UFG

UNIVERSIDADE FEDERAL DE GOIÁS
IME- INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
PROGRAMA DE PÓS GRADUAÇÃO EM MATEMÁTICA

TIAGO SOUSA MOTA

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

Goiânia
2025

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

[ ] Dissertação [X] Tese [ ] Outro*:_____________

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Tiago Sousa Mota

3. Título do trabalho

Convergence Analysis of Descent Optimization Algorithms under Polyak-Lojasiewicz- Kurdyka Conditions

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento [X] SIM [ ] NÃO[1]

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:
a) consulta ao(à) autor(a) e ao(à) orientador(a);
b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.
O documento não será disponibilizado durante o período de embargo.
Casos de embargo:
- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.

Documento assinado eletronicamente por **Tiago Sousa Mota**, **Discente**, em 03/07/2025, às 00:27, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Glaydston De Carvalho Bento**, **Professor do Magistério Superior**, em 08/08/2025, às 09:05, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5477110** e o código CRC **F0F5C7E7**.

Referência: Processo nº 23070.028857/2025-22 SEI nº 5477110

TIAGO SOUSA MOTA

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

Tese apresentada ao Programa de Pós–Graduação em matemática, do Instituto de Matemática e Estatística da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Doutor em Matemática.

Área de concentração: Otimização.

Orientador: Prof. Glaydston de Carvalho Bento

Goiânia
2025

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Mota, Tiago Sousa
Convergence Analysis of Descent Optimization Algorithms under Polyak-Lojasiewicz-Kurdyka Conditions [manuscrito] / Tiago Sousa Mota. - 2025.
86 f.: il.

Orientador: Prof. Dr. Glaydston de Carvalho Bento.
Tese (Doutorado) - Universidade Federal de Goiás, Instituto de Matemática e Estatística (IME), Programa de Pós-Graduação em Matemática, Goiânia, 2025.
Bibliografia. Apêndice.

1. Otimização não suave. 2. Métodos de descida. 3. Análise de convergência global. 4. Condições de Polyak-Lojasiewicz-Kurdyka. 5. taxa de convergência. I. Bento, Glaydston de Carvalho, orient. II. Título.

CDU 51

UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

ATA DE DEFESA DE TESE

Ata Nº **25** da sessão de Defesa de Tese de **Tiago Sousa Mota** que confere o título de Doutor em **Matemática**, na área de concentração em **Otimização.**

Ao **01/07/2025 primeiro dia do mês de julho do ano de dois mil e vinte e cinco**, a partir das **14h00**, vídeo Web conferência, realizou-se a sessão pública de Defesa de Tese intitulada **“Convergence Analysis of Descent Optimization Algorithms under Polyak-Lojasiewicz- Kurdyka Conditions”.** Os trabalhos foram instalados pelo Orientador, Professor Doutor **Glaydston de Carvalho Bento - IME/UFG** com a participação dos demais membros da Banca Examinadora: Professor Doutor **Orizon Pereira Ferreira - IME/UFG**, membro titular interno; Professor Doutor **Jefferson Divino Gonçalves de Melo - IME/UFG**, membro titular interno; Professor Doutor **João Xavier da Cruz Neto - DM/UFPI**, membro titular externo; Professor Doutor **Jurandir de Oliveira Lopes - DM/UFPI**, membro titular externo e Professor Doutor **Sándor Zoltán Németh - School of Mathematics/University of Birmingham**, membro titular externo. Durante a argüição os membros da banca **não fizeram** sugestão de alteração do título do **trabalho**. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Tese tendo sido o candidato **aprovado** pelos seus membros. Proclamados os resultados pelo Professor Doutor **Glaydston de Carvalho Bento**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, ao **01/07/2025 primeiro dia do mês de julho do ano de dois mil e vinte e cinco**.

TÍTULO SUGERIDO PELA BANCA

Documento assinado eletronicamente por **Sándor Zoltán Németh**, **Usuário Externo**, em 01/07/2025, às 16:12, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Jefferson Divino Goncalves De Melo**, **Professor do Magistério Superior**, em 01/07/2025, às 16:12, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020 .

Documento assinado eletronicamente por **Orizon Pereira Ferreira**, **Professora do Magistério Superior**, em 01/07/2025, às 16:13, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020 .

Documento assinado eletronicamente por **JOÃO XAVIER DA CRUZ NETO**, **Usuário Externo**, em 01/07/2025, às 16:14, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020 .

Documento assinado eletronicamente por **Glaydston De Carvalho Bento**, **Professor do Magistério Superior**, em 01/07/2025, às 16:14, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020 .

Documento assinado eletronicamente por **Jurandir de Oliveira Lopes**, **Usuário Externo**, em 01/07/2025, às 16:23, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020 .

A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **5472478** e o código CRC **2EAF7346**.

Referência: Processo nº 23070.028857/2025-22 SEI nº 5472478

Tiago Sousa Mota

Graduou–se em Matemática (Licenciatura) na UFES - Universidade Federal do Espírito Santo. Durante sua graduação, foi monitor no departamento de matemática pura e aplicada da UFES. Durante o Mestrado, na UNIFEI - Universidade Federal de Itajubá, foi bolsista da CAPES e da FAPEMIG.

Dedicado a:
Todos que acreditaram em mim e me apoiaram.

Agradecimentos

Primeiramente, gostaria de agradecer à CAPES pelo apoio financeiro.

Ao meu orientador, Dr. Glaydston de Carvalho Bento, pelos ensinamentos e por sua grande paciência, compreensão, incentivo, apoio, amizade e confiança, com os quais sempre pude contar ao longo da realização deste trabalho.

Aos professores membros da banca — Dr. Orizon Pereira Ferreira, Dr. Jefferson Divino Gonçalves de Melo, Dr. João Xavier da Cruz Neto, Dr. Sándor Zoltán Németh e Dr. Jurandir de Oliveira Lopes — por terem aceitado participar da defesa desta tese de doutorado, pelo tempo dedicado à leitura do manuscrito e pelas valiosas observações e sugestões.

Gostaria de expressar minha sincera gratidão aos professores Dr. Boris Mordukhovich, Dr. Antoine Soubeyran e Dr. Yurii Nesterov, que participaram como coautores em artigos que compõe a tese. Suas sugestões e críticas construtivas foram essenciais em diversas etapas da pesquisa, e a colaboração de cada um foi fundamental para que este trabalho se concretizasse com qualidade e profundidade. Muito obrigado!

Agradeço a todos os professores do IME-UFG pelo apoio e conhecimento transmitido — em especial ao Dr. Luiz Román e a todos que integram o grupo de Otimização — bem como a todos os funcionários do instituto, em especial para Ana Maria e Ulisses. Aos meus amigos do Espírito Santo — Flanderlon, Lucas, Douglas, Bruno, Weverton, Edilson — e aos(as) amigos(as) que conquistei em Goiânia — Ana Paula, Júlio Cesar, Vitória, Mayse — e a todos que estiveram por perto durante o doutorado. Agradeço também a todos que, direta ou indiretamente, contribuíram para a conclusão deste trabalho.

Um agradecimento especial à minha amiga Deysquele, por todo apoio e companheirismo ao longo do doutorado, pelos dias e noites de estudo e pela constante ajuda mútua. Agradeço também aos meus amigos Warley e Adimar, por todo o apoio e motivação ao longo dessa jornada.

Por fim, aos meus familiares, que sempre me motivaram e acreditaram em mim, e a Deus, por me conceder saúde, determinação, sabedoria e força para completar mais esta etapa.

<Nada é mais prático do que uma boa teoria>

<Kurt Lewin>,
<Field Theory in Social Science: Selected Theoretical Papers>.

Resumo

Mota, Tiago. **Convergence Analysis of Descent Optimization Algorithms under Polyak-Łojasiewicz-Kurdyka Conditions**. Goiânia, 2025. 85p. Tese de Doutorado Programa de Pós Graduação em Matemática, IME- Instituto de Matemática e Estatística, Universidade Federal de Goiás.

Esta tese apresenta uma análise de convergência abrangente de classes genéricas de algoritmos de descida em otimização não suave e não convexa sob a propriedade Polyak-Łojasiewicz-Kurdyka (PLK). Em particular, revisitamos e estendemos os resultados sobre taxas de convergência apresentadas por Khanh, Mordukhovich e Tran (J. Optim. Theory Appl., 2023), refinando a compreensão do expoente zero em funções PLK suaves e ampliando a discussão sobre a inconsistência entre a propriedade PLK de expoente baixo e a Lipschitz continuidade do gradiente para configurações mais gerais. Entre outras contribuições, estabelecemos a terminação finita de algoritmos genéricos sob condições PLK de expoente baixo. Além disso, Estabelecemos novas estimativas de taxa de convergência para métodos de gradiente inexatos e certas variantes do algoritmo na programação DC (diferença de funções convexa). Apresentamos resultados inovadores ao considerar uma condição de erro modificado, obtendo convergência finita ou superlinear para as sequências geradas. Notavelmente, revelamos que para uma ampla classe de programas de diferença de funções convexas, as condições PLK de expoente baixo são inerentemente incompatíveis com a Lipschitz continuidade do gradiente da função mais perto de um minimizador local, entretanto, monstramos que essa inconsistência pode não se manter se a continuidade de Lipschitz for substituída apenas pela continuidade do gradiente.

