$\mathbb{B}_{\mathbb{R}^m} \subseteq \nabla F(\bar{x})(\frac{\alpha}{2} \cdot \mathbb{B}_{\mathbb{R}^n}), \tag{6.14}$$

where $\mathbb{B}_{\mathbb{R}^m}$ and $\mathbb{B}_{\mathbb{R}^n}$ denote the Euclidean unit balls in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. Note that the map F is assumed to be smooth. Then there exists $\varepsilon > 0$ such that

$$\|x - \bar{x}\| < \varepsilon \Rightarrow \|\nabla F(x) - \nabla F(\bar{x})\| < \frac{1}{\alpha}.$$

We claim that for x with $\|x - \bar{x}\| < \varepsilon$,

$$\mathbb{B}_{\mathbb{R}^m} \subseteq \nabla F(x)(\alpha \mathbb{B}_{\mathbb{R}^n}). \tag{6.15}$$

Take $u \in \mathbb{B}_{\mathbb{R}^m}$. We will prove the above inclusion by constructing $v \in \alpha\mathbb{B}_{\mathbb{R}^n}$ such that $u = \nabla F(x)(v)$. By using the inclusion (6.14), there exists some $v_1 \in \frac{\alpha}{2}\mathbb{B}_{\mathbb{R}^n}$ such that $\nabla F(\bar{x})(v_1) = u$. Then we have

$$\|\nabla F(x)(v_1) - u\| = \|\nabla F(x)(v_1) - \nabla F(\bar{x})(v_1)\| \le \frac{1}{\alpha} \cdot \frac{\alpha}{2} = \frac{1}{2}.$$

Hence $u - \nabla F(x)(v_1) \in \frac{1}{2}\mathbb{B}_{\mathbb{R}^m}$ and again by (6.14) there exists $v_2 \in \frac{\alpha}{4}\mathbb{B}_{\mathbb{R}^n}$ such that $\nabla F(\bar{x})(v_2) = u - \nabla F(x)(v_1)$, which further implies

$$\begin{aligned}\|\nabla F(x)(v_1+v_2) - u\| &= \|\nabla F(x)(v_2) - (u - \nabla F(x)(v_1))\| \\ &= \|\nabla F(x)(v_2) - \nabla F(\bar{x})(v_2)\| \leq \frac{1}{4}.\end{aligned}$$

Repeating the same process one obtains a sequence $(v_n)_{n\in\mathbb{N}}$ satisfying

$$\left\|\nabla F(x)\left(\sum_{k=1}^{n} v_k\right) - u\right\| \leq \frac{1}{2^n} \text{ and } \|v_n\| \leq \frac{\alpha}{2^n}, \forall n \in \mathbb{N}.$$

The latter inequality implies there exists $v \in \mathbb{R}^n$ such that $\lim_{n\to\infty}\sum_{k=1}^{n} v_k = v$ and $\|v\| \leq \alpha$, while from the former one we have $\nabla F(x)(v) = u$, which proves our claim. Let $u \in \mathbb{B}_{\mathbb{R}^m}$ obey $\|u\| = 1$ and suppose that $u = \nabla F(x)(v)$ for some $v \in \alpha\mathbb{B}_{\mathbb{R}^n}$. Then we have

$$\begin{aligned}\frac{\|u\|}{\alpha^2} &= \frac{\langle u, u\rangle}{\alpha^2} = \frac{\langle u, \nabla F(x)(v)\rangle}{\alpha^2} = \frac{\langle (\nabla F(x))^*(u), v\rangle}{\alpha^2} \\ &\leq \frac{\|(\nabla F(x))^*(u)\|\,\|v\|}{\alpha^2} \leq \frac{\alpha\,\|(\nabla F(x))^*(u)\|}{\alpha^2} \\ &= \frac{\|(\nabla F(x))^*(u)\|}{\alpha},\end{aligned}$$

which implies that $\|u\| \leq \alpha\,\|(\nabla F(x))^*(u)\|$. Let $u = y/\|y\|$ for nonzero $y \in \mathbb{R}^m$. Then one has $\|y\| \leq \alpha\,\|(\nabla F(x))^*(y)\|$, as claimed. □

Theorem 6.11. *Suppose that $f(x) = (g \circ F)(x)$, where $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is proper lsc and $F : \mathbb{R}^n \to \mathbb{R}^m$ is a smooth map. Let $\bar{x} \in \operatorname{dom}\partial f$ and let $\varepsilon_0 > 0$. Assume that g has the generalized KL property at $F(\bar{x})$ with respect to $U_0 = \mathbb{B}(F(\bar{x}); \varepsilon_0)$, $\eta > 0$ and $\varphi \in \Phi_\eta$. If the Jacobian $\nabla F(\bar{x})$ has rank m, then there exist $\alpha > 0$ and $\varepsilon_1 > 0$ such that (6.13) holds. Furthermore, there exits $\varepsilon \in (0, \min\{\varepsilon_0, \varepsilon_1\}]$ such that f has the generalized KL property at $\bar{x}$ with respect to $U_1 = \mathbb{B}(\bar{x}; \varepsilon)$, $\eta > 0$ and $\alpha \cdot \varphi \in \Phi_\eta$.*

Proof. By assumption for $y \in \mathbb{B}(F(\bar{x}); \varepsilon_0) \cap [0 < g - g(F(\bar{x})) < \eta]$ we have

$$\varphi'_-(g(y) - g(F(\bar{x}))) \cdot \operatorname{dist}(0, \partial g(y)) \geq 1. \tag{6.16}$$

On the other hand, Lemma 6.10 implies that there exist $\alpha > 0$ and $\varepsilon_1 > 0$ such that for $x \in \mathbb{B}(\bar{x}; \varepsilon_1)$,

$$\|y\| \leq \alpha\,\|(\nabla F(x))^*(y)\|, \forall y \in \mathbb{R}^m \backslash \{0\}. \tag{6.17}$$

Moreover, by applying Fact 3.19, one has $\partial f(x) = \nabla F(\bar{x})^* \partial g(F(x)) = \{\nabla F(\bar{x})^* v : u \in \partial g(F(x))\}$. Let $v \in \partial f(x)$ be such that $\|v\| = \operatorname{dist}(0, \partial f(x))$. Then we have for some $u \in \partial g(F(x))$

$$\operatorname{dist}(0, \partial f(x)) = \|v\| = \|(\nabla F(x))^* u\| \geq \frac{\|u\|}{\alpha} \geq \frac{\operatorname{dist}(0, \partial g(F(x)))}{\alpha}, \quad (6.18)$$

where the first inequality is implied by (6.17). Suppose that $\|F(x) - F(\bar{x})\| < \varepsilon_0$ whenever $\|x - \bar{x}\| < \varepsilon_2$ for some ε_2 and set $\varepsilon = \min\{\varepsilon_0, \varepsilon_1, \varepsilon_2\}$. Then for $x \in \mathbb{B}(\bar{x}; \varepsilon) \cap [0 < f - f(\bar{x}) < \eta]$

$$\begin{aligned} &(\alpha \cdot \varphi)'_-(f(x) - f(\bar{x})) \operatorname{dist}(0, \partial f(x)) \\ &\geq \varphi'_-(g(F(x)) - g(F(\bar{x})) \operatorname{dist}(0, \partial g(F(x))) \geq 1, \end{aligned}$$

where the last inequality follows from (6.16) and (6.18). □

Remark 6.12. In [14, Theorem 3.2], Li and Pong proved (6.15) by applying Graves theorem [11, Theorem 5D.2] to the continuous map $x \mapsto \nabla F(x)$. Then the composition rule for KL exponent is obtained in a similar manner. In contrast, we provided an alternative proof for (6.15) by using Lemma 6.10. Moreover, our result applies to desingularizing functions of all forms, while only those of the form $\varphi(t) = c \cdot t^{1-\theta}$ are considered in [14, Theorem 3.2].

Corollary 6.13. *Suppose that $f(x) = g(Ax - b)$, where $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is proper lsc, $A \in \mathbb{R}^{m \times n}$ has rank m and $b \in \mathbb{R}^m$. Let $\bar{x} \in \operatorname{dom} \partial f$ and $\varepsilon > 0$. Set $r = \sqrt{\lambda_{\min}(AA^*)} > 0$. If g has the generalized KL property at $A\bar{x} - b$ with respect to $U_1 = \mathbb{B}(A\bar{x} - b; \varepsilon)$, $\eta > 0$ and $\varphi(t) \in \Phi_\eta$, then f has the generalized KL property at $\bar{x}$ with respect to $U_2 = \mathbb{B}(\bar{x}; \varepsilon / \|A\|)$, $\eta > 0$ and $\varphi(t)/r$. Note that $U_2 = \mathbb{R}^n$ if $U_1 = \mathbb{R}^m$.*

Proof. Let us first observe that $r > 0$. For $y \in \mathbb{R}^m$, one has $\langle y, AA^* y \rangle = \langle A^* y, A^* y \rangle = \|A^* y\|^2 = 0 \Leftrightarrow A^* y = 0$. Note that A is surjective. Hence A^* is injective and the kernel satisfies $\ker A^* = \{0\}$, which further implies that AA^* is positive-definite because $\langle y, AA^* y \rangle > 0$ for every non-zero y. Therefore one concludes that $r > 0$. On the other hand, by the definition, we have

$$r = \sqrt{\lambda_{\min}(AA^*)} = \sqrt{\inf_{y \neq 0} \frac{\langle y, AA^* y \rangle}{\|y\|^2}} = \inf_{y \neq 0} \frac{\|A^* y\|}{\|y\|} \leq \frac{\|A^* y\|}{\|y\|}, \forall y \in \mathbb{R}^m \backslash \{0\},$$

which means that $F(x) = Ax - b$ satisfies (6.13) for every $x \in \mathbb{R}^n$ with $\alpha = 1/r$ and $\varepsilon_1 = \infty$. Then applying a similar argument in Theorem 6.11 completes the proof. □

Chapter 7

The exact modulus of the generalized KL property

When verifying the generalized KL property of f at $\bar{x}$ with respect to neighborhood $U \ni \bar{x}$ and $\eta \in (0, \infty]$, one needs to find a desingularizing function $\varphi \in \Phi_\eta$ to "sharpen" f around $\bar{x}$. The goal of this chapter is to address the question:

What is the optimal desingularizing function with respect to U and η?

To this end, we define the *exact modulus of the generalized KL property*, which is the optimal desingularizing function for the generalized KL property. In contrast to the classic theory, where the desingularizing function usually assumes the form $\varphi(t) = c \cdot t^{1-\theta}$ with $c > 0$ and $\theta \in [0, 1)$, the exact modulus takes various forms, depending on the given function f. We will show that the exact modulus is optimal, in the sense that it is the smallest among all possible desingularizing functions with respect to given U and η. Several examples are also given to illustrate this pleasant property. Moreover, we provide various ways to compute or at least estimate the exact modulus. Furthermore, we show that there exists a function who has the generalized KL property but the desingularizing function cannot take the form $\varphi(t) = c \cdot t^{1-\theta}$.

7.1 Definition and basic properties

Definition 7.1. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc. Let $\bar{x} \in \operatorname{dom} \partial f$ and let $U \subseteq \operatorname{dom} \partial f$ be a neighborhood of $\bar{x}$. Let $\eta \in (0, \infty]$. Furthermore, define $h : (0, \eta) \to \mathbb{R}$ by

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

Suppose that $h(s) < \infty$ for $s \in (0, \eta)$. The *exact modulus of the generalized KL property* of f at $\bar{x}$ with respect to U and η is the function $\tilde{\varphi} : [0, \eta) \to \mathbb{R}_+$,

$$t \mapsto \int_0^t h(s)ds, \forall t \in (0, \eta), \tag{7.1}$$

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 7.2. (i) The exact modulus $\tilde{\varphi}(t)$ is designed to be the optimal desingularizing function, in the sense that $\tilde{\varphi}$ is the smallest possible candidate in Φ_η for the generalized KL property of f at $\bar{x}$ with respect to given $U \ni \bar{x}$ and η, see Proposition 7.3 and Figure 7.1. Moreover, we will learn from Chapter 8 that the exact modulus of the generalized KL property has significant impact on the *proximal alternating linearized minimization* (PALM) algorithm.

(ii) Note that $\lim_{s\to 0^+} h(s)$ could be infinity, in which case the function $\tilde{\varphi}(t)$ represents an improper integral. For instance, let $f(x) = x^2$ and consider the exact modulus of the generalized KL property of f at 0 with respect to $U = \mathbb{R}$ and $\eta = \infty$. Then we have $\operatorname{dist}(0, \partial f(x)) = |f'(x)| = |2x|$, and

$$\begin{aligned} h(s) &= \sup\{|2x|^{-1} : x \in \mathbb{R} \cap [0 < f < \infty], s \le x^2\} \\ &= \sup\{|2x|^{-1} : x \neq 0, |x| \ge \sqrt{s}\} = 1/\left(2\sqrt{s}\right). \end{aligned}$$

Hence $\lim_{s\to 0^+} h(s) = \infty$.

(iii) The assumption that $h(s) < \infty$ for $s \in (0, \eta)$ is essential. For example, consider the exact modulus of the generalized KL property of the function $f(x) = 1 - e^{-|x|}$ at 0. By Fact 3.18 one has for $x \neq 0$,

$$\partial f(x) = f'(x) = \operatorname{sgn}(x) \cdot e^{-|x|} \Rightarrow \operatorname{dist}^{-1}(0, \partial f(x)) = e^{|x|}.$$

Let $U = \mathbb{R}$ and $\eta_1 = 1$. Then one has

$$\begin{aligned} h_1(s) = h_{U,\eta_1}(s) &= \sup\{\operatorname{dist}^{-1}(0, \partial f(x)) : \mathbb{R} \cap [0 < f < 1], s \le f(x)\} \\ &= \sup\{e^{|x|} : x \neq 0, |x| > -\ln(1 - s)\} = \infty. \end{aligned}$$

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

$$\begin{aligned} h_2(s) = h_{U,\eta_2}(s) &= \sup\{\operatorname{dist}^{-1}(0, \partial f(x)) : x \in \mathbb{R} \cap [0 < f < \eta_2], s \le f(x)\} \\ &= \sup\{e^{|x|} : -\ln(1 - s) \le |x| < -\ln(1 - \eta_2)\} = \frac{1}{1 - \eta_2}. \end{aligned}$$

Next we show the exact modulus $\tilde{\varphi}$ is the optimal desingularzing function, in the sense that $\tilde{\varphi}$ is the smallest candidate in Φ_η such that the generalized KL property holds.

Proposition 7.3. *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 $\bar{x}$, $\eta \in (0, \infty]$ and let $\varphi(t) \in \Phi_\eta$. Suppose that f has the generalized KL property at $\bar{x}$ with respect to U, η and φ. Then the exact modulus of the generalized KL property of f at $\bar{x}$ with respect to U and η, denoted by $\tilde{\varphi}$, is well-defined and satisfies:*

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

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

Proof. **Step 1:** Let us show first that $\tilde{\varphi}(t) \leq \varphi(t)$ on $[0, \eta)$, from which the well-definedness of $\tilde{\varphi}$ readily follows. If $U \cap [0 < f - f(\bar{x}) < \eta] = \emptyset$, then by our convention $\tilde{\varphi}(t) = 0 \leq \varphi(t)$ for every $t \in [0, \eta)$. Therefore we proceed with assuming $U \cap [0 < f - f(\bar{x}) < \eta] \neq \emptyset$. By assumption, one has for $x \in U \cap [0 < f - f(\bar{x}) < \eta]$,

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

which guarantees that $\operatorname{dist}(0, \partial f(x)) > 0$. Fix $s \in (0, \eta)$ and recall from Lemma 6.1(i) that $\varphi'_-(t)$ is decreasing. Then for $x \in U \cap [0 < f - f(\bar{x}) < \eta]$ with $s \leq f(x) - f(\bar{x})$ we have

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

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

$$h(s) \leq \varphi'_-(s),$$

where $h(s) = \sup\{\operatorname{dist}^{-1}(0, \partial f(x)) : x \in U \cap [0 < f - f(\bar{x}) < \eta], s \leq f(x) - f(\bar{x})\}$, which implies that for $t \in (0, \eta)$,

$$\tilde{\varphi}(t) = \int_0^t h(s) ds \leq \int_0^t \varphi'_-(s) ds = \varphi(t) < \infty,$$

where the last equality follows from Lemma 6.1(iii).