Palavras–chave

Otimização não suave, Métodos de descida, Análise de convergência global, Condições de Polyak-Łojasiewicz-Kurdyka, Taxa de convergência, Gradiente inexato.

Abstract

Mota, Tiago. **Convergence Analysis of Descent Optimization Algorithms under Polyak-Łojasiewicz-Kurdyka Conditions**. Goiânia, 2025. 85p. PhD. Thesis. Programa de Pós Graduação em Matemática, IME- Instituto de Matemática e Estatística, Universidade Federal de Goiás.

This thesis presents a comprehensive convergence analysis of generic classes of descent algorithms in nonsmooth and nonconvex optimization under the Polyak-Łojasiewicz-Kurdyka (PLK) property. In particular, we revisit and extend the results on convergence rates presented by Khanh, Mordukhovich, and Tran (J. Optim. Theory Appl., 2023), refining the understanding of the zero exponent in smooth PLK functions and broadening the discussion on the inconsistency between the lower exponent PLK property and the Lipschitz continuity of gradients to more general settings. Among other contributions, we establish the finite termination of generic algorithms under lower exponent PLK conditions. Additionally, we derive new convergence rates for inexact reduced gradient methods and certain variants of the boosted algorithm in DC programming. We present novel results by considering a modified error condition, obtaining either finite or superlinear convergence for the generated sequences. Notably, we reveal that for a broad class of difference programs, the lower exponent PLK conditions are inherently incompatible with the Lipschitz continuity of the gradient of the plus function near a local minimizer. However, we demonstrate that this inconsistency may not hold if Lipschitz continuity is replaced solely by gradient continuity.

Keywords

Nonsmooth optimization, Descent methods, Global convergence analysis, Polyak-Łojasiewicz-Kurdyka conditions, Convergence rate, Inexact gradient.

Contents

A **85**

CHAPTER 1

Introduction

The thesis examines general classes of descent algorithms in unconstrained (usually nonsmooth and nonconvex) optimization. Our main goals are to establish global convergence and derive convergence rates for such algorithms under appropriate conditions. It is well recognized in optimization theory and in many applications that Polyak-Łojasiewicz-Kurdyka-type (PŁK) conditions play a crucial role in the study of global convergence and convergence rates of various numerical algorithms.

First, in Chapter 2, definitions and auxiliary results on the PŁK property are presented, as well as nonsmooth analyses in this context. In addition, some definitions on classical concepts and classes of functions satisfying the PŁK property are presented. In chapter 3, consider the case where $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is smooth. The convergence of descent methods under the PŁK inequality has been extensively explored in recent years. Khanh et al. in [28] introduced and investigated an abstract descent model that generates a sequence $\{x^k\}$ satisfying, for sufficiently large k:

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

where $t_k, \beta, c > 0$, for $k \in \mathbb{N}$. In [28], the authors study the convergence rate of the abstract model by ensuring its full convergence through the observation that the sequences generated by (1-1) satisfy the following conditions:

($\mathscr{H}_1$) There exists $\sigma > 0$ such that for sufficiently large $k \in \mathbb{N}$ we have

$$\phi(x^k) - \phi(x^{k+1}) \geq \sigma\|\nabla\phi(x^k)\|\|x^{k+1} - x^k\|;$$

($\mathscr{H}_2$) For sufficiently large $k \in \mathbb{N}$, we have

$$[\phi(x^{k+1}) = \phi(x^k)] \Longrightarrow [x^{k+1} = x^k].$$

In fact, with respect to these conditions, Absil et al. in [1, Theorem 4.1] proved convergence of the generated sequences, for example, when the cost function is real analytical.

This general scheme comprise new linesearch methods with inexact gradient information for finding stationary points of continuously differentiable functions by considering the following selections of stepsize rules: backtracking stepsize, constant stepsize, and diminishing stepsize. In this chapter, we consider the abstract descent method proposed in [28], and we revisit [28, Theorem 2.5] including an analysis on finite convergence. For the reasons previously mentioned, our analysis in particular contemplates the convergence rate of the new linesearch methods with inexact gradient information for finding stationary points of continuously differentiable functions. We also present a sequence of results that both:

1. improves the understanding about the exponent zero of smooth function satisfying the PŁK condition;
2. shows that the exponent PŁK conditions (2-1) fails for $q \in (0, 1/2)$ in the case where ϕ is a C^1-smooth, $\{x^k\}$ is infinitely generated and the stepsize $\{t_k\}$ either is bounded away from 0 or has zero as the accumulation point being non-summable.

In Chapter 4, we examine the nonsmooth case discussed in [8], where Attouch, Bolté, and Svaiter studied the minimization of l.s.c. functions $\phi : \mathbb{R}^n \to \overline{\mathbb{R}}$ using a generic class of descent methods that satisfy the following properties:

($\mathscr{H}1$) *Sufficient decrease*: for each $k \in \mathbb{N}$, we have

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

($\mathscr{H}2$) *Relative error*: for each $k \in \mathbb{N}$, there exists

$$w^{k+1} \in \partial\phi(x^{k+1}) \text{ with } \|w^{k+1}\| \leq b\|x^{k+1} - x^k\|, \tag{1-2}$$

expressed in terms of the convexified subdifferential (2-7), with $a, b > 0$. It is shown in [8] that a great variety of important algorithms of optimization satisfy conditions ($\mathscr{H}1$) and ($\mathscr{H}2$). Employing the basic PŁK inequality from Definition 2.7(i), a qualitative convergence analysis, with arriving at the *limiting/(M)ordukhovich-stationary point* as $0 \in \partial\phi(\bar{x})$, is developed in [8] for the general class of descent algorithms satisfying ($\mathscr{H}1$) and ($\mathscr{H}2$) while without establishing any convergence rate.[1] A detailed qualitative (with convergence rates) analysis of the generic class of descent algorithms under the exponent PŁK conditions married to ($\mathscr{H}1$) and ($\mathscr{H}2$) was developed in [25], where the the obtained results were similar to [5] established for the proximal algorithm; namely, the finite step

---

[1] In fact, an additional technical and nonrestrictive assumption, labeled as ($\mathscr{H}4$) in what follows, was imposed in [8] and the subsequent publications discussed below.

convergence if $q = 0$, linear convergence if $q \in (0, 1/2]$, and polynomial convergence if $q \in (1/2, 1)$.

A major motivation for our study came from the paper by Aragón-Artacho and Vuong [4] who observed that the relative error condition $(\mathscr{H}2)$ may fail for the *Boosted Difference of Convex Algorithm* (BDCA) proposed in [2] for DC (difference of convex) programs. As shown in [4], the convergence analysis of [8] can be given for BDCA *provided that* the relative error condition (1-2) is replaced by

$$\text{there exists } \; w^k \in \overline{\partial}\phi(x^k) \; \text{ with } \; \|w^k\| \leq b\|x^{k+1} - x^k\|, \quad k \in \mathbb{N}, \tag{1-3}$$

expressed in terms of the convexified subdifferential (2-7). To furnish this, it is assumed in [4] that the functions g and h in the DC decomposition $\phi = g - h$ are strongly convex with g being of class $\mathscr{C}^{1,1}$, and that the basic PŁK condition in (2-4) is replaced by its strong version from Definition 2.7(iii). Under these requirements, [4] establishes the convergence of BDCA iterates to a (C)*larke-stationary point* $0 \in \overline{\partial}\phi(\bar{x})$.

In Chapter 4, besides establishing new results on the convergence rate for generic algorithms satisfying conditions $(\mathscr{H}1)$ and $(\mathscr{H}2)$ under the PŁK inequality, we also investigate another generic class of algorithms where the relative error condition $(\mathscr{H}2)$ is replaced by the following one:

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

$$w^k \in \partial\phi(x^k) \; \text{ with } \; \|w^k\| \leq b\|x^{k+1} - x^k\| \tag{1-4}$$

expressed in terms of the limiting subdifferential (2-5). The convergence analysis conducted below allows us to achieve, under the basic PŁK condition, the global convergence of any algorithm satisfying $(\mathscr{H}1)$ and $(\mathscr{H}3)$ to an M-stationary point $0 \in \partial\phi(\bar{x})$. This answers in the affirmative the question posted in [4, Remark 4.5], even with the improvement (1-4) of the stronger condition in (1-3) for general l.s.c. functions. Although the global convergence results under $(\mathscr{H}3)$ are parallel to those developed under $(\mathscr{H}2)$, this is not the case when it comes to convergence rate analysis. In what follows, we prove that the *lower exponent PŁK* condition $(0 < q < 1/2)$ always yields the *finite termination* of any descent algorithm satisfying $(\mathscr{H}1)$ and $(\mathscr{H}3)$. On the other hand, we present an example showing that the finite termination *fails* under PŁK lower exponents for the proximal method, which surely satisfies $(\mathscr{H}1)$ and $(\mathscr{H}2)$. Moreover, it comes as a surprise that in the case of difference programs to minimize $\phi = g - h$, where $h \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is convex while g is of class $\mathscr{C}^{1,1}$ without any convexity assumption, there exists *no desingularizing function* of the type $\varphi(t) = t^{1-q}$ with $q \in (0, 1/2)$ for which the PŁK condition holds at *local minimizers* of ϕ.