Step 2: 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 $s_1 \leq s_2$. Then for $x \in U \cap [0 < f - f(\bar{x}) < \eta]$, one has

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

implying that $h(s_2) \leq h(s_1)$. Therefore $h(s)$ is decreasing. Invoking Fact 3.15, one concludes that $\tilde{\varphi} \in \Phi_\eta$ and for every $t \in (0, \eta)$

$$\varphi'_-(t) \geq h(t).$$

Step 3: Let $t \in (0, \eta)$. Then for $x \in U \cap [0 < f - f(\bar{x}) < \eta]$ with $t = f(x) - f(\bar{x})$ we have

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

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

Now we give several examples, beginning with the exact modulus of the generalized KL property for the log-barrier function.

Example 7.4. (Exact modulus at non-stationary point) Let $a > 0$ and $\bar{x} < a$. Consider the function $f : \mathbb{R} \to \overline{\mathbb{R}}$ given by

$$f(x) = \begin{cases} -\ln(a - x), & x < a; \\ \infty, & \text{otherwise.} \end{cases}$$

Then f admits $\tilde{\varphi}(t) = \tilde{\varphi}_{a,\bar{x}}(t) = (a - \bar{x})(1 - e^{-t})$ as the exact modulus of generalized KL property of f at $\bar{x}$ with respect to $U = \mathbb{R}$ and $\eta = \infty$.

Proof. Note that f is smooth on $(-\infty, a)$. Hence by Fact 3.18 one gets $\partial f(x) = f'(x) = 1/(a - x) > 0$ for every $x < a$. Let $s \in (0, \infty)$. Then one has $0 < f(x) - f(\bar{x}) < \infty \Leftrightarrow \bar{x} < x < a$ by the fact that $f(x)$ is increasing. Moreover, we have

$$s \leq f(x) - f(\bar{x}) = -\ln(a - x) + \ln(a - \bar{x}) = \ln\left(\frac{a - \bar{x}}{a - x}\right) \Leftrightarrow x \geq a - e^{-s}(a - \bar{x}).$$

Note that $a - e^{-s}(a - \bar{x}) > a - (a - \bar{x}) = \bar{x}$ for $s \in (0, \infty)$. Then we have

$$\begin{aligned} h(s) &= \sup\left\{|a - x| : x \in U \cap [0 < f - f(\bar{x}) < \infty], s \leq f(x) - f(\bar{x})\right\} \\ &= \sup\left\{|a - x| : a - e^{-s}(a - \bar{x}) \leq x < a\right\} = |e^{-s}(a - \bar{x})| = e^{-s}(a - \bar{x}). \end{aligned}$$

It then follows from the definition of $\tilde{\varphi}(t)$ that

$$\tilde{\varphi}(t) = \int_0^t h(s)ds = \int_0^t e^{-s}(a - \bar{x})ds = (1 - e^{-t})(a - \bar{x}),$$

as claimed. □

Remark 7.5. Note that the log-barrier function has no stationary points. Hence by using Fact 4.4 one concludes that f admits $\varphi_\theta(t) = t^{1-\theta}/(\varepsilon(1-\theta))$ at every $\bar{x} < a$ for every $\theta \in [0,1)$ and for some $\varepsilon \in (0,1]$, which means that the KL inequality is only valid for $x \in (\bar{x}-\varepsilon, \bar{x}+\varepsilon) \cap [0 < f - f(\bar{x}) < \varepsilon]$. In contrast, the exact modulus has a different form and takes all $x \in \mathbb{R} \cap [0 < f - f(\bar{x}) < \infty]$ into account.

Our next set of examples shows that the exact modulus of the generalized KL property is not necessarily differentiable, which justifies the nonsmooth extension of desingularizing functions in Definition 6.2.

Example 7.6. The following statements are true:

(i) Let $\rho > 0$. Consider the function given by

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

Define $\tilde{\varphi}_1(t)$ by

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

Then $\tilde{\varphi}_1$ is the exact modulus of the generalized KL property of f at $\bar{x} = 0$ with respect to $U = \mathbb{R}$ and $\eta = \infty$.

(ii) Consider the following function

$$g(x) = \begin{cases} 1 - e^{-|x|}, & |x| \le 1; \\ (1 - e^{-1})|x|, & |x| > 1. \end{cases}$$

Define $\tilde{\varphi}_2$ by

$$\tilde{\varphi}_2(t) = \begin{cases} e \cdot t, & t < 1 - e^{-1}; \\ \frac{e}{e-1} \cdot t + e - 2, & t \ge 1 - e^{-1}. \end{cases}$$

Then $\tilde{\varphi}_2$ is the exact modulus of generalized KL property of g at $\bar{x} = 0$ with respect to $U = \mathbb{R}$ and $\eta = \infty$.

Proof. Recall from Example 6.4(i), one has

$$\operatorname{dist}^{-1}(0, \partial f(x)) = \begin{cases} 1/|x|, & 0 < |x| \le \rho; \\ 1/2\rho, & |x| > \rho. \end{cases}$$

It follows that for $s \in (0, \rho^2/2]$,

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

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

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

Hence $\tilde{\varphi}(t) = \int_0^t h_1(s)ds = \sqrt{2t}$ for $t \le \rho^2/2$ and $\tilde{\varphi}_1(t) = \int_{\rho^2/2}^t h_1(s)ds + \int_0^{\rho^2/2} h_1(s)ds = t/(2\rho) - \rho/4 + \rho = t/(2\rho) + 3\rho/4$.

(ii) Recall from Example 6.4(ii) that for nonzero x

$$\mathrm{dist}^{-1}(0, \partial g(x)) = \begin{cases} \frac{e}{e-1}, & |x| > 1; \\ e^{|x|}, & 0 < |x| \le 1. \end{cases}$$

Hence for $s \in (0, 1 - e^{-1}]$,

$$\begin{aligned} h_2(s) &= \sup\{\mathrm{dist}^{-1}(0, \partial g(x)) : x \in \mathbb{R} \cap [0 < g < \infty], s \le g(x)\} \\ &= \sup\{\mathrm{dist}^{-1}(0, \partial g(x)) : x \neq 0, |x| \ge -\ln(1-s)\} = e, \end{aligned}$$

while for $s \in (1 - e^{-1}, \infty)$,

$$\begin{aligned} h_2(s) &= \sup\{\mathrm{dist}^{-1}(0, \partial g(x)) : x \in \mathbb{R} \cap [0 < g < \infty], s \le g(x)\} \\ &= \sup\left\{ \mathrm{dist}^{-1}(0, \partial g(x)) : x \neq 0, x \ge \frac{e}{e-1}s \right\} = \frac{e}{e-1}. \end{aligned}$$

Then $\tilde{\varphi}_2(t) = \int_0^t e ds = e \cdot t$ for $t \in (0, 1 - e^{-1}]$ and $\tilde{\varphi}_2(t) = e \cdot (1 - e^{-1}) + \int_{1-e^{-1}}^t \frac{e}{e-1} ds = \frac{e}{e-1} \cdot t + e - 2$ for $t \in (1 - e^{-1}, \infty)$. □

Remark 7.7. Note that $\tilde{\varphi}_1$ and $\tilde{\varphi}_2$ are not differentiable at $\rho^2/2$ and $1 - e^{-1}$, respectively. On the other hand, we learn from Proposition 7.3 that $\tilde{\varphi}_1$ and $\tilde{\varphi}_2$ are the optimal desingularizing functions for f and g, respectively. Hence the nonsmooth extension of desingularizing functions is necessary.

Now we revisit Proposition 5.8, from which we obtain a simple formula for the exact modulus of the generalized KL property for locally convex and C^1 functions on the line.

Proposition 7.8. *Let $f : \mathbb{R} \to \overline{\mathbb{R}}$ be proper and lsc. Let $\bar{x}$ be a stationary point. Suppose that there exists interval $(a, b) \subseteq \operatorname{int} \operatorname{dom} f$, where $-\infty \leq a < b \leq \infty$, on which f is convex on (a, b) and C^1 on $(a, b)\backslash\{\bar{x}\}$. Set $\eta = \eta_{a,b,\bar{x}}$, $f_1 = f_{a,\bar{x}}$ and $f_2 = f_{b,\bar{x}}$ (recall (5.1) and (5.2)). Furthermore, define $\tilde{\varphi} : [0, \eta) \to \mathbb{R}_+$ by*

$$\tilde{\varphi}(t) = \int_0^t \max\{(-f_1^{-1})'(s), (f_2^{-1})'(s)\} ds, \forall t \in (0, \eta) \tag{7.2}$$

and $\varphi(0) = 0$. Then $\tilde{\varphi}(t)$ is the exact modulus of the generalized KL property at $\bar{x}$ with respect to $U = (a, b)$ and η.

Proof. Replacing $f(x)$ by $g(x) = f(x + \bar{x}) - f(\bar{x})$ if necessary, we assume 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 $f'_-(0) \leq 0 \leq f'_+(0)$. We learn from Fact 3.13 that $f'_-(x)$ and $f'_+(x)$ are increasing functions. Combing the C^1 assumption, we have $f'(x) = f'_-(x) \leq f'_-(0) \leq 0$ on $(a, 0)$ and $f'(x) = f'_+(x) \geq f'_+(0) \geq 0$ on $(0, b)$. Invoking Fact 3.18, one gets for $x \in (a, b)\backslash\{0\}$

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

Hence the function $x \mapsto \operatorname{dist}(0, \partial f(x))$ is decreasing on $(a, 0)$ and increasing on $(0, b)$.

Now we work towards showing that $h(s) = \max\{(-f_1^{-1})'(s), (f_2^{-1})'(s)\}$, where $h(s)$ is the function given in Definition 7.1. Recall that $f'(x)$ is increasing on $(a, b)\backslash\{0\}$ with $f'(x) \leq 0$ on $(a, 0)$ and $f'(x) \geq 0$ on $(0, b)$. Shrinking the interval (a, b) if necessary, we only need to consider the following cases:

Case 1: Consider first the case where $f'_1(x) < 0$ 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 we have

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

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

$$s \leq f(x) = f_1(x) \Leftrightarrow f_1^{-1}(s) \geq x. \tag{7.3}$$

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

$$s \leq f(x) = f_2(x) \Leftrightarrow f_2^{-1}(s) \leq x. \tag{7.4}$$

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

$$s \leq f(x) \Leftrightarrow x \in (a, f_1^{-1}(s)] \cup [f_2^{-1}(s), b).$$

On the other hand, we have $0 < f(x) < \eta \Leftrightarrow x \in (f_1^{-1}(\eta), f_2^{-1}(\eta))\backslash\{0\}$, 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 7.1 satisfies

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

where the third equality is implied the fact that $x \mapsto \mathrm{dist}^{-1}(0, \partial f(x))$ 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 $s \in (0, \eta)$

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

where the second equality is implied by (7.4), $U \cap [0 < f < \eta] = (0, b)$ 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 $U \cap [0 < f < \eta] = \emptyset$. According to our convention, $(-f_1^{-1})'(s)$, $(f_2^{-1})'(s)$ and $\tilde{\varphi}(t)$ are set to be constant 0. Hence we have $\tilde{\varphi}(t) = \int_0^t 0 \, ds = 0$, which completes the proof. □

Example 7.9. Let $p \geq 1$. Then the following assertions hold:

(i) The exact modulus of the generalized KL property of $f(x) = |x|^p$ at 0 with respect to $U_1 = \mathbb{R}$ and $\eta_1 = \infty$ is $\tilde{\varphi}_1(t) = \sqrt[p]{t}$.

(ii) The exact modulus of the generalized KL property of $g(x) = -\ln(1 - |x|^p)$ at 0 with respect to $U_2 = (-1, 1)$ and $\eta_2 = \infty$ is $\tilde{\varphi}_2(t) = \sqrt[p]{1 - \exp(-t)}$.

(iii) The exact modulus of the generalized KL property of $h(x) = \tan|x|^p$ at 0 with respect to $U_3 = (-\pi/2, \pi/2)$ and $\eta_3 = \infty$ is $\tilde{\varphi}_3(t) = \sqrt[p]{\arctan t}$.

Proof. Note that the above functions satisfy conditions in Proposition 7.8 with $\bar{x} = 0$ and the neighborhood $(a, b) \ni \bar{x}$ being their respective domain. Hence it suffices to compute the integral given in (7.2).

(i) Define $f_1(x) = f(x)$ for $x \leq 0$ and $f_2(x) = f(x)$ for $x \geq 0$. Since $s = |x|^p \Leftrightarrow |x| = \sqrt[p]{s}$ for $s \in (0, \infty)$, we conclude that $-f_1^{-1}(s) = f_2^{-1}(s) = \sqrt[p]{s}$. Applying Proposition 7.8, one gets the following:

$$\tilde{\varphi}_1(t) = \int_0^t \max\{(-f_1^{-1})'(s), (f_2^{-1})'(s)\}ds = f_2^{-1}(t) = \sqrt[p]{t}.$$

(ii) Define $g_1(x) = g(x)$ for $x \in (-1, 0]$ and $g_2(x) = g(x)$ for $x \in [0, 1)$. Note that $s = -\ln(1 - |x|^p) \Leftrightarrow |x| = \sqrt[p]{1 - \exp(1 - s)}$ for $s \in (0, \infty)$. Then one has $-g_1^{-1}(s) = g_2^{-1}(s) = \sqrt[p]{1 - \exp(-t)}$. Consequently, we have

$$\tilde{\varphi}_2(t) = \int_0^t \max\{(-g_1^{-1})'(s), (g_2^{-1})'(s)\}ds = \sqrt[p]{1 - \exp(-t)}.$$

(iii) Define $h_1(x) = h(x)$ for $x \in (-\pi/2, 0]$ and $h_2(x) = h(x)$ for $x \in [0, \pi/2)$. For $s \in (0, \infty)$, one has $-h_1^{-1}(s) = h_2^{-1}(s) = \sqrt[p]{\arctan(s)}$ because $s = \tan(|x|^p) \Leftrightarrow |x| = \sqrt[p]{\arctan(s)}$. It then follows that

$$\tilde{\varphi}_3(t) = \int_0^t \max\{(-h_1^{-1})'(s), (h_2^{-1})'(s)\}ds = \sqrt[p]{\arctan(t)},$$

which completes the proof. □

Remark 7.10. Recall from Example 4.15 that the above functions satisfy the KL property at 0 with $\varphi(t) = p \cdot t^{1/p}$. According to Proposition 7.3 we have $\tilde{\varphi}_i(t) \leq \varphi(t)$ on $[0, \eta_i)$ for each $i \in \{1, 2, 3\}$.

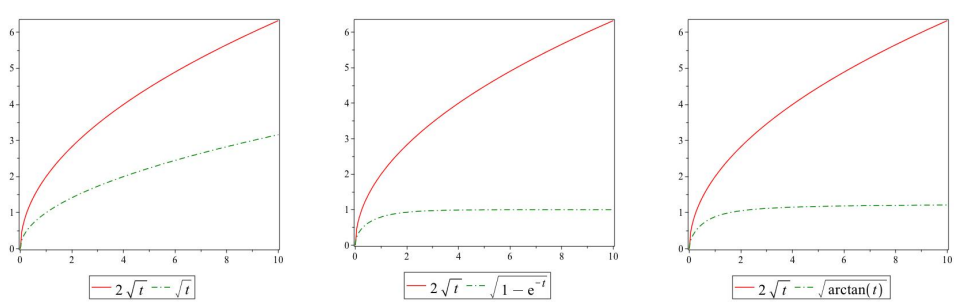

Figure 7.1: Compare the exact modulus $\tilde{\varphi}_i$ to $\varphi(t)$ when $p = 2$, see Example 7.9.

Now we work towards showing that there exists $f : \mathbb{R} \to \mathbb{R}$ who has the generalized KL property at 0 but the desingularizing function cannot have the form $\varphi(t) = c \cdot t^{1-\theta}$.

Lemma 7.11. *Let $\varphi : [0, \infty) \to \mathbb{R}_+$ be such that $\varphi(0) = 0$ and let $\theta \in [0, 1)$. The following statements are equivalent:*

(i) There exist $c > 0$ and $\eta \in (0, \infty]$ such that $\varphi(t) \leq c \cdot t^{1-\theta}, \forall t \in [0, \eta)$.

(ii) $\limsup_{t \to 0^+} \varphi(t)/t^{1-\theta} < \infty$.

Proof. "$(i) \Rightarrow (ii)$": It is easy to see $\limsup_{t \to 0^+} \phi(t)/t^{1-\theta} \leq c < \infty$.

"$(ii) \Rightarrow (i)$": Let $\varepsilon > 0$. Then there exists $\eta > 0$ such that

$$\limsup_{t \to 0^+} \frac{\varphi(t)}{t^{1-\theta}} = \inf_{\eta > 0} \sup_{0 < t < \eta} \frac{\varphi(t)}{t^{1-\theta}} > \sup_{0 < t < \eta} \frac{\varphi(t)}{t^{1-\theta}} - \varepsilon \geq \frac{\varphi(t)}{t^{1-\theta}} - \varepsilon, \forall 0 < t < \eta.$$

Set $c = \limsup_{t \to 0^+} \varphi(t)/t^{1-\theta} + \varepsilon < \infty$. Then one has $\varphi(t) < c \cdot t^{1-\theta}$ for every $t \in (0, \eta)$. □

Corollary 7.12. *Let $\varphi : [0, \infty) \to \mathbb{R}_+$ be such that $\varphi(0) = 0$ and let $\theta \in [0, 1)$. The following statements are equivalent:*

(i) For every $c > 0$ and $\eta \in (0, \infty]$, there exists $t \in (0, \eta)$ such that $\varphi(t) > c \cdot t^{1-\theta}$.

(ii) $\limsup_{t \to 0^+} \varphi(t)/t^{1-\theta} = \infty$.

Proof. Apply Lemma 7.11. □

Proposition 7.13. *Define $f(x)$ by $f(x) = e^{-1/x^2}$ for $x \neq 0$ and $f(0) = 0$. Then the following statements hold:*

(i) The exact modulus of the generalized KL property of f at 0 with respect to $U = (-\sqrt{2/3}, \sqrt{2/3})$ and $\eta = e^{-3/2}$ is $\tilde{\varphi}(t) = \sqrt{-1/\ln(t)}$ for $t > 0$ and $\tilde{\varphi}(0) = 0$.

(ii) For every $c > 0$ and $\theta \in [0, 1)$, the function $\varphi(t) = c \cdot t^{1-\theta}, 0 \leq t < \infty$, cannot be a desingularizing function of the generalized KL property of f at 0 with respect to any neighborhood $U \ni 0$ and $\eta \in (0, \infty]$.

Proof. (i) Apply Example 5.11 and Proposition 7.3.

(ii) Suppose to the contrary that there were $c > 0$ and $\theta \in [0, 1)$ such that f has the generalized KL property at 0 with respect to some $U \ni 0$ and $\eta > 0$ and $\varphi(t) = c \cdot t^{1-\theta}$. Taking the intersection if necessary, assume without loss

of generality that $U \cap [0 < f < \eta] \subseteq (-\sqrt{2/3}, \sqrt{2/3}) \cap [0 < f < e^{-3/2}]$. Hence by using Proposition 7.8, the exact modulus of f with respect to U and η is $\tilde{\varphi}(t)$. Furthermore, Proposition 7.3 implies that

$$\tilde{\varphi}(t) \leq \varphi(t) = c \cdot t^{1-\theta}, \forall t \in (0, \min\{\eta, e^{-3/2}\}). \tag{7.5}$$

Let $s > 0$. Then one has $s = \tilde{\varphi}(t) \Leftrightarrow t = e^{-1/s^2}$, which further implies that

$$\limsup_{t \to 0^+} \frac{\varphi(t)}{t^{1-\theta}} = \limsup_{s \to 0^+} \frac{s}{e^{-(1-\theta)/s^2}} = \limsup_{s \to 0^+} \frac{e^{(1-\theta)/s^2}}{s^{-1}} = \infty,$$

which by Corollary 7.12 contradicts to (7.5). □

The next example shows that the exact modulus may have an unusual form.

Example 7.14. (The Lambert $\mathcal{W}$ function as exact modulus) Let $f(x) = x^2 - \ln(1 - x^2)$. Then the exact modulus of f at 0 with respect to $U = (-1, 1)$ and $\eta = \infty$ is $\tilde{\varphi}(t) = \sqrt{1 - \mathcal{W}(e^{1-t})}$, where $\mathcal{W}$ denotes the Lambert $\mathcal{W}$ function.

Proof. Define $f_1(x) = f(x)$ for $x \in (-1, 0)$ and $f_2(x) = f(x)$ for $x \in (0, 1)$. For $s \in (0, \eta)$, $s = x^2 - \ln(1 - x^2) \Leftrightarrow e^{-x^2}(1 - x^2) = e^{-s} \Leftrightarrow e^{1-x^2}(1 - x^2) = e^{1-s}$. Note that $e^{1-s} > 0$ for $s > 0$. Hence by using Fact 2.11(ii), one has $1 - x^2 = \mathcal{W}(e^{1-s}) \Leftrightarrow |x| = \sqrt{1 - \mathcal{W}(e^{1-s})}$, from which one concludes that $-f_1^{-1}(s) = f_2^{-1}(s) = \sqrt{1 - \mathcal{W}(e^{1-s})}$. By applying Proposition 7.8, we have $\tilde{\varphi}(t) = \int_0^t \max\{(-f_1^{-1})'(s), (f_2^{-1})'(s)\} ds = \int_0^t (f_2^{-1})'(s) ds = f_2^{-1}(t) = \sqrt{1 - \mathcal{W}(e^{1-t})}$. □

Remark 7.15. Note that $\operatorname{argmin} f = \{0\}$ and $x^2 - \ln(1 - x^2) \geq 2x^2 = 2\operatorname{dist}(x, \operatorname{argmin} f)^2$ for every $x \in \mathbb{R}$. Therefore f satisfies (4.3) with $r = c = 2$. By invoking Fact 4.13, we conclude that f has the KL property at 0 with respect to $U = \mathbb{R}$, $\eta = \infty$ and $\varphi(t) = \sqrt{2t}$. By Proposition 7.3, we have $\tilde{\varphi}(t) \leq \varphi(t)$, see Figure 7.2.

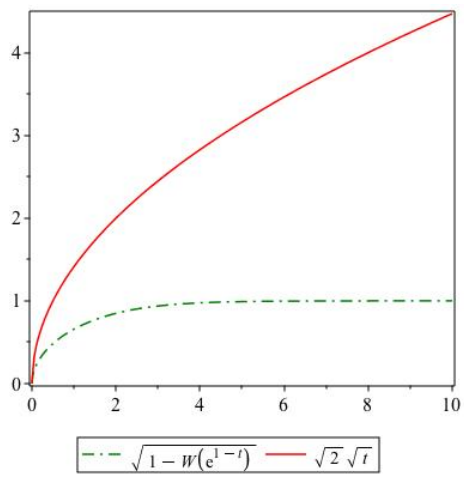

Figure 7.2: Compare the exact modulus $\tilde{\varphi}(t) = \sqrt{1-\mathcal{W}(e^{1-t})}$ to $\varphi(t) = \sqrt{2t}$, see Remark 7.15.

7.2 Computing the exact modulus

Unlike the one-dimensional case, it is usually difficult to compute directly the exact modulus of the generalized KL property for multi-variable functions. In this section, we provide several ways to determine or at least estimate the exact modulus for functions defined on $\mathbb{R}^n$, beginning with a formula for the exact modulus under scalar multiplication.

Proposition 7.16. *Let $\varepsilon_1 > 0$ and $\eta \in (0, \infty]$. Let $a \in \mathbb{R}\backslash\{0\}$ and $b \in \mathbb{R}^n$. Suppose that $f(x) = g(a^{-1}x + b)$, where $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper and lsc. If $\tilde{\varphi}_1(t)$ is the exact modulus of the generalized KL property of g at $\bar{y} = a^{-1}\bar{x}+b$ with respect to $U = \mathbb{B}(\bar{y}; \varepsilon_1)$ and $\eta > 0$, then $\tilde{\varphi}_2(t) = |a|\tilde{\varphi}_1(t)$ is the exact modulus of the generalized KL property of f at $\bar{x}$ with respect to $U = \mathbb{B}(\bar{x}; \varepsilon_2)$, where $\varepsilon_2 = |a|\varepsilon_1$, and $\eta > 0$.*

Proof. Suppose that $\tilde{\varphi}_1(t) = \int_0^t h_1(s)ds$, where $h_1(s) = \sup\{\mathrm{dist}^{-1}(0, \partial g(y)) : y \in K_1(s)\}$ and $K_1(s) = \{y \in \mathbb{R} : \mathbb{B}(\bar{y}; \varepsilon_1) \cap [0 < g - g(\bar{y}) < \eta], s \leq g(y) - g(\bar{y})\}$. Note that $g(y) = f(a(y-b))$. Hence by applying Fact 3.19, one gets $\partial g(y) = a \cdot \partial f(a(y-b))$. Take $v \in \partial g(y)$ with $\|v\| = \mathrm{dist}(0, \partial g(y))$. Then one has

$$\mathrm{dist}(0, \partial g(y)) = \|v\| = \|au\| = |a|\,\|u\| \geq |a|\,\mathrm{dist}(0, \partial f(a(y-b))),$$

for some $u \in \partial f(a(y-b))$ satisfying $v = au$. Similarly, by choosing $u \in \partial f(a(y-b))$ with $\|u\| = \operatorname{dist}(0, \partial f(a(y-b)))$ we have for some $v \in \partial g(y)$ with $v = au$

$$\operatorname{dist}(0, \partial f(a(y-b))) = \|u\| = |a|^{-1} \|v\| \geq |a|^{-1} \operatorname{dist}(0, \partial g(y)),$$

Hence we conclude that

$$\operatorname{dist}(0, \partial g(y)) = |a| \operatorname{dist}(0, \partial f(a(y-b))). \tag{7.6}$$

Define for $s \in (0, \eta)$, $h_2(s) = \sup\{\operatorname{dist}^{-1}(0, \partial f(x)) : x \in K_2(s)\}$, where $K_2(s) = \{x \in \mathbb{R} : x \in \mathbb{B}(\bar{x}; \varepsilon_2) \cap [0 < f - f(\bar{x}) < \eta], s \leq f(x) - f(\bar{x})\}$. Observe that $g(y) - g(\bar{y}) = f(a(y-b)) - f(a(\bar{y}-b)) = f(a(y-b)) - f(\bar{x})$ and

$$\|y - \bar{y}\| < \varepsilon_1 \Leftrightarrow \|a(y-b) - \bar{x}\| = \|ay - (\bar{x} + ab)\| = |a| \|y - \bar{y}\| < |a|\varepsilon_1 = \varepsilon_2.$$

Then one concludes that

$$y \in K_1(s) \Leftrightarrow a(y-b) \in K_2(s). \tag{7.7}$$

By definition, $h_2(s) \geq \operatorname{dist}^{-1}(0, \partial f(x))$ for every $x \in K_2(s)$. Taking $y \in K_1(s)$, the equivalence (7.7) ensures that there exists some $x \in K_2(s)$ such that $x = a(y-b)$. Therefore we have for every $y \in K_1(s)$

$$\begin{aligned} h_2(s) &\geq \frac{1}{\operatorname{dist}(0, \partial f(a(y-b)))} = |a| \frac{1}{|a| \operatorname{dist}(0, \partial f(a(y-b)))} \\ &= |a| \frac{1}{\operatorname{dist}(0, \partial g(y))}, \end{aligned}$$

where the last equality follows from (7.6), which means $h_2(s) \geq |a| h_1(s)$ and $\tilde{\varphi}_2(t) \geq |a| \tilde{\varphi}_1(t)$.

On the other hand, applying Corollary 6.13 shows $\tilde{\varphi}_2(t) \leq |a| \tilde{\varphi}_1(t)$. □

Next we provide a counterexample, showing that the sum of exact modulus is not necessarily the exact modulus of the sum.

Example 7.17. (Failure of the sum rule) Let $f_1(x) = x^2$ and $f_2(x) = -\ln(1 - x^2)$. Then f_1 and f_2 admit $\tilde{\varphi}_1(t) = \sqrt{t}$ and $\tilde{\varphi}_2(t) = \sqrt{1 - e^{-t}}$ as their exact modulus of the generalized KL property at 0 with respect to $U_i = \operatorname{dom} f_i$ and $\eta_i = \infty$, respectively. Define $f(x) = f_1(x) + f_2(x)$. Then the exact modulus of f at 0 with respect to $U = U_1 \cap U_2 = (-1, 1)$ and $\eta = \infty$ is $\tilde{\varphi}(t) = \sqrt{1 - \mathcal{W}(e^{1-t})}$, where $\mathcal{W}$ denotes the Lambert $\mathcal{W}$ function. Clearly, $\tilde{\varphi}(t) \neq \tilde{\varphi}_1(t) + \tilde{\varphi}_2(t)$.

Proof. See Example 7.14 and Example 7.9. □

With calculus rules of the generalized KL property in mind, we can estimate the exact modulus following the divide-and-conquer strategy below:

Step 1: Represent the given function f by functions $f_i, i \in I$, where I is an index set, whose exact modulus $\tilde{\varphi}_i$ is easier to be computed.

Step 2: Apply calculus rules of the generalized KL property to obtain a desingularizing function for f, denoted by φ, which is determined by $\tilde{\varphi}_i$.

Step 3: Invoke Proposition 7.3 to get an upper bound for the exact modulus of the given function, namely $\tilde{\varphi}(t) \leq \varphi(t)$, where $\tilde{\varphi}$ is the exact modulus of f.

Using the above methodology, we now provide several ways to estimate the exact modulus.