Chapter 5, besides pointing out some research directions, is dedicated to presenting some results obtained when revisiting the class of alternating algorithms, which was explored in [7] for non-convex structured functions of type $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ given by,

$$L(x,y) = f(x) + Q(x,y) + g(y),$$

where f and g are proper lower semicontinuous functions defined on Euclidean spaces, not necessarily convex, and Q is a smooth function that couples the variables x and y. The algorithm's dynamics are described as follows: let $(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$, the sequence is generated iteratively of the form $(x_k, y_k) \to (x_{k+1}, y_k) \to (x_{k+1}, y_{k+1})$:

$$x_{k+1} \in \arg\min \left\{ L(u, y_k) + \frac{1}{2\lambda_k} \|u - x_k\|^2 : u \in \mathbb{R}^n \right\}, \tag{1-5}$$

$$y_{k+1} \in \arg\min \left\{ L(x_{k+1}, v) + \frac{1}{2\mu_k} \|v - y_k\|^2 : v \in \mathbb{R}^m \right\}, \tag{1-6}$$

where $\{\lambda_k\}$ and $\{\mu_k\}$ are positive sequences. This proposal was originally explored for the convex case in [6, 9] and, in a co-substantiated manner, the authors in [7] substantiate its important connection with several important topics of classical literature, among which we highlight, the two-block Gauss-Seidel method and the classical alternating minimization algorithm, which can be seen as a proximal regularization of the von Neumann algorithm (see [48]).

In our analysis, we revisit the convergence rate result presented in [7] and present a valuable refinement concerning the speed of convergence for low-exponent Polyak-Łojasiewicz-Kurdyka functions. More precisely, we have established a finite or superlinear convergence rate in the case where the bifunction L satisfies the PŁK property with lower exponents. Additionally, we establish a convergence rate for the sequence of functional values.

We conclude this introduction by emphasizing that the present work is a compilation of three research papers. Two of these have been submitted to leading indexed journals of international scope in my field of research: one has already been accepted for publication (see [12]), while the other has undergone peer review, and we are currently awaiting the journal's final decision. The third paper is in the final stages of preparation and will be submitted to another prominent journal in the field of Optimization.

The first paper, developed jointly with G. C. Bento, constitutes the content of Chapter 3. The second paper, in collaboration with G. C. Bento, B. Mordukhovich, and Y. Nesterov, encompasses the material presented in Chapter 4. The third paper, written in collaboration with G. C. Bento, B. Mordukhovich, and A. Soubeyran, forms the basis of

Chapter 5. The final chapter outlines future research directions.

CHAPTER 2

Preliminaries Results

In this chapter we present some definition, notation, and preliminary results that will be used throughout Chapters 3 and 4.

Throughout this thesis, the inner product of two vectors $x, y \in \mathbb{R}^n$ is denoted by $\langle x, y\rangle$, while $\|\cdot\|$ denotes the induced norm, defined by $\|x\| = \sqrt{\langle x, x\rangle}$. The open ball of center x and radius $r > 0$ is denoted by $B(x, r)$.

Theorem 2.1 (Schwarz inequality) *For any vectors u and v in an inner product space, the following inequality holds:*

$$|\langle u, v\rangle| \leq \|u\| \cdot \|v\|.$$

Definition 2.2 *Let $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a real extended-valued function.*

a) Denote by

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

*its effective domain. We recall that ϕ is said to be **proper** when $dom(\phi) \neq \emptyset$, i. e., ϕ is not identically $+\infty$.*

b) The graph of a real-extended-valued function ϕ is defined by

$$\mathrm{Graph}\,\phi = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} \mid y = \phi(x)\}.$$

c) ϕ is a closed function if $\mathrm{Graph}\,\phi$ is a closed set.

*d) ϕ is **lower semicontinuous** (l.s.c.) if,*

$$\liminf_{u \to x} \phi(u) \geq \phi(x),$$

*e) ϕ is **convex** if,*

$$\phi(\lambda x + (1-\lambda)y) \leq \lambda\phi(x) + (1-\lambda)\phi(y),$$

for all $x, y \in \mathbb{R}^n$ and $\lambda \in]0, 1[$.

f) The function ϕ *is called* ***strongly convex*** *with modulus* $\rho > 0$ *if for all* $x, y \in \mathbb{R}^n$ *and* $\lambda \in]0,1[$,

$$\phi(\lambda x + (1-\lambda)y) \leq \lambda\phi(x) + (1-\lambda)\phi(y) - \frac{1}{2}\rho\lambda(1-\lambda)\|x-y\|^2$$

or, equivalently, when $\phi - \frac{\rho}{2}\|\cdot\|^2$ *is convex.*

g) The function ϕ *is called Lipschitz continuous if there is some constant* $L \geq 0$ *such that*

$$\|\phi(x) - \phi(y)\| \leq L\|x-y\|$$

for all $x, y \in \mathbb{R}^n$*, and* ϕ *is said to be locally Lipschitz continuous if, for every* x *in* $\mathbb{R}^n$*, there exists a neighborhood* U *of* x *such that* ϕ *restricted to* U *is Lipschitz continuous.*

Proposition 2.3 *A function* $\phi : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ *is l.s.c. if and only if the set* $\operatorname{Graph}\phi$ *is closed.*

Definition 2.4 *Let* $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *be a convex extended real-valued function. The convex subdifferential, denoted by* $\partial\phi(\bar{x})$*, of a function* ϕ *at* $\bar{x} \in \mathbb{R}^n$ *is defined at any point* $\bar{x} \in$ *dom* ϕ *by*

$$\partial\phi(\bar{x}) = \{u \in \mathbb{R}^n \mid \phi(x) - \phi(\bar{x}) \geq \langle u, x - \bar{x}\rangle, \forall\, x \in \mathbb{R}^n\},$$

and is empty otherwise.

If $T : \mathbb{R}^n \to P(\mathbb{R}^n)$ denotes a multivalued mapping on $\mathbb{R}^n$, the domain of T is the set of elements $x \in \mathbb{R}^n$ such that $T(x) \neq \emptyset$. The Fréchet-subdifferential (also known as regular-subdifferential) of ϕ at $x \in \mathbb{R}^n$ is defined by

$$\hat{\partial}\phi(x) := \begin{cases} \left\{x^* \in \mathbb{R}^n \;:\; \liminf\limits_{y \to x; y \neq x} \dfrac{1}{\|x-y\|}(\phi(y) - \phi(x) - \langle x^*, y - x\rangle) \geq 0\right\}, & if\ x \in \mathrm{dom}(\phi), \\ \varnothing, & if\ x \notin \mathrm{dom}(\phi). \end{cases}$$

and the (Mordukhovich or limiting) subdifferential of ϕ at $x \in \mathbb{R}^n$ is defined by

$$\partial_L\phi(x) := \begin{cases} \left\{x^* \in \mathbb{R}^n \;:\; \exists x_n \to x, \phi(x_n) \to \phi(x), x_n^* \in \hat{\partial}\phi(x_n); x_n^* \to x^*\right\}, & if\ x \in \mathrm{dom}(\phi), \\ \varnothing, & if\ x \notin \mathrm{dom}(\phi). \end{cases}$$

The symmetric subdifferential of ϕ at x, defined by

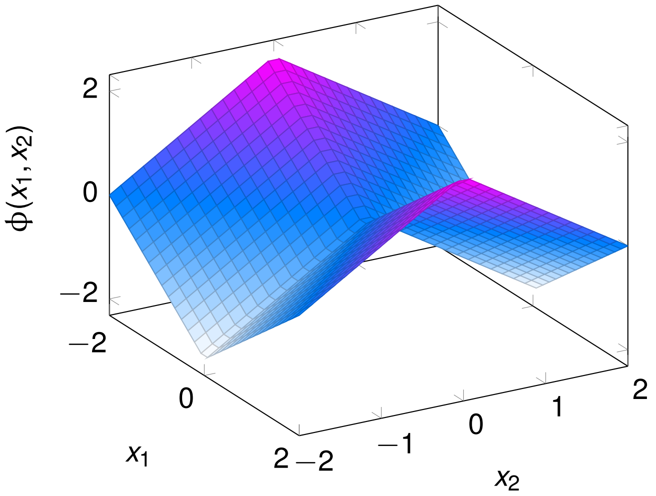
$$\partial^0\phi(x) := \partial_L\phi(x) \cup [-\partial_L(-\phi(x))],$$

It is clear that $\partial_L\phi(x) \subset \partial^0\phi(x)$.

An important property of the symmetric subdifferential is that $\partial^0(-\phi(x)) = -\partial^0\phi(x)$.

The example below highlights the relationship between the subdifferentials, see [37, pag 92]

Example 1 *For* $\phi : \mathbb{R}^2 \longrightarrow \mathbb{R}$ *given be* $\phi(x_1, x_2) = |x_1| - |x_2|$,



we have,

$$\partial_L\phi(0) = \{(v,1) \mid -1 \le v \le 1\} \cup \{(v,-1) \mid -1 \le v \le 1\},$$

$$-\partial_L(-\phi)(0) = \{(-1,v) \mid -1 \le v \le 1\} \cup \{(1,v) \mid -1 \le v \le 1\}.$$

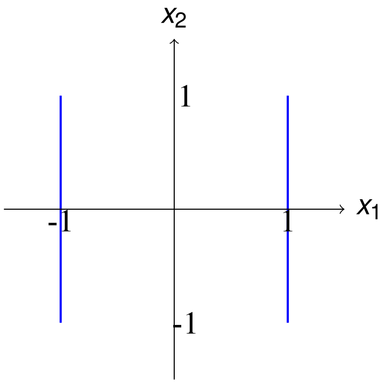


Figure 2.1: — $-\partial_L(-\phi)(0)$.

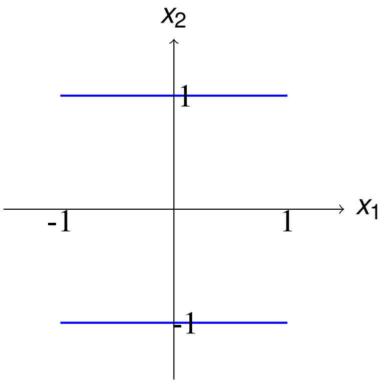


Figure 2.2: — $\partial_L\phi(0)$.

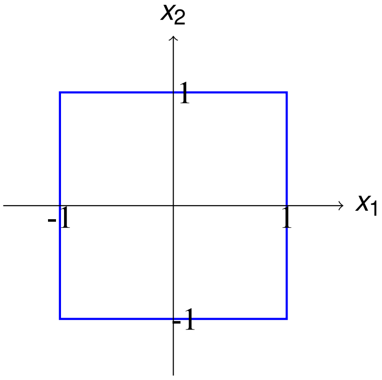


Figure 2.3: — $\partial^0\phi(0)$.

A point $x \in \mathbb{R}^n$ such that $0 \in \partial_L\phi(x)$ is called *limiting-critical* or simply *critical point* of ϕ.

The following lemma, whose proof idea appears in the proof of [1, Theorem 3.2], is important for the development of the convergence rate results.

Lemma 2.5 (estimates for monotone sequences) *Let* $\{a_k\} \subset [0,+\infty)$ *be a sequence monotone decreasing and let* $q \in (0,1)\cup(1,2)$. *Then,*

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

Proof. Since $a_k \subset [0,+\infty)$ is monotonically decreasing, we have $a_j \ge t$ for all $t \in [a_{j+1}, a_j]$, $j \in \mathbb{N}$. From this, it follows that

$$\frac{1}{(a_j)^q} \le \frac{1}{t^q}, \qquad t \in [a_{j+1}, a_j], \quad q \in (0,1)\cup(1,2),$$

which implies that

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

But this tell us that

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

from which the desired result follows. $\square$

2.1 Polyak-Łojasiewicz-Kurdyka Inequality

For functions ϕ of class $C^{1,1}$ (i.e., continuously differentiable functions with Lipschitzian gradients), Polyak introduced in 1962 (the English translation of his paper was published in [41]) the condition

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

and used it to prove a linear convergence of the gradient descent method in Hilbert spaces (Theorem 4 in [41]). Independently, Łojasiewicz [33] introduced the inequality

$$\|\nabla\phi(x)\| \geq c|\phi(x)-\phi(\bar{x})|^{q}, \quad c := 1/M(1-q), \quad q\in[0,1), \tag{2-1}$$

The gradient inequality (2-1), originally studied in the context of real analytic functions within the finite-dimensional framework of semialgebraic geometry and unrelated to optimization, has recently gained prominence under the name of the *Polyak–Łojasiewicz* (PL) condition—particularly in the literature on machine learning and computer science; see, e.g., [27]. A notable algebraic-geometric extension of (2-1) was developed by Kurdyka [30] for differentiable functions definable on o-minimal structures in $\mathbb{R}^n$. Subsequently, a nonsmooth generalization of this condition was introduced by Bolté et al. [15], formalized in Definition 2.7(ii), and referred to as the *Kurdyka–Łojasiewicz* (KL) inequality—though without reference to Polyak's original work. We suggest using the name of *Polyak-Łojasiewicz-Kurdyka* (PŁK) conditions for the properties of this type formulated below. More details on the employed subgradient constructions and their calculi can be found in [34, 36, 37]. We also refer the reader to [32] for important calculus rules for PŁK exponents and their applications.

Definition 2.6 *A differentiable function* $\phi:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ *is said to have the PŁK property at* $\bar{x}$ *iff there exist* $\eta\in]0,+\infty]$, *a neighborhood* U *of* $\bar{x}$, *and continuous concave function* $\varphi:[0,\eta[\to\mathbb{R}_+$ *(called desingularizing function) such that*

$$\varphi(0)=0, \quad \varphi\in C^1(]0,\eta[), \quad \varphi'(s)>0, \quad s\in]0,\eta[;$$

$$\varphi'(\phi(x)-\phi(\bar{x}))\|\nabla\phi(x)\|\geq 1, \quad x\in U\cap[\phi(\bar{x})<\phi<\phi(\bar{x})+\eta]. \tag{2-2}$$

- $[\phi(\bar{x})<\phi<\phi(\bar{x})+\eta] := \{x\in\mathbb{R}^n : \phi(\bar{x})<\phi(x)<\phi(\bar{x})+\eta\}$.

For more details see, for instance, [7, 30] where the authors present a wide list of classes of functions satisfying the KL inequality where the desingularizing function φ is given by $\varphi(t)=Mt^{1-q}$, $M>0$, $q\in[0,1)$.

For the nonsmooth case, we have the following definition,

Definition 2.7 (PŁK conditions) Let $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be an extended-real-valued lower semicontinuous (l.s.c.) function with the domain $\mathrm{dom}\,\phi \neq \emptyset$. We say that the function ϕ satisfies:

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

$$\varphi(0) = 0, \quad \varphi \in C^1(0, \eta), \quad \varphi'(s) > 0 \text{ for all } s \in (0, \eta), \text{ and} \tag{2-3}$$

$$\varphi'\big(\phi(x) - \phi(\bar{x})\big)\mathrm{dist}\big(0, \partial\phi(x)\big) \geq 1 \text{ for all } x \in U \cap [\phi(\bar{x}) < \phi(x) < \phi(\bar{x}) + \eta] \tag{2-4}$$

- *dist*$(0, \partial\phi(x)) :=$ *inf*$\{\|v\| : v \in \partial\phi(x)\}$,
- $[\phi(\bar{x}) < \phi < \phi(\bar{x}) + \eta] := \{x \in \mathbb{R}^n : \phi(\bar{x}) < \phi(x) < \phi(\bar{x}) + \eta\}$.

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

$$\partial\phi(\bar{x}) := \left\{ v \in \mathbb{R}^n \,\middle|\, \exists x_k \xrightarrow{\phi} \bar{x},\, v_k \in \widehat{\partial}\phi(x_k), \quad v_k \to v \text{ as } k \to \infty \right\}, \tag{2-5}$$

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

$$\liminf_{u \to x_k} \frac{\phi(u) - \phi(x_k) - \langle v_k, u - x_k \rangle}{\|u - x_k\|} \geq 0,\ k \in \mathbb{N} := \{1, 2, \ldots\}.$$

(ii) The *symmetric PŁK condition* at $\bar{x}$ if ϕ is continuous around $\bar{x}$ and $\partial\phi$ is replaced in (2-4) by the *symmetric subdifferential* of ϕ at $\bar{x}$ defined by

$$\partial^0\phi(\bar{x}) := \partial\phi(\bar{x}) \cup \big(-\partial(-\phi)(\bar{x})\big). \tag{2-6}$$

(iii) The *strong PŁK condition* if ϕ is Lipschitz continuous around $\bar{x}$ and ∂ is replaced in (2-4) by the *Clarke/convexified subdifferential* of ϕ at $\bar{x}$ given by

$$\overline{\partial}\phi(\bar{x}) := \mathrm{co}\,\partial\phi(\bar{x}), \tag{2-7}$$

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

(iv) The *exponent* versions of the *PŁK conditions* in (i)–(iii) if the desingularizing function in (2-3) and (2-4) is selected in the form $\varphi(t) = Mt^{1-q}$, where M is a positive constant, and where $q \in [0, 1)$. We refer to the case where $q \in (0, 1/2)$ as the *PŁK conditions with lower exponents.*

Remark 1 *For differentiable functions definable on an o-minimal structure defined over* $\mathbb{R}$ *(for more details on definable functions on an o-minimal structure, see [47]), the inequality* (2-4) *retrieves the inequality* (2-2) *through the following result:*

"Given $U \subset \mathbb{R}^n$ *a bounded open set and* $h : U \to \mathbb{R}_+$ *a differentiable function definable on a o-minimal structure, there exist* $c, \eta > 0$ *and a strictly increasing positive function definable* $\varphi : \mathbb{R}_+ \to \mathbb{R}$ *of class* C^1*, such that*

$$\|\nabla(\varphi \circ h)(x)\| \geq c, \qquad x \in U \cap h^{-1}(0, \eta)."$$

Using inequality (2-4) *as a reference, it was assumed in this result that* $h(\bar{x}) = 0$. *Note that taking* $\varphi(t) = t^{1-q}$, $q \in [0, 1)$, *the last inequality yields*

$$\|\nabla h(x)\| \geq c|h(x)|^q,$$

where $c = 1/(1-q)$.