Proposition 7.18. *Let $n_i \in \mathbb{N}, i = 1, \ldots, m$, and let $n = \sum_{i=1}^m n_i$. For each i, let $f_i : \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$ be a proper and lsc function that is continuous on $\operatorname{dom} \partial f_i$. Furthermore, suppose that each f_i admits $\tilde{\varphi}_i : [0, \eta_i) \to \mathbb{R}_+$ as the exact modulus of the generalized KL property at $\bar{x}_i \in \operatorname{dom} \partial f_i$ with respect to $U_i = \mathbb{B}(\bar{x}_i; \varepsilon_i)$ for $\varepsilon_i > 0$ and $\eta_i > 0$. Let $\varepsilon = \min \varepsilon_i$ and let η be a real number with $\eta < m \cdot \min \eta_i$. If the function $t \mapsto (\tilde{\varphi}_i)'_-(t)$ is invertible for each $i \in \{1, \ldots, m\}$, then the exact modulus of the separable sum $f(x) = \sum_{i=1}^m f_i(x_i)$ at $\bar{x}$ with respect to $U = \mathbb{B}(\bar{x}; \varepsilon)$ and η satisfies*

$$\tilde{\varphi}(t) \leq \int_0^t \max_{1 \leq i \leq m} \left[(\tilde{\varphi}_i)'_- \left(\frac{s}{m} \right) \right] ds, \forall t \in (0, \eta].$$

Proof. Apply Theorem 6.7 and Proposition 7.3. □

Proposition 7.19. *Suppose that $f(x) = (g \circ F)(x)$, where $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is proper lsc and $F : \mathbb{R}^n \to \mathbb{R}^m$ is a smooth map. Let $\bar{x} \in \operatorname{dom} \partial f$ and let $\varepsilon_0 > 0$. Assume that g admits $\tilde{\varphi}_0$ as the exact modulus of the generalized KL property at $F(\bar{x})$ with respect to $U_0 = \mathbb{B}(F(\bar{x}); \varepsilon_0)$ and $\eta > 0$. If the Jacobian $\nabla F(\bar{x})$ has rank m, then there exist $\alpha > 0$ and $\varepsilon_1 > 0$ such that (6.13) holds. Furthermore, there exits $\varepsilon \in (0, \min\{\varepsilon_0, \varepsilon_1\}]$ such that f has the generalized KL property at $\bar{x}$ with respect to $U_1 = \mathbb{B}(\bar{x}; \varepsilon)$, $\eta > 0$ and $\alpha \cdot \tilde{\varphi}_0$. The exact modulus of the generalized KL property of f at $\bar{x}$ with respect to U_1 and η satisfies $\tilde{\varphi}(t) \leq \alpha \cdot \tilde{\varphi}_0$.*

Proof. Apply Theorem 6.11 and Proposition 7.3. □

Proposition 7.20. *Suppose that $f(x) = g(Ax - b)$, where $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is proper lsc, $A \in \mathbb{R}^{m \times n}$ has rank m and $b \in \mathbb{R}^m$. Let $\bar{x} \in \operatorname{dom} \partial f$ and $\varepsilon > 0$.*

Set $r = \sqrt{\lambda_{\min}(AA^*)} > 0$. If $\tilde{\varphi}_0$ is the exact modulus $\tilde{\varphi}$ of the generalized KL property of g at $A\bar{x} - b$ with respect to $U_1 = \mathbb{B}(A\bar{x} - b; \varepsilon)$ and $\eta > 0$, then the exact modulus of f at $\bar{x}$ with respect to $U_2 = \mathbb{B}(\bar{x}; \varepsilon / \|A\|)$ and $\eta > 0$ satisfies $\tilde{\varphi}(t) \leq \tilde{\varphi}/r$. Note that if $U_1 = \mathbb{R}^m$, then $U_2 = \mathbb{R}^n$.

Proof. Apply Corollary 6.13 and Proposition 7.3. □

Example 7.21. Let $p > 1$ and $\lambda > 0$. Suppose that $A \in \mathbb{R}^{m \times n}$ has rank m and $b \in \mathbb{R}^m$. The following statements hold:

(i) Define $f(x) = \lambda \|x\|_p^p$ for $x \in \mathbb{R}^n$, where $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$. Then the exact modulus of f at $\bar{x}$ with respect to $U_1 = \mathbb{R}^n$ and $\eta_1 = \infty$, denoted by $\tilde{\varphi}_1$, satisfies $\tilde{\varphi}_1(t) \leq \lambda^{-\frac{1}{p}} n^{1-\frac{1}{p}} \cdot t^{\frac{1}{p}}$.

(ii) Define $g(x) = \lambda \|Ax - b\|_p^p$ for $x \in \mathbb{R}^n$. Let $\bar{x} \in \mathbb{R}^n$ obey $A\bar{x} - b = 0$ and set $r = \sqrt{\lambda_{\min}(AA^*)}$. Then the exact modulus of g at $\bar{x}$ with respect to $U_2 = \mathbb{R}^n$ and $\eta_2 = \infty$, denoted by $\tilde{\varphi}_2$, satisfies $\tilde{\varphi}_2(t) \leq \frac{1}{r} \lambda^{-\frac{1}{p}} m^{1-\frac{1}{p}} t^{\frac{1}{p}}$.

(iii) Define $h(x) = \sum_{i=1}^n -\ln(\lambda - x_i)$ for $x = (x_1, \ldots, x_n)$ with $x_i < \lambda$ for every i. Then the exact modulus of h at $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)$ with respect to $U_3 = \mathbb{R}^n$ and $\eta_3 = \infty$, denoted by $\tilde{\varphi}_3$, satisfies $\varphi_3(t) \leq n(\lambda - \bar{x}_*)(1 - e^{-\frac{t}{n}})$, where $\bar{x}_* = \min_i \bar{x}_i$.

Proof. (i) Define $f_i(x_i) = \lambda |x_i|^p$. Recall from Example 7.9 that the function $x_i \mapsto |x_i|^p$ admits the exact modulus $t \mapsto t^{1/p}$ at 0 with respect to $U_1 = \mathbb{R}$ and $\eta_1 = \infty$ for each i. Applying Proposition 7.16 shows that f_i admits $\tilde{\varphi}_i(t) = \lambda^{-1/p} t^{1/p}$ as the exact modulus of the generalized KL property at 0 with respect to $U = \mathbb{R}$ and $\eta = \infty$, because $f_i(x_i) = |\lambda^{1/p} x_i|^p$. Note that $p > 1$. Hence the function $\tilde{\varphi}_i(t)$ is strictly concave. It then follows from Proposition 7.18 that

$$\tilde{\varphi}_1(t) \leq \int_0^t \max_{1 \leq i \leq n} \tilde{\varphi}_i'\left(\frac{s}{n}\right) ds = \lambda^{-\frac{1}{p}} \int_0^t \frac{1}{p} \left(\frac{s}{n}\right)^{\frac{1}{p}-1} ds = \lambda^{-\frac{1}{p}} n^{1-\frac{1}{p}} t^{\frac{1}{p}}.$$

(ii) Define $F : \mathbb{R}^m \to \mathbb{R}$ by $F(y) = \lambda \|y\|_p^p$. We learn from (i) that the exact modulus of F at $\bar{y} = A\bar{x} - b = 0$ with respect to $U_1 = \mathbb{R}^m$ and $\eta_1 = \infty$ satisfies $\varphi_1(t) \leq \lambda^{-\frac{1}{p}} m^{1-\frac{1}{p}} t^{\frac{1}{p}}$. Invoking Proposition 7.20, one concludes that the exact modulus of $g(x) = F(Ax - b)$ at $\bar{x}$ with respect to $U_2 = \mathbb{R}^n$ and $\eta_2 = \infty$ satisfies

$$\varphi_2(t) \leq \frac{1}{r} \tilde{\varphi}_1(t) \leq \frac{1}{r} \lambda^{-\frac{1}{p}} n^{1-\frac{1}{p}} t^{\frac{1}{p}}.$$

(iii) Define $h_i(x_i) = -\ln(\lambda - x_i)$ for $x_i < \lambda$ for each i. Then by Example 7.4 we conclude that h_i admits $\tilde{\varphi}_i(t) = (\lambda - \bar{x}_i)(1 - e^{-t})$ as the exact modulus of the generalized KL property at $\bar{x}_i$ with respect to $U_3 = \mathbb{R}$ and $\eta_3 = \infty$ for each i. Note that the derivative $(\tilde{\varphi}_i)'_-(t) = e^{-t}(\lambda - \bar{x}_i)$ is strictly decreasing hence invertible. By using Proposition 7.18, we conclude that the exact modulus of h at $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)$ with respect to $U_3 = \mathbb{R}^n$ and $\eta_3 = \infty$ satisfies

$$\tilde{\varphi}_3(t) \leq \int_0^t \max_{1 \leq i \leq n} (\lambda - \bar{x}_i) e^{-\frac{s}{n}} ds = \int_0^t (\lambda - \bar{x}_*) e^{-\frac{s}{n}} ds = n(\lambda - \bar{x}_*)(1 - e^{-\frac{t}{n}}),$$

where the second equality holds because $\max_{1 \leq i \leq n}(\lambda - \bar{x}_i) = \lambda - \bar{x}_*$, where $\bar{x}_* = \min_{1 \leq i \leq n} \bar{x}_i$. □

Remark 7.22. (i) Recall a result by Bolte et al. [10, Corollary 9]: Piecewise polynomial convex functions on $\mathbb{R}^n$ with degree m admit KL exponent

$$\theta = 1 - \frac{1}{1 + (m-1)^n},$$

which implies that f and g considered in Example 7.21 have KL exponent $1 - 1/\left(1 + (p-1)^n\right)$ if $p \in \mathbb{N}$. Let us observe a *limitation* of their result. Suppose that $p > 2$. Then it is easy to see that

$$\theta = 1 - \frac{1}{1 + (p-1)^n} \to 1, n \to \infty,$$

which means that the formula above becomes an *overestimation* of the KL exponent when the dimension of data is high. In contrast, the upper bounds of exact modulus in Example 7.21 imply that f and g admit KL exponent $1 - 1/p$, which is *dimension-independent*, where p is a real number.

(ii) Note that $h(x)$ considered in Example 7.21 has no stationary points. Hence according to Fact 4.4 one can show that h has desingularizing function $\varphi_\theta = t^{1-\theta}/(\varepsilon(1-\theta))$ for every $\theta \in [0, 1)$ with respect to $U = \mathbb{B}(\bar{x}; \varepsilon)$ and $\eta = \varepsilon$ for some $\varepsilon \in (0, 1]$ sufficiently small, which means the KL inequality (4.1) is only valid for $x \in \mathbb{B}(\bar{x}; \varepsilon) \cap [0 < h - h(\bar{x}) < \varepsilon]$, while our result takes all $x \in \mathbb{R}^n \cap [0 < h - h(\bar{x}) < \infty]$ into account.

7.3 Uniformizing the generalized KL property over a compact set

In this section, we show that the generalized KL property can be uniformized over a compact set, a technique that plays a key role in Chapter 8.

To this end, it is convenient to introduce the uniform versions of the generalized KL property and its exact modulus.

Definition 7.23. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper lsc and let $\mu \in \mathbb{R}$. Let $V \subseteq \operatorname{dom} \partial f$ and suppose that $f(x) = \mu$ for all $x \in V$.

(i) We say that f has *the uniform generalized KL property* on V with respect to $U \supset V$, $\eta > 0$ and $\varphi \in \Phi_\eta$, if for every $x \in U \cap [0 < f - \mu < \eta]$

$$\varphi'_-(f(x) - \mu) \cdot \operatorname{dist}(0, \partial f(x)) \geq 1. \tag{7.8}$$

(ii) Define $h : (0, \eta) \to \mathbb{R}$ by

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

Then *the exact modulus of the uniform generalized KL property* on V with respect to U and η is the function $\tilde{\varphi} : [0, \eta) \to \mathbb{R}_+$,

$$t \mapsto \int_0^t h(s)ds, \forall t \in (0, \eta),$$

and $\tilde{\varphi}(0) = 0$. If $U \cap [0 < f - \mu < \eta] = \emptyset$, then we set $\tilde{\varphi}(t) \equiv 0$.

Remark 7.24. (i) Note that the uniform generalized KL property on V with respect to U and η reduces to the generalized KL property at $\bar{x}$ with respect to U and η, when $V = \{\bar{x}\}$ for some $\bar{x} \in \operatorname{dom} \partial f$. Similarly, the exact modulus of the uniform generalized KL property on V reduces to the exact modulus of the generalized KL property at $\bar{x}$ if $V = \{\bar{x}\}$.

(ii) Suppose that f is constant on a compact set Ω. We will show that the generalized KL property at each $\bar{x} \in \Omega$ implies the uniform generalized KL property on Ω, see Proposition 7.27.

The next proposition shows that the exact modulus of the uniform generalized KL property is the smallest possible function in Φ_η such that (7.8) holds. Note that the proof below is similar to Proposition 7.3. A proof in detail is provided below for the sake of completeness.

Proposition 7.25. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc and let $\mu \in \mathbb{R}$. Let $V \subseteq \operatorname{dom} \partial f$ be a nonempty set on which $f(x) = \mu$ for all $x \in V$. Furthermore, let $U \supset V$ be a nonempty set, $\eta \in (0, \infty]$ and let $\varphi \in \Phi_\eta$. Suppose that f has the uniform generalized KL property on V with respect to U, η and φ. Then the exact modulus of the uniform generalized KL property $\tilde{\varphi} : [0, \eta) \to \overline{\mathbb{R}}$ exists and satisfies*

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

Moreover, the function f has the uniform generalized KL property on V with respect to U, η and $\tilde{\varphi}$.

Proof. Let us show first that $\tilde{\varphi}(t) \leq \varphi(t)$ on $[0, \eta)$, which implies that $\tilde{\varphi}(t) < \infty$. If $U \cap [0 < f - \mu < \eta] = \emptyset$ then clearly $\tilde{\varphi}(t) = 0 \leq \varphi(t)$ for every $t \in (0, \eta)$ according to the convention. Therefore we proceed the proof assuming that $U \cap [0 < f - \mu < \eta] \neq \emptyset$. Fix $s \in (0, \eta)$ and recall from Lemma 6.1(i) that φ'_- is decreasing. Then for $x \in U \cap [0 < f - \mu < \eta]$ with $s \leq f(x) - \mu$ we have

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

Taking the supremum over all such x yields

$$h(s) \leq \varphi'_-(s),$$

where $h(s) = \sup\{\operatorname{dist}^{-1}(0, \partial f(x)) : x \in U \cap [0 < f - \mu < \eta], s \leq f(x) - \mu\}$, which implies that for $t \in (0, \eta)$,

$$\tilde{\varphi}(t) = \int_0^t h(s)ds \leq \int_0^t \varphi'_-(s)ds = \varphi(t) < \infty.$$

Note that the last equality is implied by Lemma 6.1.

Let $0 < s_1 \leq s_2 < \eta$. Then for $x \in U \cap [0 < f - \mu < \eta]$

$$s_2 \leq f(x) - \mu \Rightarrow s_1 \leq f(x) - \mu,$$

meaning that $h(s_2) \geq h(s_1)$. Hence $h(s)$ is a positive-valued decreasing function. By applying Fact 3.15, one concludes that $\tilde{\varphi}(t) \in \Phi_\eta$ and $\tilde{\varphi}'_-(t) \geq h(t)$. Consequently, for $x \in U \cap [0 < f - \mu < \eta]$ with $t = f(x) - \mu$ we have

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

which completes the proof. □

For nonempty subset $U \subseteq \mathbb{R}^n$ and $\varepsilon \in (0, \infty]$, define

$$U_\varepsilon = \{x \in \mathbb{R}^n : \operatorname{dist}(x, U) < \varepsilon\}.$$

Note that $U_\infty = \mathbb{R}^n$ and $U_\varepsilon = \mathbb{B}(\bar{x}; \varepsilon)$ if $U = \{\bar{x}\}$ for some $\bar{x} \in \mathbb{R}^n$. The following technical lemma is known as the Lebesgue number lemma [19, Theorem 55].