Lemma 2.8 *Suppose that* ϕ *is a proper closed function,* $\bar{x} \in \operatorname{dom} \partial\phi$ *and* $0 \notin \partial\phi(\bar{x})$. *Then, for any* $q \in [0, 1)$*, the function* ϕ *satisfies the Polyak–Kurdyka–Łojasiewicz (PŁK) property at* $\bar{x}$ *with exponent* q.

Proof. See [32, Lemma 2.1]. □

Semialgebraic functions: Recall that a subset of $\mathbb{R}^n$ is called *semialgebraic* if it can be expressed as a finite union of sets of the form

$$\{x \in \mathbb{R}^n \mid p_i(x) = 0,\, q_i(x) < 0,\, i = 1, \ldots, p\},$$

where each p_i and q_i is a real polynomial function.

A function $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called *semialgebraic* if its graph is a semialgebraic subset of $\mathbb{R}^{n+1}$. Such functions satisfy the *PŁK property* (see Bolte et al. [10, 16]) with

$$\varphi(s) = Ms^{1-q},$$

for some $q \in [0, 1) \cap \mathbb{Q}$ and some constant $M > 0$. Stability properties of semialgebraic functions are numerous (see, e.g., Benedetti and Risler [11], Bochnak et al. [14]); the following few facts might help the reader to understand how they impact optimization matters:

- Finite sums and products of semialgebraic functions are semialgebraic;
- Scalar products are semialgebraic;
- Indicator functions of semialgebraic sets are semialgebraic;
- Generalized inverses of semialgebraic mappings are semialgebraic;
- Composition of semialgebraic functions or mappings are semialgebraic;

– Functions of the type $\phi : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\phi(x) = \sup_{y \in C} g(x,y) \quad \text{(resp. } \phi(x) = \inf_{y \in C} g(x,y)\text{)},$$

where g and C are semialgebraic, are semialgebraic.

Below we can see some examples where the desingularizing function of the PŁK property is presented for some classes of functions by [7].

Growth condition for convex functions: consider a convex function ϕ satisfying the following growth condition: exist U neighborhood of $\bar{x}$, $\eta > 0, c > 0, r > 0$, such that

$$\forall\, x \in U \cap [\min\phi < \phi < \min\phi + \eta], \quad \phi(x) \geq \phi(\bar{x}) + c\,\mathrm{dist}(x, arg\min\phi)^r, \tag{2-8}$$

here $\bar{x} \in arg\min\phi \neq \emptyset$. Then ϕ complies with (2-2) at point $\bar{x}$ (for $\varphi(s) = rc^{-\frac{1}{r}}s^{\frac{1}{r}}$) on $U \cap [\min\phi < \phi < \min\phi + \eta]$.

Example 2 *Let* $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ *given by* $\phi(x) = |x|^p$, *with* $p \in \mathbb{N}, p \geq 3$ *and let* $U = (-\epsilon, \epsilon)$ *for* $\epsilon > 0$. ϕ *satisfies* (2-8) *with* $\bar{x} = 0$, *in fact*

$$\phi(x) \geq \phi(\bar{x}) + c|x - \bar{x}|^p \Longleftrightarrow |x|^p \geq c|x|^p, \quad c \in (0,1].$$

Uniform convexity: If ϕ is uniformly convex, i.e., satisfies

$$\phi(y) \geq \phi(x) + \langle \nabla\phi(x), y - x \rangle + K\|y - x\|^p, \quad p \geq 1,$$

for all $x, y \in \mathbb{R}^n$, then ϕ satisfies the PŁK inequality on $\mathrm{dom}\,\phi$ for $\varphi(s) = pK^{-\frac{1}{p}}s^{\frac{1}{p}}$.

Preliminares do paper 3

2.2 The Quotient Convergence Factors

This section is based in [39, Chapter 9]. Let $\mathscr{F}$ be an iterative process generating a sequence $\{x_k\}$, and let $\bar{x}$ be one of its limit points. For the purposes of this section, it will suffice to consider such a process simply as a collection of sequences in $\mathbb{R}^n$ and to disregard how these sequences are generated.

Our goal is to associate with $\mathscr{F}$ precise indicators that characterize the asymptotic convergence rate of the process at $\bar{x}$. Specifically, the construction of these indicators is motivated by the relevance of estimates of the form

$$\|x_{k+1} - \bar{x}\| \leq \gamma \|x_k - \bar{x}\|^p, \quad \forall k \geq 0, \quad \gamma > 0 \text{ and } p \in [1, \infty),$$

such estimates frequently emerge naturally in the analysis of certain iterative processes. To begin, we define these indicators for a general convergent sequence, without assuming that it originates from any specific iterative scheme.

Definition 2.9 *Let $\{x_k\} \subset \mathbb{R}^n$ be any convergent sequence with limit $\bar{x}$. Then the quantities*

$$Q_p(x_k) = \begin{cases} 0, & \text{if } x_k = \bar{x}, \text{ for all but finitely many } k; \\ \limsup_{k \to \infty} \frac{\|x_{k+1} - \bar{x}\|}{\|x_k - \bar{x}\|^p}, & \text{if } x_k \neq \bar{x}, \text{ for all but finitely many } k; \\ +\infty, & \text{otherwise}, \end{cases}$$

defined for all $p \in [1, \infty)$, are the quotient convergence factors, or Q-factors, for short, of $\{x_k\}$ with respect to the norm $||.||$ on $\mathbb{R}^n$.

When analyzing an iterative process $\mathscr{F}$, rather than a single sequence, it is desirable for the rate-of-convergence indicator to capture the worst-case asymptotic convergence rate among all sequences generated by $\mathscr{F}$ that converge to the same limit point.

Definition 2.10 *Let $C(\mathscr{F}, \bar{x})$ denote the set of all sequences with limit $\bar{x}$ generated by an iterative process $\mathscr{F}$. Then*

$$Q_p(\mathscr{F}, \bar{x}) = \sup\left\{Q_p(x_k) \mid \{x_k\} \in C(\mathscr{F}, \bar{x})\right\}, \quad 1 \leq p < +\infty,$$

are the Q-factors of $\mathscr{F}$ at $\bar{x}$ with respect to the norm in which the $Q_p(x_k)$ are computed.

It is worth emphasizing that, in general, the Q-factors of a process depend on the specific limit point $\bar{x}$, and may vary across different limit points of $\mathscr{F}$.

The primary motivation for introducing the Q-factors of an iterative process is to have a precise means of comparing the rate of convergence of different iterations. We utilize the Q-factors in this connection as follows:

Definition 2.11 *Let $\mathscr{F}_1$ and $\mathscr{F}_2$ denote two iterative processes with the same limit point $\bar{x}$, and let $Q_p(\mathscr{F}_1, \bar{x})$ and $Q_p(\mathscr{F}_2, \bar{x})$ be the corresponding Q-factors computed in the same norm on $\mathbb{R}$. Then $\mathscr{F}_1$ is* Q-faster *than $\mathscr{F}_2$ at $\bar{x}$ if there is a $p \in [1, \infty)$ such that $Q_p(\mathscr{F}_1, \bar{x}) \leq Q_p(\mathscr{F}_2, \bar{x})$.*

Iterative processes with Q-orders 1, 2, or 3 play an especially important role in the theory, and we introduce some additional terminology that will be useful on occasion. Whenever $Q(\mathscr{F}, \bar{x}) = 0$ $(p = 1)$, we say that the process has **Q-superlinear convergence** at $\bar{x}$, while, if $0 < Q(\mathscr{F}, \bar{x}) < 1$, the convergence is called **Q-linear**. Any process $\mathscr{F}$ for which $Q(\mathscr{F}, \bar{x}) \geq 1$ is called **Q-sublinear**.

Similarly, any process of Q-order two for which

$$0 < Q_2(\mathscr{F}, \bar{x}) < +\infty,$$

is called **Q-quadratic** at $\bar{x}$, while, for $Q_2(\mathscr{F}, \bar{x}) = 0$ or $Q_2(\mathscr{F}, \bar{x}) = +\infty$, we refer to **Q-superquadratic** or **Q-subquadratic** convergence, respectively. Finally, an analogous terminology is sometimes applied to Q-order three convergence with "quadratic" replaced by "cubic."

CHAPTER 3

Convergence Rate of Inexact Gradient Methods for Functions Satisfying the Polyak-Łojasiewicz-Kurdyka Inequality

In this chapter, we consider the abstract descent method proposed and developed by Khanh, Mordukhovich, Tran (J Optim Theory Appl https:// doi.org/10.1007/s10957-023-02319-9, 2023). In our analysis, we revisited the result about convergence rate presented by Khanh, Mordukhovich, Tran and presented new results that both enhance the understanding of the zero exponent in smooth PŁK functions and broaden the discussion on the inconsistency between the lower exponent PŁK property and Lipschitz continuity of gradients to more general situations.