Lemma 7.26. *Let $U \subseteq \mathbb{R}^n$ be a nonempty compact subset. Suppose that $\{U_i\}_{i=1}^p$ is a finite open cover of U. Then there exists $\varepsilon > 0$, which is called the Lebesgue number of U, such that*

$$U \subseteq U_\varepsilon \subseteq \bigcup_{i=1}^p U_i.$$

Proof. We claim that there exists $\varepsilon > 0$ such that for every $x \in U$, $\mathbb{B}(x;\varepsilon) \subseteq U_{i_0}$ for some i_0. Otherwise, for each $n \in \mathbb{N}$ we would have a sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \in U$ such that

$$\mathbb{B}(x_n; 1/n) \not\subseteq U_i \tag{7.9}$$

for all $i \in \{1, \dots, p\}$. By the compactness, taking subsequence if necessary, assume without loss of generality that $x_n \to \bar{x} \in U$. Then there exists an open set U_{i_0} such that $\bar{x} \in U_{i_0}$ and consequently $\mathbb{B}(\bar{x};\delta) \subseteq U_{i_0}$ for some $\delta > 0$. Take $N \in \mathbb{N}$ sufficiently large so that $x_N \in \mathbb{B}(\bar{x};\delta/2)$ and $1/N < \delta/2$. Hence by the triangle inequality, one has $\|x - \bar{x}\| \le \|x_N - x\| + \|x_N - \bar{x}\| \le 1/N + \delta/2 < \delta$ for every $x \in \mathbb{B}(x_N; 1/N)$. Consequently, we have

$$\mathbb{B}(x_N; 1/N) \subseteq \mathbb{B}(\bar{x};\delta) \subseteq U_{i_0},$$

which is a contradiction to (7.9).

Take $y \in U_\varepsilon$. Then there exists $x \in U$ such that $\|x - y\| = \operatorname{dist}(y, U) < \varepsilon$, meaning that $y \in \mathbb{B}(\bar{x};\varepsilon) \subseteq U_{i_0}$ for some i_0. Hence one concludes that $U_\varepsilon \subseteq \bigcup_{i=1}^p U_i$. □

We are now ready for the following proposition, which connects the generalized KL property on a compact set to its uniform counterpart.

Proposition 7.27. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc. Let $\Omega \subseteq \operatorname{dom} \partial f$ be a nonempty compact set on which $f(x) = \mu$ for some $\mu \in \mathbb{R}$. Suppose that f satisfies the generalized KL property at each $x \in \Omega$. Then the following statements are true:*

(i) There exist $\varepsilon > 0, \eta \in (0, \infty]$ and $\varphi(t) \in \Phi_\eta$ such that f has the uniform generalized KL property on Ω with respect to $U = \Omega_\varepsilon$, η and φ, i.e., for every $x \in \Omega_\varepsilon \cap [0 < f - \mu < \eta]$,

$$\varphi'_-(f(x) - \mu) \cdot \operatorname{dist}(0, \partial f(x)) \ge 1. \tag{7.10}$$

(ii) Denote by $\tilde{\varphi}(t)$ the exact modulus of the uniform generalized KL property of f on Ω with respect to $U = \Omega_\varepsilon$ and η. Then for every $t \in [0, \eta)$,

$$\tilde{\varphi}(t) \le \varphi(t) < \infty.$$

Moreover, the function f has the uniform generalized KL property on $V = \Omega$ with respect to $U = \Omega_\varepsilon$, η and $\tilde{\varphi}$.

Proof. (i) For each $x \in \Omega$, there exist $\varepsilon = \varepsilon(x) > 0$, $\eta = \eta(x) \in (0, \infty]$ and $\varphi(t) = \varphi_x(t) \in \Phi_\eta$ 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 $\Omega \subseteq \bigcup_{x\in\Omega} \mathbb{B}(x;\varepsilon)$. By the compactness, there exist $x_1, \ldots, x_p \in \Omega$ 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

$$(\varphi_i)'_-(f(x) - \mu) \cdot \operatorname{dist}(0, \partial f(x)) \geq 1. \tag{7.11}$$

Define $\varphi(t) = \sum_{i=1}^{p} \varphi_i(t)$ and $\eta = \min_{1\leq i\leq p} \eta_i$. It is easy to see that φ belongs to Φ_η. By using Lemma 7.26, 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 $f(x_i) = \mu$ and $\eta \leq \eta_i$ 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

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

where the first inequality holds because $(\varphi_i)'_-(t) > 0$, see Lemma 6.1(i); the last one is implied by (7.11).

(ii) Apply Proposition 7.25. □

Chapter 8

Applications: The PALM algorithm revisited

In this chapter, we revisit the celebrated PALM algorithm and investigate the algorithmic impact of the exact modulus of the generalized KL property. Our analysis focuses on two aspects:

- Let $(z_k)_{k\in\mathbb{N}}$ be a sequence generated by the PALM algorithm. By assuming that the objective function is KL, Bolte et al. [9] showed that $(z_k)_{k\in\mathbb{N}}$ has the *finite length property*, i.e.,

$$\sum_{k=1}^{\infty} \|z_{k+1} - z_k\| < \infty,$$

  and the sequence $(z_k)_{k\in\mathbb{N}}$ converges to a critical point of the objective function. We will show that the exact modulus of the generalized KL property provides *the sharpest upper bound* for $\sum_{k=1}^{\infty} \|z_{k+1} - z_k\|$, which is an improvement of [9, Theorem 1].

- Recall from Proposition 4.22 that the KL exponent of the objective function provides an estimation of the convergence rate of certain algorithm of interest. Taking the PALM algorithm as an example, we will explore the interplay between the exact modulus of the generalized KL property and the convergence rate of PALM.

The structure of Chapter 8 is as follows: The PALM algorithm is introduced in Section 8.1. In Section 8.2, we collect basic properties of PALM. The sharpest upper bound of $\sum_{k=1}^{\infty} \|z_{k+1} - z_k\|$ is given in Section 8.3. Finally, we study the interplay between the convergence rate of the PALM algorithm and the exact modulus of generalized KL property in Section 8.4.

8.1 An introduction to the PALM algorithm

Consider the following nonconvex and nonsmooth optimization problem:

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

where $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ are proper and lsc functions and $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is C^1. We need the following blanket assumptions:

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

(A2) For every fixed y, the function $x \mapsto F(x,y)$ is $C^{1,1}_{L_1(y)}$, i.e.,

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

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

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

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

(A4) ∇F is Lipschitz continuous on bounded subsets of $\mathbb{R}^n \times \mathbb{R}^m$, i.e., 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 $(x_i, y_i) \in B_1 \times B_2$, $i = 1,2$,

$$\begin{aligned}&\|(\nabla_x F(x_1,y_1), \nabla_y F(x_1,y_1)) - (\nabla_x F(x_2,y_2), \nabla_y F(x_2,y_2))\|\\ &\le M\,\|(x_1 - x_2, y_1 - y_2)\|\,.\end{aligned}$$

Bolte et al. [9] proposed the following algorithm to solve the aforementioned 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:
   2.1. Take $\gamma_1 > 1$, set $c_k = \gamma_1 L_1(y_k)$ and compute

$$x_{k+1} \in \operatorname{prox}^f_{c_k}\big(x_k - \tfrac{1}{c_k}\nabla_x F(x_k, y_k)\big). \tag{8.1}$$

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

$$y_{k+1} \in \operatorname{prox}^g_{d_k}\big(y_k - \tfrac{1}{d_k}\nabla_y F(x_{k+1}, y_k)\big). \tag{8.2}$$

Remark 8.1. Let $g(y) \equiv 0$ and $F(x,y) = F(x)$. Then one has $\Psi(x,y) = \Psi(x) = f(x)+F(x)$. Assume in addition that f and F are convex. Then the PALM algorithm reduces to the well-known *proximal gradient method* (PGM), in which case (8.1) and (8.2) reduce to

$$x_{k+1} = \operatorname{prox}_{c_k}^f \left(x_k - \tfrac{1}{c_k}\nabla F(x_k)\right),\ y_{k+1} = \operatorname{prox}_{d_k}^g(y_k) = y_k.$$

8.2 Basic properties

In this section, we collect several nice properties of the PALM algorithm that will be used in the sequel. Note that these results can be found in [9]. We provided proof in detail for these results for the sake of self-containedness.

Lemma 8.2. *Let $h : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz smooth with modulus $L_h > 0$, and let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and lsc with $\inf_{\mathbb{R}^n} f > -\infty$. Fix $t > L_h$. Then for every $u \in \operatorname{dom} f$ and $u^+ \in \mathbb{R}^n$ given by*

$$u^+ \in \operatorname{prox}_t^f \left(u - \tfrac{1}{t}\nabla h(u)\right),$$

we have

$$h(u^+) + f(u^+) \le h(u) + f(u) - \frac{t - L_h}{2}\left\|u^+ - u\right\|^2. \tag{8.3}$$

Proof. By Fact 3.28, u^+ is well-defined and

$$u^+ \in \operatorname{argmin}_{v\in\mathbb{R}^n}\left\{\langle v-u, \nabla h(u)\rangle + \frac{t}{2}\|v-u\|^2 + f(v)\right\}.$$

Hence by taking $v = u$, we have $\langle u^+ - u, \nabla h(u)\rangle + \frac{t}{2}\left\|u^+ - u\right\|^2 + f(u^+) \le f(u)$. Equivalently,

$$f(u^+) \le f(u) - \langle u^+ - u, \nabla h(u)\rangle - \frac{t}{2}\left\|u^+ - u\right\|^2. \tag{8.4}$$

Invoking Fact 3.12 gives

$$h(u^+) \le h(u) + \langle \nabla h(u), u^+ - u\rangle + \frac{L_h}{2}\left\|u^+ - u\right\|^2.$$

It then follows that

$$\begin{aligned} f(u^+) + h(u^+) &\le f(u^+) + h(u) + \langle \nabla h(u), u^+ - u\rangle + \frac{L_h}{2}\left\|u^+ - u\right\| \\ &\le f(u) + h(u) - \frac{t - L_h}{2}\left\|u^+ - u\right\|^2, \end{aligned}$$

where the second inequality is implied by (8.4). □

Lemma 8.3. *Suppose that (A1)-(A3) 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

$$\frac{\rho_1}{2}\left\|z_{k+1}-z_k\right\|^2 \leq \Psi(z_k)-\Psi(z_{k+1}), \forall k \geq 0, \tag{8.5}$$

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

(ii) $\sum_{k=1}^{\infty}\left\|z_{k+1}-z_k\right\|^2 < \infty$, and hence $\lim_{k\to\infty}\left\|z_{k+1}-z_k\right\| = 0$.

Proof. (i) Let $k \geq 0$. Recall from (8.1) that $(x_k)_{k\in\mathbb{N}}$ satisfies

$$x_{k+1} \in \operatorname{prox}_{c_k}^f\left(x_k - \tfrac{1}{c_k}\nabla_x F(x_k, y_k)\right),$$

and according to assumptions (A2) and (A3), the map $x \mapsto \nabla F(x, y_k)$ is Lipschitz continuous with modulus $L_1(y_k) \geq \lambda_1^-$. Hence by applying Lemma 8.2 to $x \mapsto \nabla F(x, y_k)$ with $t = c_k = \gamma_1 L_1(y_k) > L_1(y_k)$, one has

$$\begin{aligned}
&F(x_{k+1}, y_k) + f(x_{k+1}) \leq F(x_k, y_k) + f(x_k) - \frac{c_k - L_1(y_k)}{2}\left\|x_{k+1}-x_k\right\|^2 \\
&= F(x_k, y_k) + f(x_k) - \frac{1}{2}(\gamma_1-1)L_1(y_k)\left\|x_{k+1}-x_k\right\|^2.
\end{aligned}$$

Similarly, invoking Lemma 8.2 to $y \mapsto \nabla F(x_{k+1}, y)$ with $t = d_{k+1} = \gamma_2 L_2(x_{k+1}) > L_2(x_{k+1})$, one has

$$\begin{aligned}
&F(x_{k+1}, y_{k+1}) + g(y_{k+1}) \leq F(x_{k+1}, y_k) + g(y_k) \\
&- \frac{d_{k+1} - L_2(x_{k+1})}{2}\left\|y_{k+1}-y_k\right\|^2 \\
&= F(x_{k+1}, y_k) + g(y_k) - \frac{1}{2}(\gamma_2-1)L_2(x_{k+1})\left\|y_{k+1}-y_k\right\|^2.
\end{aligned}$$

Adding above inequalities together, one gets

$$\begin{aligned}
\Psi(z_k) - \Psi(z_{k+1}) &\geq \frac{1}{2}(\gamma_1-1)L_1(y_k)\left\|x_{k+1}-x_k\right\|^2 \\
&\quad + \frac{1}{2}(\gamma_2-1)L_2(x_{k+1})\left\|y_{k+1}-y_k\right\|^2 \\
&\geq \frac{1}{2}(\gamma_1-1)\lambda_1^-\left\|x_{k+1}-x_k\right\|^2 + \frac{1}{2}(\gamma_2-1)\lambda_2^-\left\|y_{k+1}-y_k\right\|^2 \\
&\geq \frac{\rho_1}{2}\left\|z_{k+1}-z_k\right\|^2 \geq 0,
\end{aligned}$$

where the second inequality is implied by the assumption (A3) that $L_1(x_k) \geq \lambda_1^-$ and $L_2(y_k) \geq \lambda_2^-$.

(ii) Fix $N \in \mathbb{N}$. Adding (8.5) from $k = 0$ to $k = N - 1$ one obtains

$$\begin{aligned}\frac{\rho_1}{2}\sum_{k=0}^{N-1}\|z_{k+1} - z_k\|^2 &\leq \sum_{k=0}^{N-1}[\Psi(z_k) - \Psi(z_{k+1})] \\ = \Psi(z_0) - \Psi(z_N) &\leq \Psi(z_0) - \inf_{\mathbb{R}^n}\Psi < \infty,\end{aligned}$$

which by taking $N \to \infty$ further implies that $\sum_{k=0}^{\infty}\|z_{k+1} - z_k\|^2 < \infty$ and $\lim_{k\to\infty}\|z_{k+1} - z_k\| = 0$. □

Lemma 8.4. *Suppose that (A1)-(A4) hold. 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*

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

Then $(A_x^k, A_y^k) \in \partial\Psi(x_k, y_k)$. Let $M > 0$ be the Lipschitz constant given in (A4). Then we have

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

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

Proof. Recall from (8.1) that $x_k \in \operatorname{prox}_{c_{k-1}}^f\big(x_{k-1} - \nabla_x F(x_{k-1}, y_{k-1})/c_{k-1}\big)$. Applying Fact 3.28 one gets

$$x^k \in \operatorname{argmin}_{x\in\mathbb{R}^n}\Big\{\langle x - x_{k-1}, \nabla_x F(x_{k-1}, y_{k-1})\rangle + \tfrac{c_{k-1}}{2}\|x - x_{k-1}\|^2 + f(x)\Big\},$$

which by Fact 3.25 further implies that $0 \in \nabla_x F(x_{k-1}, y_{k-1}) + c_{k-1}(x_k - x_{k-1}) + \partial f(x_k)$. Hence for some $u_k \in \partial f(x_k)$, one has

$$u_k = c_k(x_{k-1} - x_k) - \nabla_x F(x_{k-1}, y_{k-1}).$$

Similarly, we have for some $v_k \in \partial g(y_k)$

$$v_k = d_{k-1}(y_{k-1} - y_k) - \nabla_y F(x_k, y_{k-1}).$$

Hence invoking Fact 3.18(iv) gives

$$\Big(A_x^k, A_y^k\Big) = (u_k + \nabla_x F(x_k, y_k), v_k + \nabla_y F(x_k, y_k)) \in \partial\Psi(x_k, y_k).$$