3.1 Inexact Gradient Methods

The convergence of descent methods under the PŁK inequality has been extensively explored in recent years. Khanh et al. in [28] introduced and investigated an abstract descent model that generates a sequence $\{x^k\}$ satisfying, for sufficiently large k:

$$\phi(x^k) - \phi(x^{k+1}) \geq \frac{\beta}{t_k}\|x^{k+1} - x^k\|^2 \text{ and } \|\nabla\phi(x^k)\| \leq \frac{c}{t_k}\|x^{k+1} - x^k\|, \tag{3-1}$$

where $t_k, \beta, c > 0$, for $k \in \mathbb{N}$. In [28], the authors study the convergence rate of the abstract model by ensuring its full convergence through the observation that the sequences generated by (3-1) satisfy the following conditions:

($\mathscr{H}_1$) There exists $\sigma > 0$ such that for sufficiently large $k \in \mathbb{N}$ we have

$$\phi(x^k) - \phi(x^{k+1}) \geq \sigma\|\nabla\phi(x^k)\|\|x^{k+1} - x^k\|;$$

($\mathscr{H}_2$) For sufficiently large $k \in \mathbb{N}$, we have

$$[\phi(x^{k+1}) = \phi(x^k)] \Longrightarrow [x^{k+1} = x^k].$$

In fact, with respect to these conditions, Absil et al. in [1, Theorem 4.1] proved convergence of the generated sequences, for example, when the cost function is real analytical. As a particular instance of this abstract model, the authors in [28] still consider the following general scheme for inexact reduced gradient methods:

Method 1

- ***Step 0*** *(initialization) Select an initial point* $x^1 \in \mathbb{R}^n$*, initial radii* $\epsilon_1, r_1 > 0$*, radius reduction factors* $\mu, \theta \in (0,1)$.
- ***Step 1*** *(inexact gradient and stopping criterion) Choose* g^k *such that*

$$\|g^k - \nabla\phi(x^k)\| \leq \epsilon_k.$$

- ***Step 2*** *(radius update) If* $\|g^k\| \leq r_k + \epsilon_k$*, then set* $r_{k+1} := \mu r_k$*,* $\epsilon_{k+1} := \theta\epsilon_k$*,* $d^k := 0$*, and go to Step 3. Otherwise, set* $r_{k+1} := r_k$*,* $\epsilon_{k+1} := \epsilon_k$*, and*

$$d^k := -\frac{\|g^k\| - \epsilon_k}{\|g^k\|} g^k.$$

- ***Step 3*** *(stepsize) Choose* $t_k > 0$ *by a specific rule.*
- ***Step 4*** *(iteration update) Set* $x^{k+1} := x^k + t_k d^k$.

This general scheme comprise new linesearch methods with inexact gradient information for finding stationary points of continuously differentiable functions by considering the following selections of stepsize rules: backtracking stepsize, constant stepsize, and diminishing stepsize. In this paper, we consider the abstract descent method proposed in [28], and we revisit [28, Theorem 2.5] including an analysis on finite convergence. For the reasons previously mentioned, our analysis in particular contemplates the convergence rate of the new linesearch methods with inexact gradient information for finding stationary points of continuously differentiable functions. We also present a sequence of results that both:

1. improves the understanding about the exponent zero of smooth PŁK functions;
2. shows that the exponent KL property fails for $q \in (0, 1/2)$ in the case where f is a C^1-smooth, $\{x^k\}$ is infinitely generated and the stepsize $\{t_k\}$ either is bounded away from 0 or has zero as the accumulation point being non-summable.

3.2 Convergence Rate

In [28] the authors present a good discussion showing that they are the first to consider the convergence rate for methods satisfying condition (3-1). They note that some particular instances have been considered, for example in [27, 41], where the authors analyze the convergence of the exact gradient method for smooth functions under the PŁK condition. Since exact gradient methods is a special case of the inexact linesearch method satisfying (3-1) and the PL property is a special case of the PŁK property (it holds by considering q=1/2 in the desingularizing function $\varphi(t) = Mt^{1-q}$), in fact, [28, Theorem 2.5] extends the scope of applicability of the results in [27, 41].

In the following result, we show an improvement in the convergence rate of [28, Theorem 2.5], indicating that for $q \in [0, 1/2)$, the sequence $\{x^k\}$ converges within a finite number of steps.

Theorem 3.1 *Let the sequences* $\{x^k\} \subset \mathbb{R}^n$, $\{t_k\} \subset \mathbb{R}_+$ *and the numbers* $\beta > 0$, $c > 0$ *satisfying* (3-1) *for* $k \in \mathbb{N}$ *sufficiently large. Suppose that there exists* $\bar{t} > 0$ *such that* $t_k > \bar{t}$, $k \in \mathbb{N}$, *that* $\bar{x}$ *is an accumulation point of* $\{x^k\}$, *and that* f *satisfies the PŁK property at* $\bar{x}$ *with* $\varphi(t) = Mt^{1-q}$ *for some* $M > 0$ *and* $q \in [0, 1)$. *The following convergence rates are guaranteed:*

i) if $q \in [0, \frac{1}{2})$, *then the sequence* $\{x^k\}$ *terminates at* $\bar{x}$ *in finite steps;*

Moreover, when $\{x^k\}$ *is such that* $x^{k+1} \neq x^k$ *for all* $k \in \mathbb{N}$,

ii) if $q = \frac{1}{2}$, *then the sequence* $\{x^k\}$ *converges linearly to* $\bar{x}$;

iii) if $q \in (\frac{1}{2}, 1)$, *then there exists a positive constant* Q *such that*

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

Proof. As emphasized in the introduction of the paper, the sequence $\{x^k\}$ generated from (3-1) converges to $\bar{x}$. From the decreasing condition in (3-1), we have the sequence $\{\phi(x^k)\}$ is monotone decreasing. Because $\bar{x}$ is an accumulation point of the sequence $\{x^k\}$, then $\{\phi(x^k)\}$ converges to $\phi(\bar{x})$ as k goes to infinity. Defining $\Psi(x) := \phi(x) - \phi(\bar{x})$, it follows that $\{\Psi(x^k)\} \subset [0, +\infty)$ is monotone decreasing. From Lemma 2.5 with $a_k = \Psi(x^k)$, we have

$$\sum_{j=k}^{k+l} \frac{\Psi(x^j) - \Psi(x^{j+1})}{\Psi(x^j)^q} \leq \frac{1}{1-q}\left(\Psi(x^k)^{1-q} - \Psi(x^{k+l+1})^{1-q}\right), \qquad k, l \in \mathbb{N}. \tag{3-2}$$

On the other hand, combining the two inequality in (3-1), we get

$$\phi(x^{k+1}) \leq \phi(x^k) - \frac{\beta}{t_k}\|x^{k+1} - x^k\|^2 \leq \phi(x^k) - \frac{\beta t_k}{c^2}\|\nabla\phi(x^k)\|^2. \tag{3-3}$$

From (2-2) with $\varphi(t) = Mt^{1-q}$, we have

$$\|\nabla\phi(x^k)\| \geq \frac{1}{M(1-q)}(\phi(x^k) - \phi(\bar{x}))^q. \tag{3-4}$$

Combining (3-3)-(3-4) and using that $\Psi(x^k) = \phi(x^k) - \phi(\bar{x})$, it follow that

$$\Psi(x^{k+1}) \leq \Psi(x^k) - \frac{\beta t_k}{M^2(1-q)^2c^2}(\Psi(x^k))^{2q}. \tag{3-5}$$

For the proof of item i), let us first assume that $q \in (0, 1/2)$ and suppose, by contradiction, that the sequence $\{x^k\}$ is infinitely generated. But this tell us that $\Psi(x^k) > 0$ for all $k \in \mathbb{N}$. Because $\Psi(x^k)$ converges to 0 as k goes to infinity and taking into account that $2q - 1 < 0$, from the equality

$$\Psi(x^k) - \frac{\beta t_k}{M^2(1-q)^2c^2}(\Psi(x^k))^{2q} = \Psi(x^k)\left[1 - \frac{\beta t_k}{M^2(1-q)^2c^2}(\Psi(x^k))^{2q-1}\right],$$

combined with (3-5), there exists $k_0 \in \mathbb{N}$ such that

$$\begin{aligned} \Psi(x^{k_0+1}) &\leq \Psi(x^{k_0})\left[1 - \frac{\beta t_{k_0}}{M^2(1-q)^2c^2}(\Psi(x^{k_0}))^{2q-1}\right] \\ &\leq \Psi(x^{k_0})\left[1 - \frac{\beta \bar{t}}{M^2(1-q)^2c^2}(\Psi(x^{k_0}))^{2q-1}\right] < 0, \end{aligned}$$

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

To conclude item *i*), let us now assume the case $q = 0$ and suppose, by contradiction, that the sequence is infinitely generated. Using (3-4) with $q = 0$, for $k \in \mathbb{N}$ sufficiently large and that $t_k > \bar{t} > 0$ it follows that

$$\|\nabla\phi(x^k)\| \geq \frac{1}{M},$$

which, combined with (3-1) and taking into account that $t_k > \bar{t}$ for all $k \in \mathbb{N}$, yields

$$\frac{\bar{t}}{Mc} \leq \frac{t_k}{Mc} \leq \|x^{k+1} - x^k\|.$$

But the previous inequality contradicts the fact that $\{x^k\}$ converges to $\bar{x}$, which conclude the proof of item *i*). For proving item *ii*), note that from (3-5) we obtain

$$\Psi(x^{k+1}) \leq \left(1 - \frac{4\beta t_k}{M^2c^2}\right)\Psi(x^k) \leq \left(1 - \frac{4\beta \bar{t}}{M^2c^2}\right)\Psi(x^k), \tag{3-6}$$

where in the last inequality we have used the fact that $t_k \geq \bar{t}$, $k \in \mathbb{N}$. Because $\Psi(x^k) \geq 0$, $k \in \mathbb{N}$, from (3-6) it follows, in partircular, that $4\beta\bar{t}/(M^2c^2) \in (0, 1)$ and, hence, that $\{\phi(x^k)\}$ converges linearly for $\{\phi(\bar{x})\}$. On the other hand, from (3-1), we have

$$\phi(x^{k+1}) \leq \phi(x^k) - \frac{\beta}{c}\|\nabla\phi(x^k)\|\|x^{k+1} - x^k\|, \tag{3-7}$$

which, combined with (3-4) and definition of Ψ yields

$$\Psi(x^{k+1}) \leq \Psi(x^k) - \frac{2\beta}{cM}\|x^{k+1} - x^k\|(\Psi(x^k))^{\frac{1}{2}}.$$

From the last inequality we obtain

$$\|x^{k+1} - x^k\| \leq \frac{cM}{2\beta}\frac{\Psi(x^k) - \Psi(x^{k+1})}{\Psi(x^k)^{\frac{1}{2}}},$$

which, combined with (3-2), yields

$$\sum_{j=k}^{k+l}\|x^{j+1} - x^j\| \leq \frac{cM}{2\beta}\sum_{j=k}^{k+l}\frac{\Psi(x^j) - \Psi(x^{j+1})}{\Psi(x^j)^{\frac{1}{2}}} \leq \frac{cM}{\beta}(\Psi(x^k)^{\frac{1}{2}} - \Psi(x^{k+l+1})^{\frac{1}{2}}).$$

Taking limits when l goes to infinity in the last inequality, it follows that

$$s_k := \sum_{j=k}^{+\infty}\|x^{j+1} - x^j\| \leq \frac{cM}{\beta}(\Psi(x^k))^{\frac{1}{2}}.$$

Combining the definition of s_k with the last inequality, (3-4), (3-1), and using that $t_k \geq \bar{t}$, $k \in \mathbb{N}$, we obtain

$$s_{k+1} \leq \frac{cM}{\beta}(\Psi(x^k))^{\frac{1}{2}} \leq \frac{cM^2}{2\beta}\|\nabla\phi(x^k)\| \leq \frac{c^2M^2}{2\beta t_k}\|x^{k+1} - x^k\| \leq \frac{c^2M^2}{2\beta\bar{t}}\|x^{k+1} - x^k\|.$$

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

$$s_{k+1} \leq \frac{c^2M^2}{2\beta\bar{t}}(s_k - s_{k+1}),$$

from which it follows that

$$s_{k+1} \leq \frac{\tau}{1+\tau}s_k, \qquad \tau := \frac{c^2M^2}{2\beta\bar{t}}.$$

But this tells us that the sequence $\{s_k\}$ converges linearly to 0. On the other hand, using the triangle inequality we have $\|x^k - x^{k+l}\| \leq \sum_{j=k}^{k+l}\|x^{j+1} - x^j\|$, from which it follows that $\|x^k - \bar{x}\| \leq s_k$. Therefore, the linear convergence of $\{x^k\}$ to $\bar{x}$ can be deduced from the linear convergence of $\{s_k\}$ to 0, which conclude the proof of item ii). The proof of item iii) remains as in [28] and thus we conclude the proof of the theorem. $\square$

Next claim improves the understanding about the exponent 0 of smooth PŁK functions.

Proposition 3.2 *Let* $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *be* C^1 *around some stationary point* $\bar{x} \in \mathbb{R}^n$*. If* ϕ *satisfies the PŁK property at* $\bar{x}$ *with exponent 0, then* ϕ *is locally constant around* $\bar{x}$*.*

Proof. Since ϕ is C^1 around $\bar{x}$ and satisfies the PŁK property at $\bar{x}$ with exponent 0, we find some neighborhood U of $\bar{x}$ such that $\nabla\phi$ is smooth on U and

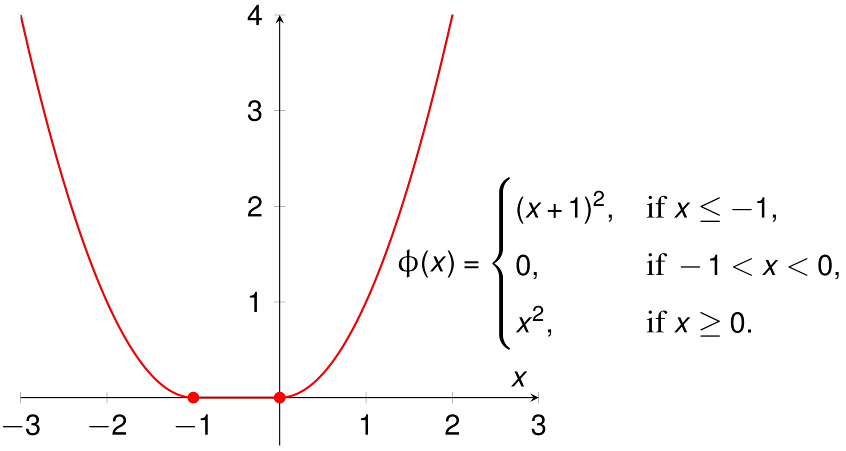
$$M|\phi(x) - \phi(\bar{x})|^0 \le \|\nabla\phi(x)\| \quad \text{for all } x \in U. \tag{3-8}$$

Assume that ϕ is not locally constant around $\bar{x}$. Then we can construct some sequence $\{x^k\} \subset U$, $x^k \to \bar{x}$ as $k \to \infty$ and $\phi(x^k) \neq \phi(\bar{x})$. It follows from (3-8) that

$$\|\nabla\phi(x^k)\| \ge M \quad \text{for all } k \in \mathbb{N}.$$

Letting $k \to \infty$ gives us $\|\nabla\phi(\bar{x})\| \ge M$, which contradicts with the stationarity of $\bar{x}$. ☐

Let us now look at an example to illustrate the above proposition.



The next result shows that the exponent KL property at $\bar{x}$ fails whenever $q \in (0, 1/2)$ for any algorithm applied to a C^1-smooth objective function that generates an infinite sequence satisfying (3-1) with a sequence of stepsize $\{t_k\}$ bounded away from 0.

Proposition 3.3 *Let* $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *be* C^1 *and suppose that the sequence* $\{x^k\}$*, with* $x^{k+1} \neq x^k$ *for all* $k \in \mathbb{N}$*, satisfies the condition* (3-1) *and has an accumulation point* $\bar{x}$*. If there exists* $\bar{t}$ *such that* $t_k > \bar{t} > 0$ *for all* $k \in \mathbb{N}$*, then the exponent PŁK property of* ϕ *in* $\bar{x}$ *fails whenever* $q \in (0, 1/2)$*.*

Proof. First, note that if $x^k \neq x^{k+1}$ holds for all $k \in \mathbb{N}$, as

$$\phi(x^k) - \phi(x^{k+1}) \ge \frac{\beta}{t_k}\|x^{k+1} - x^k\|^2,$$

it follows that $\phi(x^k) \neq \phi(x^{k+1})$ for all $k \in \mathbb{N}$ and suppose, by contradiction, that ϕ satisfies the PŁK property in $\bar{x}$ with exponent $q \in (0, 1/2)$. Combining the two inequalities in (3-1), we obtain

$$\phi(x^k) - \phi(x^{k+1}) \geq \frac{\beta t_k}{c^2}\|\nabla\phi(x^k)\|^2. \tag{3-9}$$

Using (3-4) and also that $t_k > \bar{t} > 0$, for all $k \in \mathbb{N}$, it follows that,

$$\phi(x^k) - \phi(\bar{x}) \geq \phi(x^k) - \phi(x^{k+1}) \geq c_1\left(\phi(x^k) - \phi(\bar{x})\right)^{2q}, \qquad c_1 := \frac{\beta\bar{t}}{c^2M^2(1-q)^2} > 0,$$

and, consequently, we have

$$\left(\phi(x^k) - \phi(\bar{x})\right)^{1-2q} \geq c_1.$$

Since $\phi(x^k) - \phi(\bar{x})$ converges to zero, the inequality above leads to a contradiction for $q \in (0,1/2)$, which concludes the proof of the proposition. □

The next result shows that even when $\{t_k\}$ is not summable but has zero as an accumulation point, the exponent PŁK property of f at $\bar{x}$ still fails whenever $q \in (0,1/2)$.