Recall from (A4) that $\nabla F(x,y)$ is Lipschitz continuous on every bounded subsets of $\mathbb{R}^n \times \mathbb{R}^m$ with modulus $M > 0$. Hence we have

$$\begin{aligned}\left\|A_x^k\right\| &\leq c_{k-1}\left\|x_{k-1} - x_k\right\| + \left\|\nabla_x F(x_k, y_k) - \nabla_x F(x_{k-1}, y_{k-1})\right\| \\ &\leq \left\|(\nabla_x F(x_k, y_k) - \nabla_x F(x_{k-1}, y_{k-1}), \nabla_y F(x_k, y_k) - \nabla_y F(x_{k-1}, y_{k-1}))\right\| \\ &+ c_{k-1}\left\|x_{k-1} - x_k\right\| \\ &\leq c_{k-1}\left\|x_{k-1} - x_k\right\| + M\left\|(x_k - x_{k-1}, y_k - y_{k-1})\right\| \\ &\leq (c_{k-1} + M)\left\|x_{k-1} - x_k\right\| + M\left\|z_k - z_{k-1}\right\|.\end{aligned}$$

Moreover, according to (A3), we have $c_{k-1} = \gamma_1 L_1(y_{k-1}) \leq \gamma_1 \lambda_1^+$, which means that

$$\left\|A_x^k\right\| \leq (M + \gamma_1\lambda_1^+)\left\|x_k - x_{k-1}\right\| + M\left\|z_k - z_{k-1}\right\| \leq (2M + \rho_2)\left\|z_k - z_{k-1}\right\|.$$

A similar argument shows that

$$\begin{aligned}\left\|A_y^k\right\| &\leq d_{k-1}\left\|y_k - y_{k-1}\right\| + \left\|\nabla_y F(x_k, y_k) - \nabla_y F(x_k, y_{k-1})\right\| \\ &\leq d_{k-1}\left\|y_k - y_{k-1}\right\| \leq 2\rho_2\left\|z_k - z_{k-1}\right\|.\end{aligned}$$

Altogether, we conclude that

$$\left\|(A_x^k, A_y^k)\right\| \leq \left\|A_x^k\right\| + \left\|A_y^k\right\| \leq (2M + 3\rho_2)\left\|z_k - z_{k-1}\right\|,$$

as claimed. □

Remark 8.5. Several sufficient conditions for $(z_k)_{k\in\mathbb{N}}$ to be bounded are discussed in [1, Remark 5]. For example, $(z_k)_{k\in\mathbb{N}}$ is bounded, if f and g have compact lower-level sets and $F(x,y) = \frac{1}{2}\left\|x - y\right\|^2$.

Let $(z_k)_{k\in\mathbb{N}}$ be a sequence generated by PALM with starting point z_0. The set of limit points is denoted by $\omega(z_0)$, i.e.,

$$\omega(z_0) = \{z \in \mathbb{R}^n \times \mathbb{R}^m : \exists (z_{k_q})_{q\in\mathbb{N}} \subseteq (z_k)_{k\in\mathbb{N}}, z_{k_q} \to z, \text{ as } q \to \infty\}.$$

Lemma 8.6. *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^*, one has

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

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

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

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

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

Proof. (i) Let $z^* = (x^*, y^*) \in \omega(z_0)$ and suppose that $\left(z_{k_q}\right)_{n\in\mathbb{N}}$ converges to z^* as $q \to \infty$, where $z_{k_q} = (x_{k_q}, y_{k_q})$ for each $q \in \mathbb{N}$. Then by lower semi-continuity, we conclude that $f(x^*) \leq \liminf_{q\to\infty} f(x_{k_q})$ and $g(y^*) \leq \liminf_{q\to\infty} g(y_{k_q})$. On the other hand, by using (8.1) and Fact 3.28, one gets

$$x_{k+1} \in \operatorname{argmin}_{x\in\mathbb{R}^n} \left\{\langle x - x_k, \nabla_x F(x_k, y_k)\rangle + \frac{c_k}{2}\|x - x_k\|^2 + f(x)\right\}.$$

Then taking $x = x^*$ gives

$$\begin{aligned}
&\langle x_{k+1} - x_k, \nabla_x F(x_k, y_k)\rangle + \frac{c_k}{2}\|x_{k+1} - x_k\|^2 + f(x_{k+1}) \\
&\leq \langle x^* - x_k, \nabla_x F(x_k, y_k)\rangle + \frac{c_k}{2}\|x^* - x_k\|^2 + f(x^*).
\end{aligned}$$

Choosing $k = k_q - 1$ and rearranging the inequality above, we have

$$\begin{aligned}
f(x_{k_q}) &\leq \langle x^* - x_{k_q}, \nabla_x F(x_{k_q-1}, y_{k_q-1})\rangle + \frac{c_{k_q-1}}{2}\left\|x^* - x_{k_q-1}\right\|^2 \\
&\quad - \frac{c_{k_q-1}}{2}\left\|x_{k_q} - x_{k_q-1}\right\|^2 + f(x^*) \\
&\leq \left\|x^* - x_{k_q}\right\| \left\|\nabla_x F(x_{k_q-1}, y_{k_q-1})\right\| + \frac{c_{k_q-1}}{2}\left\|x^* - x_{k_q-1}\right\|^2 \\
&\quad - \frac{c_{k_q-1}}{2}\left\|x_{k_q} - x_{k_q-1}\right\|^2 + f(x^*),
\end{aligned}$$

which by Lemma 8.3(ii) and the assumption $x_{k_q} \to x^*$ implies that

$$\limsup_{q\to\infty} f(x_{k_q}) \leq f(x^*).$$

Hence we conclude that $\lim_{q\to\infty} f(x_{k_q}) = f(x^*)$. A similar argument shows that $\lim_{q\to\infty} g(y_{k_q}) = g(y^*)$. Altogether, one concludes that $\Psi(z_{k_q}) \to \Psi(z^*)$ as $q \to \infty$.

Invoking Lemma 8.4, we have

$$\left\|(A_x^{k_q}, A_y^{k_q})\right\| \leq (2M + 3\rho_2)\left\|z_{k_q} - z_{k_q-1}\right\| \to 0,\ q \to \infty.$$

Therefore we conclude that $0 \in \partial\Psi(z^*)$ by using Fact 3.24.

(ii) For every subsequence $(z_{k_q})_{q\in\mathbb{N}}$, taking subsequence if necessary, assume that $z_{k_q} \to z^*$ for some $z^* \in \omega(z_0)$. Then $\operatorname{dist}(z_{k_q}, \omega(z_0)) \leq \left\|z_{k_q} - z^*\right\| \leq 0$ as $q \to \infty$.

(iii) Clearly $\omega(z_0)$ is nonempty as $(z_k)_{k\in\mathbb{N}}$ is assumed to be bounded. We claim that

$$\omega(z_0) = \cap_{n\in\mathbb{N}} \operatorname{cl} \cup_{k>n} \{z_k\},$$

which implies that $\omega(z_0)$ is closed and bounded, hence it is compact. Now we prove the claim. Let $z^* \in \omega(z_0)$ and suppose that $z_{k_q} \to z^*$ as $q \to \infty$. Equivalently, for every $\varepsilon > 0$, there exists $q_\varepsilon \in \mathbb{N}$ such that $\left\|z^* - z_{k_q}\right\| < \varepsilon$ whenever $q > q_\varepsilon$. Fix $n \in \mathbb{N}$ and assume that $q_\varepsilon > n$ for each $\varepsilon > 0$. Then for every ε, there exists $q_\varepsilon > n$ such that for $q > q_\varepsilon$, $\left\|z_{k_q} - z^*\right\| < \varepsilon$, meaning that

$$\mathbb{B}(z^*; \varepsilon) \cap (\cup_{k>n}\{z_k\}) \neq \emptyset.$$

Hence $z^* \in \operatorname{cl} \cup_{k>n}\{z_k\}$ and $z^* \in \cap_{n\in\mathbb{N}} \operatorname{cl} \cup_{k>n}\{z_k\}$. On the other hand, take $z^* \in \cap_{n\in\mathbb{N}} \operatorname{cl} \cup_{k>n}\{z_k\}$. Then for every n and ε there exists $k > n$ such that $\|z_k - z^*\| < \varepsilon$, which means $z_k \to z^*$.

Now we prove that $\omega(z_0)$ is connected. Suppose to the contrary that there were nonempty closed disjoints A and B in $\mathbb{R}^n \times \mathbb{R}^m$ such that $\omega(z_0) = A \cup B$. Define $\gamma : \mathbb{R}^n \times \mathbb{R}^m :\to \mathbb{R}$ by

$$\gamma(z) = \frac{\operatorname{dist}(z, A)}{\operatorname{dist}(z, A) + \operatorname{dist}(z, B)}.$$

By [21, Example 1.20], γ is well-defined and continuous on $\mathbb{R}^n \times \mathbb{R}^m$. Clearly $A = [\gamma = 0]$ and $B = [\gamma = 1]$. Let $U = [\gamma < 1/4]$ and $V = [\gamma > 3/4]$. Then U and V are open sets containing A and B, respectively. We claim that there exists $k_0 \in \mathbb{N}$ such that either $z_k \in U$ or $z_k \in V$ for $k > k_0$. Otherwise we would have a subsequence $(z_{k_q})_{q\in\mathbb{N}}$ with $z_{k_q} \in (U \cup V)^c$. Suppose that $z_{k_q} \to z^*$. Then by the openness of $U \cup V$, one gets $z^* \notin \omega(z_0)$, which is absurd. Set $r_k = \gamma(z_k)$ for $k \in \mathbb{N}$. We claim that the following statements hold true:

(a) $r_k \notin [1/4, 3/4]$ for all $k > k_0$.

(b) There are infinitely many k such that $r_k < 1/4$.

(c) There are infinitely many k such that $r_k > 3/4$.

(d) $|r_{k+1} - r_k| \to 0$ as $k \to \infty$.

The claim (a) follows immediately from our previous analysis. Statements (b) and (c) hold as $A, B \neq \emptyset$. The claim (d) follows from the fact that γ is uniformly continuous on the compact set $R \cdot \mathbb{B}_{\mathbb{R}^n \times \mathbb{R}^m}$, where $R = \sup_{k \in \mathbb{N}} \|z_k\|$ is finite by assumption, and $\|z_{k+1} - z_k\| \to 0$ as $k \to \infty$. Observe that (a)-(d) are mutually exclusive. Indeed, by using (a)-(c), we can take a sequence $(r_k)_{k \in \mathbb{N}}$ such that $r_k < 1/4$ for even k and $r_k > 3/4$ for odd k. Then one has $|r_{2n} - r_{2n+1}| \geq |r_{2n+1}| - |r_{2n}| \geq 3/4 - 1/4 = 1/2, \forall n \in \mathbb{N}$, contradicting to (d).

(iv) Note that $(\Psi(z_k))_{k \in \mathbb{N}}$ is a decreasing sequence with $\inf_{k \in \mathbb{N}} \Psi(z_k) \geq \inf \Psi > -\infty$. Hence $\lim_{k \to \infty} \Psi(z_k) = l$. Take $z^* \in \omega(z_0)$ and assume that $z_{k_q} \to z^*$. From assertion (i), we have $\Psi(z^*) = \lim_{q \to \infty} \Psi(z_{k_q}) = \lim_{k \to \infty} \Psi(z_k) = l$, which completes the proof. □

Remark 8.7. It is worth noting that statement(iii) can be proved alternatively, by using Lemma 8.3 and a result by Ostrowski (see [18, Theorem 26.1]), which asserts that for a bounded sequence $(x_k) \subseteq \mathbb{R}^n$ with $x_{k+1} - x_k \to 0$, the set of limit points is compact and connected.

8.3 Finite length property

In this section, we obtain the sharpest upper bound for the sum of consecutive iterates gap $\sum_{k=1}^{\infty} \|z_{k+1} - z_k\|$, which improves a result by Bolte et al. [9, Theorem 1].

Theorem 8.8. *Suppose that the objective function Ψ is a generalized 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 $(z_k)_{k \in \mathbb{N}}$ has finite length. To be specific, there exist $l \in \mathbb{N}$, $\eta \in (0, \infty]$ and $\tilde{\varphi} \in \Phi_\eta$, such that for $p \geq l + 1$ and every $q \in \mathbb{N}$

$$\sum_{k=p}^{p+q} \|z_{k+1} - z_k\| \leq C \cdot \tilde{\varphi}(\Psi(z_p) - \Psi(z^*)) + \|z_p - z_{p-1}\| . \tag{8.6}$$

Therefore

$$\sum_{k=1}^{\infty} \|z_{k+1} - z_k\| \leq A + C \cdot \tilde{\varphi}(\Psi(z_{l+1}) - \Psi(z^*)) < \infty, \tag{8.7}$$

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

(ii) The sequence $(z_k)_{k\in\mathbb{N}}$ converges to a critical point $z^* = (x^*, y^*)$ of objective function Ψ.

Proof. (i) Suppose that $z_{k_q} \to z^* \in \omega(z_0)$. Then Lemma 8.6(i) implies that $\lim_{q\to\infty} \Psi(z_{k_q}) = \Psi(z^*)$. Note that $(\Psi(z_k))_{k\in\mathbb{N}}$ is a decreasing sequence with $\inf_{k\in\mathbb{N}} \Psi(z_k) > -\infty$ according to (A1). Hence one has $\lim_{k\to\infty} \Psi(z_k) = \lim_{q\to\infty} \Psi(z_{k_q}) = \Psi(z^*)$. Now we continue the proof 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 (8.5). 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 8.6 and assumption, Ψ is a generalized KL function that is constant on compact set $\omega(z_0)$. Invoking Lemma 7.27 shows that there exist $\varepsilon > 0$ and $\eta > 0$ such that the exact modulus of the uniform generalized 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 $z \in \Omega_\varepsilon \cap [0 < \Psi - \Psi(z^*) < \eta]$,

$$\tilde{\varphi}'_-(\Psi(z) - \Psi(z^*)) \cdot \operatorname{dist}(0, \partial\Psi(z)) \geq 1. \tag{8.8}$$

By the fact that $\lim_{k\to\infty} \Psi(z_k) = \Psi(z^*)$, there exists some $l_1 > 0$ such that $0 < \Psi(z_k) - \Psi(z^*) < \eta$ for $k > l_1$. On the other hand, Lemma 8.6(ii) shows that there exists $l_2 > 0$ such that $\operatorname{dist}(z_k, \omega(z_0)) < \varepsilon$ for $k > l_2$. Altogether, we conclude that for $k > l$, where $l = \max\{l_1, l_2\}$, $z_k \in \Omega_\varepsilon \cap [0 < \Psi - \Psi(z^*) < \eta]$ and

$$\tilde{\varphi}'_-(\Psi(z_k) - \Psi(z^*)) \cdot \operatorname{dist}(0, \partial\Psi(z_k)) \geq 1. \tag{8.9}$$

It follows from Lemma 8.4 that $\operatorname{dist}(0, \partial\Psi(z_k)) \leq \left\|\left(A_x^k, A_y^k\right)\right\| \leq (2M + 3\rho_2)\|z_k - z_{k-1}\|$. Hence one has from (8.9) that for $k > l$,

$$\tilde{\varphi}'_-(\Psi(z_k) - \Psi(z^*)) \geq \operatorname{dist}^{-1}(0, \partial\Psi(z_k)) \geq \frac{1}{2M + 3\rho_2}\|z_k - z_{k-1}\|^{-1}. \tag{8.10}$$

Note that $\|z_k - z_{k-1}\| \neq 0$. Otherwise Lemma 8.4 would imply that

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

which contradicts to (8.9). Applying Lemma 6.1(ii) to $\tilde{\varphi}$ with $s = \Psi(z_{k+1}) -$

$\Psi(z^*)$ and $t = \Psi(z_k) - \Psi(z^*)$, one obtains for $k > l$

$$\begin{aligned}\frac{\tilde{\varphi}(\Psi(z_k) - \Psi(z^*)) - \tilde{\varphi}(\Psi(z_{k+1}) - \Psi(z^*))}{\Psi(z_k) - \Psi(z_{k+1})} &\geq \tilde{\varphi}'_-(\Psi(z^k) - \Psi(z^*)) \\ &\geq \frac{1}{2M + 3\rho_2} \|z_k - z_{k-1}\|^{-1}.\end{aligned} \tag{8.11}$$

For the sake of simplicity, we introduce the following shorthand:

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

Then (8.11) can be rewritten as

$$\Psi(z_k) - \Psi(z_{k+1}) \leq \|z_k - z_{k-1}\| \cdot \Delta_{k,k+1} \cdot (2M + 3\rho_2). \tag{8.12}$$

Furthermore, Lemma 8.3(i) gives

$$\|z_{k+1} - z_k\|^2 \leq \frac{2}{\rho_1} [\Psi(z_k) - \Psi(z_{k+1})] \leq C\Delta_{k,k+1} \|z_k - z_{k-1}\|, \tag{8.13}$$

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

$$2\|z_{k+1} - z_k\| \leq C\Delta_{k,k+1} + \|z_k - z_{k-1}\|. \tag{8.14}$$

Let $p \geq l + 1$. For every $q \in \mathbb{N}$, summing up the above inequality from p up 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 one has 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 (8.6). By taking $q \to \infty$, one has

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

from which (8.7) readily follows by setting $p = l + 1$.

(ii) 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 $k \to \infty$. Hence invoking (8.6), one concludes 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. The rest of the proof follows from Lemma 8.6(i), where we showed that $\omega(z_0) \subseteq \text{crit } \Psi$. □

Remark 8.9. (i) We claim that (8.7) is the sharpest upper bound for the sum of consecutive iterates gap. Assuming that the objective function Ψ is KL, Bolte et al. [9, Theorem 1] showed that

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

where $\varphi(t)$ is a desingularizing function for the uniform KL property of Ψ on $\Omega = \omega(z_0)$ with respect to Ω_ε and $\eta > 0$, whose existence is given in [9, Lemma 6]. According to Proposition 7.25, we conclude that $\tilde{\varphi}(t)$ is the smallest among all $\varphi(t)$ given by [9, Lemma 6], which means that (8.7) is indeed the sharpest upper bound.

(ii) See [1, Remark 5] for conditions guaranteeing the boundedness of $(z_k)_{k\in\mathbb{N}}$.

8.4 The interplay between convergence rate and g-KL modulus

In this section, we study how the exact modulus of the generalized KL property affects the convergence rate of PALM algorithm. The following fact is given in [9, Remark 6], which is a specific version of Proposition 4.22.

Fact 8.10. *Let $(z_k)_{k\in\mathbb{N}}$ be a sequence generated by PALM. If the desingularizing function φ of the objective function Ψ is of the form $\varphi(t) = c \cdot t^{1-\theta}$ for $c > 0$ and $\theta \in [0, 1)$, then the following estimations hold:*

(i) If $\theta = 0$ then the sequence $(z_k)_{k\in\mathbb{N}}$ converges in finitely many steps.

(ii) If $\theta \in (0, 1/2]$ then there exists $\omega > 0$ and $\tau \in [0, 1)$ such that $\|z_k - z^*\| \leq \omega\tau^k$.

(iii) If $\theta \in (1/2, 1)$ then there exists $\omega > 0$ such that

$$\|z_k - z^*\| \leq \omega k^{-\frac{1-\theta}{2\theta-1}}.$$

Our goal is to improve the above fact by using the exact modulus of generalized KL property. To this end, we approximate the exact modulus $\tilde{\varphi}$ with functions of the form $t \mapsto mt^{1-\theta}$ for $\theta \in [0, 1)$ and $m > 0$. Define for $k \in \mathbb{N}$,

$$\sigma_k = \sum_{i=k}^{\infty} \|z_{i+1} - z_i\| < \infty,$$

where $(z_k)_{k\in\mathbb{N}}$ is a sequence generated by PALM. Note that $(\sigma_k)_{k\in\mathbb{N}}$ is a well-defined decreasing sequence with $\lim_{k\to\infty} \sigma_k = 0$ due to Theorem 8.8. Recall the following quantities:

$$\rho_1 = \min\{(\gamma_1 - 1)\lambda_1^-, (\gamma_2 - 1)\lambda_2^-\}, \rho_2 = \max\{\gamma_1\lambda_1^+, \gamma_2\lambda_2^+\},$$

where $\gamma_1, \gamma_2 > 1$ and $\lambda_i^{\pm}$ are given in (A3). Furthermore, recall that $M > 0$ is the Lipschitz constant given in (A4) and $C = 2(2M + 3\rho_2)/\rho_1$.

Theorem 8.11. *Let $\varphi(t) = m \cdot t^{1-\theta}$ for $\theta \in [0, 1)$ and $m > 0$. Let $\tilde{\varphi}$ be the exact modulus of the uniform generalized KL property of Ψ on $\Omega = \omega(z_0)$ with respect to $U = \Omega_\varepsilon$ for $\varepsilon > 0$ and $\eta > 0$. Let $\gamma = 2m(2M + 3\rho_2)^{\frac{1}{\theta}}\rho_1^{-1} > 0$ and let $R > 1$. Suppose that all assumptions in Theorem 8.8 hold. Furthermore, let $l \in \mathbb{N}$ and suppose that $z_k \in \Omega_\varepsilon \cap [0 < \Psi - \Psi(z^*) < \eta]$ for $k > l$. If $\tilde{\varphi}(t) \leq \alpha \cdot \varphi(t)$ for some $\alpha \in (0, 1]$ on $[0, \eta)$, then the following assertions hold:*

(i) If $\theta \in (0, \frac{1}{2}]$ then for $k \geq l + 1$

$$\|z_k - z^*\| \leq c_1 \cdot Q_1^{k-l-1},$$

where $Q_1 = \frac{\alpha^{\frac{1}{\theta}}\gamma+1}{\alpha^{\frac{1}{\theta}}\gamma+2} \in (0, 1)$ and $c_1 = \sigma_{l+1} > 0$.

(ii) *If* $\theta \in (\frac{1}{2}, 1)$ *then for* $k \geq l+1$

$$\|z_k - z^*\| \leq Q_2 \cdot k^{-\frac{1-\theta}{2\theta-1}},$$

where $Q_2 = (\frac{c_2}{l+1})^{-\frac{1-\theta}{2\theta-1}} > 0$ *and* $c_2 = \min\{(R^{\frac{2\theta-1}{\theta}} - 1)\sigma_l^{\frac{1-2\theta}{1-\theta}}, \frac{2\theta-1}{(1-\theta)R}(1+\alpha^{\frac{1}{\theta}}\gamma)^{-\frac{\theta}{1-\theta}}\} > 0$.

(iii) *If* $\theta = 0$ *then the sequence* $(z_k)_{k\in\mathbb{N}}$ *converges in finitely many steps. The number of iteration* $n_0 \in \mathbb{N}$ *satisfies*

$$n_0 \leq \alpha \cdot Q_3 + l,$$

where $Q_3 = m(2M + 3\rho_2)\sigma_{l+1} > 0$.

Proof. Assume without loss of generality that $\Psi(z^*) = 0$. Since $\|z_k - z^*\| = \lim_{i\to\infty} \|z_i - z_k\| \leq \sigma_k$ for each k, it suffices to find upper bound of σ_k.

For $t > 0$, we learn from Lemma 6.1(i) that

$$\tilde{\varphi}'_-(t) \leq \frac{\tilde{\varphi}(t)}{t} \leq \alpha \cdot mt^{-\theta}. \tag{8.15}$$

From now on, assume that $\theta > 0$. By using Lemma 8.6, there exists $l \in \mathbb{N}$ such that $z_k \in \Omega_\varepsilon \cap [0 < \Psi - \Psi(z^*) < \eta]$ for $n > l$ and hence

$$1 \leq \tilde{\varphi}'_-(\Psi(z_k)) \cdot \operatorname{dist}(0, \partial\Psi(z_k)) \leq \alpha \cdot m(\Psi(z_k))^{-\theta} \cdot \operatorname{dist}(0, \partial\Psi(z_k)),$$

where the second inequality is implied by (8.15). Applying Lemma 8.4, one obtains

$$\begin{aligned} 1 &\leq \alpha \cdot m(\Psi(z_k))^{-\theta} \cdot \left\|(A_x^k, A_y^k)\right\| \\ &\leq \alpha \cdot m(\Psi(z_k))^{-\theta} \cdot (2M + 3\rho_2)\|z_k - z_{k-1}\| \\ &= \alpha \cdot m(\Psi(z_k))^{-\theta} \cdot (2M + 3\rho_2)(\sigma_{k-1} - \sigma_k), \end{aligned}$$

which further implies that

$$\Psi(z_k) \leq [\alpha \cdot m(2M + 3\rho_2)(\sigma_{k-1} - \sigma_k)]^{\frac{1}{\theta}}. \tag{8.16}$$

On the other hand, (8.6) shows that for $k \geq l+1$, $\sigma_k = \sum_{i=k}^{\infty} \|z_{k+1} - z_k\| \leq C \cdot \tilde{\varphi}(\Psi(z_k)) + \|z_k - z_{k-1}\|$, where $C = 2(2M + 3\rho_2)/\rho_1$. Combining with (8.16), one gets

$$\begin{aligned} \sigma_k &\leq \alpha \cdot Cm(\Psi(z_k))^{1-\theta} + \sigma_{k-1} - \sigma_k \\ &\leq \alpha^{\frac{1}{\theta}} \cdot \gamma(\sigma_{k-1} - \sigma_k)^{\frac{1-\theta}{\theta}} + \sigma_{k-1} - \sigma_k, \end{aligned} \tag{8.17}$$

where the first inequality follows from the assumption that $\tilde{\varphi}(t) \leq \alpha \cdot m^{1-\theta}$ and the second one is implied by (8.16), where $\gamma = \frac{2m}{\rho_1}(2M + 3\rho_2)^{1/\theta}$. Assume without loss of generality that $\sigma_{k-1} - \sigma_k < 1$ for $k \geq l+1$ because $\lim_{k\to\infty} \|z_{k+1} - z_k\| = 0$. We are now ready to analyze the convergence rate.

(i): If $\theta \in (0, \frac{1}{2}]$, then $\frac{1-\theta}{\theta} \geq 1$ and $(\sigma_{k-1} - \sigma_k)^{\frac{1-\theta}{\theta}} \leq \sigma_{k-1} - \sigma_k$. Hence (8.17) implies that for $k \geq l+1$,

$$\sigma_k \leq \alpha^{\frac{1}{\theta}} \cdot \gamma(\sigma_{k-1} - \sigma_k) + (\sigma_{k-1} - \sigma_k) \Rightarrow \sigma_k \leq \frac{\alpha^{\frac{1}{\theta}}\gamma + 1}{\alpha^{\frac{1}{\theta}}\gamma + 2}\sigma_{k-1}.$$

Let $Q_1 = \frac{\alpha^{\frac{1}{\theta}}\gamma+1}{\alpha^{\frac{1}{\theta}}\gamma+2}$. Then we have for $k \geq l+1$

$$\sigma_k \leq Q_1^{k-l-1}\sigma_{l+1},$$

which proves the desired inequality.

Case 2: If $\theta \in (\frac{1}{2}, 1)$, then $0 < \frac{1-\theta}{\theta} < 1$ and $(\sigma_{k-1} - \sigma_k)^{\frac{1-\theta}{\theta}} \geq \sigma_{k-1} - \sigma_k$. It then follows from (8.17) that for $k \geq l+1$

$$\begin{aligned}\sigma_k &\leq \alpha^{\frac{1}{\theta}} \cdot \gamma(\sigma_{k-1} - \sigma_k)^{\frac{1-\theta}{\theta}} + (\sigma_{k-1} - \sigma_k)^{\frac{1-\theta}{\theta}} \\ &= (1 + \alpha^{\frac{1}{\theta}}\gamma)(\sigma_{k-1} - \sigma_k)^{\frac{1-\theta}{\theta}},\end{aligned}$$

which means

$$\sigma_k^{\frac{\theta}{1-\theta}} \leq (1 + \alpha^{\frac{1}{\theta}}\gamma)^{\frac{\theta}{1-\theta}}(\sigma_{k-1} - \sigma_k). \tag{8.18}$$

Define $H(t) = t^{-\frac{\theta}{1-\theta}}$. Then (8.18) can rewritten as

$$1 \leq (1 + \alpha^{\frac{1}{\theta}}\gamma)^{\frac{\theta}{1-\theta}}(\sigma_{k-1} - \sigma_k)H(\sigma_k).$$

Note that $H(t)$ is strictly decreasing as is the sequence $(\sigma_k)_{k\in\mathbb{N}}$. If $H(\sigma_k) \leq R \cdot H(\sigma_{k-1})$, where $R > 1$ is a fixed real number, then we have

$$\begin{aligned}1 &\leq (1 + \alpha^{\frac{1}{\theta}}\gamma)^{\frac{\theta}{1-\theta}}(\sigma_{k-1} - \sigma_k)H(\sigma_k) \leq R(1 + \alpha^{\frac{1}{\theta}}\gamma)^{\frac{\theta}{1-\theta}}(\sigma_{k-1} - \sigma_k)H(\sigma_{k-1}) \\ &\leq R(1 + \alpha^{\frac{1}{\theta}}\gamma)^{\frac{\theta}{1-\theta}} \int_{\sigma_k}^{\sigma_{k-1}} H(t)dt = R(1 + \alpha^{\frac{1}{\theta}}\gamma)^{\frac{\theta}{1-\theta}} \frac{1-\theta}{1-2\theta}(\sigma_{k-1}^{\frac{1-2\theta}{1-\theta}} - \sigma_k^{\frac{1-2\theta}{1-\theta}}).\end{aligned}$$

Set $v = \frac{1-2\theta}{1-\theta} < 0$. Then the inequality above implies that

$$\sigma_{k-1}^v - \sigma_k^v \geq \frac{2\theta - 1}{(1-\theta)R}(1 + \alpha^{\frac{1}{\theta}}\gamma)^{-\frac{\theta}{1-\theta}} > 0. \tag{8.19}$$

If $H(\sigma_k) > R \cdot H(\sigma_{k-1})$, which is equivalent to $\sigma_k^{-\frac{\theta}{1-\theta}} > R \cdot \sigma_{k-1}^{-\frac{\theta}{1-\theta}}$, one has $\sigma_k < q \cdot \sigma_{k-1}$, where $q = (\frac{1}{R})^{\frac{1-\theta}{\theta}} \in (0,1)$. Recall that $v < 0$. Hence we have

$$\sigma_k < q \cdot \sigma_{k-1} \Rightarrow \sigma_k^v > q^v \cdot \sigma_{k-1}^v \Leftrightarrow \sigma_k^v - \sigma_{k-1}^v > (q^v - 1) \cdot \sigma_{k-1}^v > 0.$$

Set $\mu = (q^v - 1)\sigma_l^v > 0$. Then we have for $k \geq l+1$

$$\sigma_k^v - \sigma_{k-1}^v \geq (q^v - 1)\sigma_{k-1}^v \geq \mu > 0, \tag{8.20}$$

where the second inequality holds because $v < 0$ and $\sigma_l \geq \sigma_{k-1}$ for $k \geq l+1$.

Define $c_2 = \min\{\mu, \frac{2\theta-1}{(1-\theta)R}(1+\alpha^{\frac{1}{\theta}}\gamma)^{-\frac{\theta}{1-\theta}}\} > 0$. Combining (8.19) and (8.20), we conclude that for $k \geq l+1$,

$$\sigma_k^v - \sigma_{k-1}^v \geq c_2 > 0, \tag{8.21}$$

Summing (8.21) from $l+1$ up to some $k > l+1$ gives $\sigma_k^v - \sigma_{l+1}^v \geq (k-l)c_2$. Furthermore, one has

$$\sigma_k^v \geq (k-l)c_2 + \sigma_{l+1}^v \geq (k-l)c_2 \geq \frac{k}{l+1}c_2,$$

where the last inequality holds because $k - l \geq \frac{k}{l+1}$ for $k \geq l+1$. Hence we have

$$\|z_k - z^*\| \leq \sigma_k \leq \left(\frac{k}{l+2}\right)^{\frac{1-\theta}{1-2\theta}} = Q_2 \cdot k^{-\frac{1-\theta}{2\theta-1}},$$

where $Q_2 = (\frac{c_2}{l+1})^{-\frac{1-\theta}{2\theta-1}}$.

Case 3: If $\theta = 0$ then (8.15) becomes $\tilde{\varphi}'_-(t) \leq \alpha \cdot m$ for $t \in (0, \eta)$. Set $I = \{k \in \mathbb{N} : z_{k+1} \neq z_k\}$ and suppose that I has infinite elements. Invoking Lemma 8.4, one has $(2M + 3\rho_2)\|z_{k+1} - z_k\| \geq \operatorname{dist}(0, \partial\Psi(z_{k+1}))$. Hence for $k \in I$ with $k > l$, we have

$$\alpha \cdot m(2M + 3\rho_2)\|z_{k+1} - z_k\| \geq \tilde{\varphi}'_-(\Psi(z_{k+1})) \cdot \operatorname{dist}(0, \partial\Psi(z_{k+1})) \geq 1. \tag{8.22}$$

Recall from Lemma 8.3 that $\lim_{k\to\infty}\|z_{k+1} - z_k\| = 0$. Then by taking $k \to \infty$, the above inequality yields $0 \geq 1$, which is absurd. Hence one

concludes that there exists n_0 such that $I = \{1, \ldots, n_0\}$ and $z_{k+1} = z_k$ for all $k > n_0$. If $n_0 > l$, the (8.22) implies that for $l + 1 \leq k \leq n_0$,

$$\|z_{k+1} - z_k\| \geq \frac{1}{\alpha \cdot m(2M + 3\rho_2)}. \tag{8.23}$$

which further implies that

$$\begin{aligned}\sigma_1 &= \sum_{k=1}^{l} \|z_{k+1} - z_k\| + \sum_{k=l+1}^{n_0} \|z_{k+1} - z_k\| \\ &\geq \frac{(n_0 - l)}{\alpha \cdot m(2M + 3\rho_2)} + \sum_{k=1}^{l} \|z_{k+1} - z_k\|.\end{aligned}$$

Hence by rearranging one has $n_0 \leq \alpha \cdot (2M + 3\rho_2)\sigma_{l+1} + l$. If $n_0 \leq l$, then the inequality $n_0 \leq \alpha \cdot (2M + 3\rho_2)\sigma_{l+1} + l$ holds trivially. □

Remark 8.12. (i) Lemma 7.11 guarantees that the exact modulus $\tilde{\varphi}$ satisfies $\tilde{\varphi}(t) \leq \alpha \cdot mt^{1-\theta}$ for some $\theta \in [0, 1)$ and $m > 0$ if and only if $\limsup_{t \to 0^+} \varphi(t)/t^\theta < \infty$.

(ii) When $\theta \in [0, 1/2]$, it is easy to see that the upper bounds provided in assertions (i) and (iii) decrease as one of α and θ decreases.

(iii) It is difficult to determine analytically the behavior of $Q_2 = Q_2(\alpha, \theta)$ given in statement (ii) when α or θ changes. By setting irrelevant quantities $R = l = \sigma_l = (2M + 3\rho_2) = 2$ and $2m\rho_1^{-1} = 1$, we obtain the following plots describing the behavior of the upper bound $Q_2 \cdot k^{-\frac{1-\theta}{2\theta-1}}$ when one of α and θ increases and another one is fixed, see Figure 8.1.

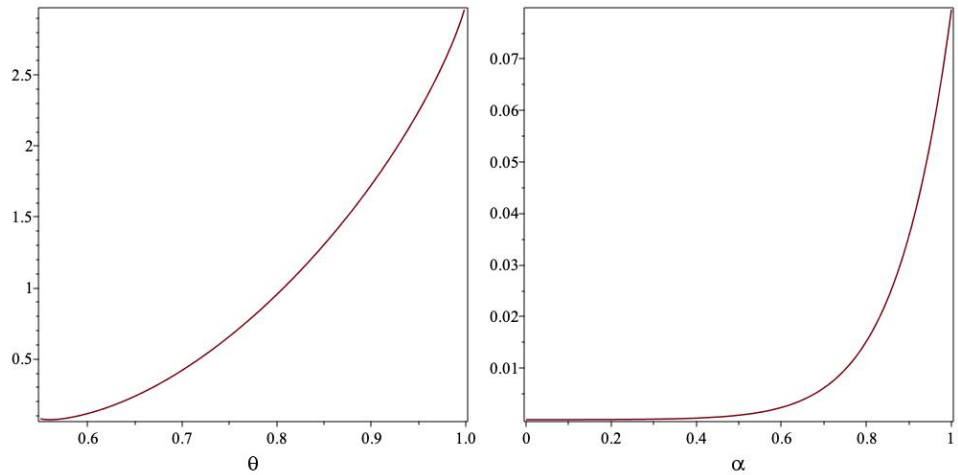

Figure 8.1: The upper bound $Q_2 \cdot k^{-\frac{1-\theta}{2\theta-1}}$, where $k = 1000$. We set $\alpha = 1$ in the left figure and $\theta = 0.55$ in the right one. See Theorem 8.11(ii).

A simple but illustrative one-dimension example is given below, from which we will learn that Theorem 8.11 provides a more accurate estimation of the convergence rate. Recall from Remark 8.1 that PALM reduces to PGM if $F(x, y) = F(x)$, $g(y) \equiv 0$, and F and f are both convex, in which case $\Psi(x, y) = \Psi(x) = f(x) + F(x)$ and iteration (8.1) reduces to

$$x_{k+1} = \operatorname{prox}_{c_k}^{f}\left(x_k - \tfrac{1}{c_k}\nabla F(x_k)\right),\ y_{k+1} = y_k. \tag{8.24}$$

Example 8.13. Consider the following one-dimensional optimization problem:

$$\min_{x\in\mathbb{R}} \Psi(x) = f(x) + F(x),$$

where $f(x) = |x|$ and $F(x) = x^2/2$. Let $\gamma_1 > 1$ and $x_0 \in \mathbb{R}$. Then the following statements hold:

(i) The exact modulus of the generalized KL property of Ψ at $x^* = 0$ with respect to $U = \mathbb{R}$ and $\eta = \infty$ is $\tilde{\varphi}(t) = \sqrt{2t+1} - 1$.

(ii) The objective function Ψ satisfies assumptions (A1)-(A4), and the sequence generated by (8.24) satisfies

$$x_{k+1} = \left[(1 - 1/\gamma_1)|x_k| - 1/\gamma_1\right]_+ \cdot \operatorname{sgn}(x_k), k \in \mathbb{N},$$

where $[u]_+ = \max\{0, u\}$ for $u \in \mathbb{R}$.

(iii) The sequence $(x_k)_{k\in\mathbb{N}}$ converges to $x^* = 0$ in n_0 steps, where $n_0 \leq (2+3\gamma_1)\sigma_2 + 1$.

Proof. (i) For $s > 0$, one has $s = |x| + x^2/2 \Leftrightarrow |x| = \sqrt{2s+1} - 1$. Define $\Psi_1(x) = \Psi(x)$ for $x \leq 0$ and $\Psi_2(x) = \Psi(x)$ for $x \geq 0$. Then we have $-\Psi_1^{-1}(s) = \Psi_2^{-1}(s) = \sqrt{2s+1} - 1$. Applying Proposition 7.8, one concludes that $\tilde{\varphi}(t) = \sqrt{2t+1} - 1$.

(ii) Clearly (A1) holds. As $F'(x) = x$, F is Lipschitz smooth with modulus 1, meaning that assumptions (A2)-(A4) hold with $L_1(y_k) \equiv 1$ and $M = \lambda_1^- = \lambda_1^+ = 1$. Note that $g(y) \equiv 0$. Then $\lambda_2^+ = \lambda_2^- = 0$. It follows that $\rho_1 = \max\{(\gamma_1 - 1)\lambda_1^-, (\gamma_2 - 1)\lambda_2^-\} = \gamma_1 - 1$ and $\rho_2 = \max\{\gamma_1\lambda_1^+, \gamma_2\lambda_2^+\} = \gamma_1\lambda_1^+ = \gamma_1$. Applying Example 3.29 and (8.24) yields

$$x_{k+1} = \operatorname{prox}_{\gamma_1}^f\left((1 - \tfrac{1}{\gamma_1})x_k\right) = \left[(1 - 1/\gamma_1)|x_k| - 1/\gamma_1\right]_+ \cdot \operatorname{sgn}(x_k).$$

(iii) By assertions (i) we have $\tilde{\varphi}'(t) = 1/\sqrt{2t+1} \leq 1$ for $t \in [0, \infty)$, implying that $\tilde{\varphi}(t) \leq \alpha \cdot mt^{1-\theta}$ with $\alpha = m = 1$ and $\theta = 0$. Invoking Theorem 8.11 yields that $(x_k)_{k\in\mathbb{N}}$ converges in finite steps. On the other hand, since $[0 < \Psi(x) - \Psi(x^*) < \infty] = \mathbb{R}\backslash\{0\}$, we have $x_k \in \mathbb{R}\backslash\{0\}$ for every $k > l = 1$. Hence by using Theorem 8.11. one concludes that $Q_3 = m(2M + 3\rho_2)\sigma_{l+1} = (2 + 3\gamma_1)\sigma_2$, and the number of iteration n_0 satisfies

$$n_0 \leq (2 + 3\gamma_1)\sigma_2 + 1,$$

as claimed. □

Remark 8.14. Note that Ψ is a piecewise polynomial convex function with degree 2. Hence Ψ has KL exponent $\theta = 1/2$ by [10, Corollary 9]. Then we learn from Fact 8.10 that $(x_k)_{k\in\mathbb{N}}$ converges in infinitely many steps and there exit $\omega > 0$ and $\tau \in [0, 1) > 0$ such that

$$\|x_k - x^*\| \leq \omega\tau^k.$$

In contrast, out estimation is more accurate and provides an explicit upper bound for the number of iterations.

Chapter 9

Conclusion and future research

We have provided new characterizations for the KL property by using tools from nonsmooth analysis, including formulas for the desingularizing functions of locally convex C^1 functions and polynomials on the line. These results provide a new prospective to understand the KL property because our formulas do not assume the usual form $\varphi(t) = c \cdot t^{1-\theta}$. On the other hand, by introducing the generalized KL property and its exact modulus, we extended the classic theory of the KL property and provided a complete answer to the question: "*What is the optimal desingularizing function?*", which fills a gap in the current literature. Moreover, our work showed that there exists a function $f : \mathbb{R} \to \mathbb{R}$ who has the KL property but the desingularizing function cannot take the form $\varphi(t) = c \cdot t^{1-\theta}$. We have also showed that the exact modulus has significant impact on optimization because it provides the sharpest upper bound for the trajectory of a sequence generated by PALM. To end this thesis, we provide several directions for the future work:

- Compute or at least estimate the exact modulus of the generalized KL property for concrete optimizations models.
- There is no general sum rule for the KL property in the current literature, therefore another direction is to provide a sum rule, which will fill a gap in this area. Furthermore, a sum rule will allow us to study the KL property of concrete optimization models, for instance estimating the exact modulus of $f(x) = \frac{1}{2} \|Ax - b\|_2^2 + \|x\|_0$.
- When studying how the exact modulus affects the convergence rate of PALM, we assume that the exact modulus is dominated by $\varphi(t) = c \cdot t^{1-\theta}$. However, the exact modulus may have various forms depending on the given function. Hence, a possible direction for future work is to analyze the convergence rate of PALM without this assumption.

Bibliography

[1] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438–457, 2010. → pages 22, 24, 25, 30, 32, 90, 96

[2] J.-P. Aubin and H. Frankowska. *Set-valued Analysis*. Springer Science & Business Media, 2009. → pages 7

[3] S. Banert and R. I. Bot. A general double-proximal gradient algorithm for d.c. programming. *Mathematical Programming*, 178(1):301–326, 2019. → pages 34

[4] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert spaces*. Springer, 2011. → pages 10, 11, 12, 16, 21

[5] H. H. Bauschke and S. B. Lindstrom. Proximal averages for minimization of entropy functionals. *arXiv:1807.08878*, 2018. → pages 7

[6] J. Bolte, A. Daniilidis, and A. Lewis. The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM Journal on Optimization*, 17(4):1205–1223, 2007. → pages 23, 24, 30

[7] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota. Clarke subgradients of stratifiable functions. *SIAM Journal on Optimization*, 18(2):556–572, 2007. → pages 26

[8] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet. Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity. *Transactions of the American Mathematical Society*, 362(6):3319–3363, 2010. → pages 23, 31

[9] J. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1):459–494, 2014. → pages 1, 20, 26, 31, 34, 85, 86, 87, 93, 96

[10] J. Bolte, T. P. Nguyen, J. Peypouquet, and B. Suter. From error bounds to the complexity of first-order descent methods for convex functions. *Mathematical Programming*, 165(2):471–507, 2016. → pages 35, 49, 80, 103

[11] A. Dontchev and R. T. Rockafellar. *Implicit Functions and Solution Mappings*. Springer, 2009. → pages 64

[12] E. Kreyszig. *Introductory Functional Analysis with Applications*. John Wiley & Sons, 1978. → pages 5

[13] K. Kurdyka. On gradients of functions definable in o-minimal structures. *Annales de l'institut Fourier*, 48(3):769–783, 1998. → pages 23

[14] G. Li and T. K. Pong. Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods. *Foundations of Computational Mathematics*, 18(5):1199–1232, 2018. → pages 25, 34, 35, 57, 61, 64

[15] S. Łojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. *Les équations aux dérivées partielles*, 117:87–89, 1963. → pages 23

[16] Z. Q. Luo, J. S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, 1996. → pages 34

[17] B. S. Mordukhovich. *Variational Analysis and Applications*. Springer, 2018. → pages 18

[18] A. M. Ostrowski. *Solutions of Equations in Euclidean and Banach Spaces*. Academic Press, 1973. → pages 93

[19] C. C. Pugh. *Real Mathematical Analysis*. Springer, 2015. → pages 82

[20] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970. → pages 13

[21] R. T. Rockafellar and R. J. B. Wets. *Variational Analysis*. Springer-Verlag, 1998. → pages 16, 17, 18, 20, 21, 92

[22] K. A. Ross. *Elementary Analysis: The Theory of Calculus*. Springer Science & Business Media, 2013. → pages 4, 5

[23] P. Yu, T. K. Pong, and Z. S. Lu. Convergence rate analysis of a sequential convex programming method with line search for a class of constrained difference-of-convex optimization problems. *arXiv:2001.06998*, 2020. → pages 34