Proposition 3.4 *Let $\phi : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ be C^1 and suppose that the sequence $\{x^k\}$, with $x^{k+1} \neq x^k$ for all $k \in \mathbb{N}$, satisfies condition (3-1) and has an accumulation point $\bar{x}$. If $\{t_k\}$ has zero as an accumulation point with $\sum_{j=k}^{+\infty} t_j = \infty$, then the exponent PŁK property of f at $\bar{x}$ fails whenever $q \in (0,1/2)$.*

Proof. Note that $\phi(x^k) \neq \phi(x^{k+1})$ for all $k \in \mathbb{N}$ and suppose, by contradiction, that f satisfies the PŁK property in $\bar{x}$ with exponent $q \in (0,1/2)$. Since $\{t_k\}$ has zero as an accumulation point, without loss of generality, we can assume that $\{t_k\}$ converges to zero as k goes to infinity. Combining (3-9) with (3-4) we have,

$$\phi(x^k) - \phi(x^{k+1}) \geq \frac{\beta t_k}{c^2M^2(1-q)^2}\left(\phi(x^k) - \phi(\bar{x})\right)^{2q}, \qquad k = 0,1,\ldots.$$

Setting $\Psi(x) := \phi(x) - \phi(\bar{x})$, we obtain

$$\frac{\Psi(x^k) - \Psi(x^{k+1})}{\Psi(x^k)^{2q}} \geq \frac{\beta t_k}{c^2M^2(1-q)^2}$$

and, for $l > 1$, summing on both sides of the last inequality from an arbitrary index k, we have

$$\sum_{j=k}^{k+l}\frac{\Psi(x^j) - \Psi(x^{j+1})}{\Psi(x^j)^{2q}} \geq \frac{\beta}{c^2M^2(1-q)^2}\sum_{j=k}^{k+l} t_j.$$

Because $q \in (0,1/2)$, using Lemma 2.5 with $a_k = \Psi(x^k)$, it follows that

$$\frac{1}{1-2q}\left(\Psi(x^k)^{1-2q} - \Psi(x^{k+l+1})^{1-2q}\right) \geq \frac{\beta}{c^2M^2(1-q)^2}\sum_{j=k}^{k+l} t_j,$$

Hence, letting l goes to infinity, the last inequality yields

$$\Psi(x^k)^{1-2q} \geq \frac{\beta(1-2q)}{c^2M^2(1-q)^2}\sum_{j=k}^{+\infty} t_j,$$

which is a contradiction for $q \in (0,1/2)$ given that $\sum_{j=k}^{+\infty} t_j = \infty$. Therefore, the proof of the proposition is complete. ☐

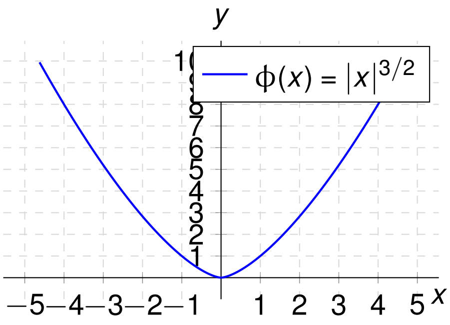
Next, we can see an example that leads to the conclusion that it is not possible to exchange the hypothesis of non-summable step length for summable in Propositions 3.3 and 3.4.

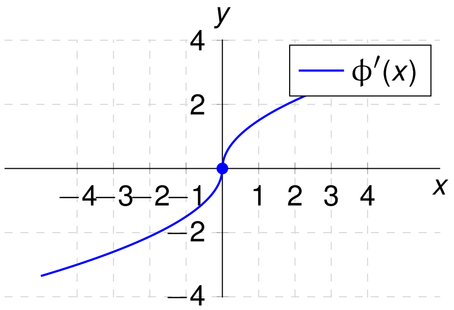
Example 3 $\min_{x\in\mathbb{R}} \phi(x)$, *where* $\phi(x) = |x|^{\frac{3}{2}}$.

This problem has only one critical point, $x^* = 0$, *and*

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

Below, we can see the graphical representation of ϕ *and* ϕ',





Consider the following steepest descent method: Take $x_0 > 0$, *and, for* $k = 0,1,2,\ldots$, *define*

$$x_{k+1} = x_k - t_k\phi'(x_k), \tag{3-10}$$

where $t_k \geq 0$ *is such that*

$$\phi(x_{k+1}) \leq \phi(x_k) - \delta t_k[\phi'(x_k)]^2, \tag{3-11}$$

for some fixed $\delta \in (0,1)$.

Taking $t_k = \frac{2\sqrt{x_k}}{9}$ *and* $\delta = 1 - \left(\frac{2}{3}\right)^{\frac{3}{2}}$, *we get:*

- $x_{k+1} = x_k - t_k\phi'(x_k) = x_k - \frac{3t_k}{2}\sqrt{x_k} = \sqrt{x_k}\left(\sqrt{x_k} - \frac{3t_k}{2}\right) = \frac{2}{3}x_k$;
- $\phi(x_k) - \delta t_k[\phi'(x_k)]^2 = x_k^{\frac{3}{2}} - \delta t_k\frac{9}{4}x_k = x_k\left[\sqrt{x_k} - \delta\frac{9t_k}{4}\right] = x_k^{\frac{3}{2}}\left(1 - \frac{\delta}{2}\right) \geq \left(\frac{2}{3}x_k\right)^{\frac{3}{2}} = \phi(x_{k+1})$.

Conclusion: *The sequence* $\left\{\left(\frac{2}{3}\right)^k x_0\right\}$, *where* $x_0 > 0$, *is generated by the algorithm* (3-10)-(3-11) *with parameter* $\delta = 1 - \left(\frac{2}{3}\right)^{\frac{3}{2}}$ *when we set* $t_k = \frac{2\sqrt{x_k}}{9}$. *This sequence has infinitely many terms.*

Similarly, we can conveniently choose the initial point and parameters so that the sequence converges in finite time. Just take $x_0 = 1/4$, $\delta = \frac{2}{3}$ *and* $t_1 = \frac{1}{2}$, *then* $x_1 = 0$.

Remark 2 *It is known that the classical Łojasiewicz gradient inequality always holds for any* analytic *function* $\phi : \mathbb{R}^n \to \mathbb{R}$ *with some exponent* $q \in [1/2, 1)$; *see, for instance, [7]. Although it does not exclude the fulfillment of the PŁK property with lower exponents, recently the authors in [12, Theorem 4] showed, in particular, that this property fails for local minimizers of smooth functions with Lipschitzian gradients, i.e., the class* $\mathscr{C}^{1,1}$. *It is worth highlighting that this fact has also been observed by the authors in [42, Remark 2.21]. By considering the univariate function* $\phi_{1/2}(x) = |x|^{3/2}$, *at [12, Example 1] the authors observed that in fact the* $\mathscr{C}^{1,1}$ *property is essential for inconsistency with lower exponent PŁK. Note that* $\phi_\alpha(x) = |x|^{1+\alpha}$, $0 < \alpha < 1$, *has also global minimizer* $\bar{x} = 0$, *is* C^1 *and satisfies the PŁK property with* $q \in (0, 1/2)$. *Hence, by taking into account the content of Propositions 3.3 and 3.4, either the sequence* $\{x^k\}$ *satisfying* (3-1) *is finite or* $\{t_k\}$ *is summable.*

CHAPTER 4

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

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

In [8], Attouch, Bolté and Svaiter studied minimization of l.s.c. functions $\phi : \mathbb{R}^n \to \overline{\mathbb{R}}$ by using a *generic* class of *descent methods* satisfying the following properties:

($\mathscr{H}1$) *Sufficient decrease*: for each $k \in \mathbb{N}$, we have

$$\phi(x^{k+1}) + a\|x^{k+1} - x^k\|^2 \le \phi(x^k). \tag{4-1}$$

($\mathscr{H}2$) *Relative error*: for each $k \in \mathbb{N}$, there exists

$$w^{k+1} \in \partial\phi(x^{k+1}) \text{ with } \|w^{k+1}\| \le b\|x^{k+1} - x^k\|, \tag{4-2}$$

with $a, b > 0$. It is shown in [8] that a great variety of important algorithms of optimization satisfy conditions ($\mathscr{H}1$) and ($\mathscr{H}2$). Employing the basic PŁK property from Definition 2.7(i), a qualitative convergence analysis, with arriving at the *limiting/(M)ordukhovich-stationary point* as $0 \in \partial\phi(\bar{x})$, is developed in [8] for the general class of descent algorithms satisfying ($\mathscr{H}1$) and ($\mathscr{H}2$) while without establishing any convergence rate.[1] A detailed qualitative (with convergence rates) analysis of the

---

[1] In fact, an additional technical and nonrestrictive assumption, labeled as ($\mathscr{H}4$) in what follows, was imposed in [8] and the subsequent publications discussed below.

generic class of descent algorithms under the exponent PŁK conditions married to ($\mathscr{H}1$) and ($\mathscr{H}2$) was developed in [25], where the the obtained results were similar to [5] established for the proximal algorithm; namely, the finite step convergence if $q = 0$, linear convergence if $q \in (0, 1/2]$, and polynomial convergence if $q \in (1/2, 1)$.

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

$$\text{there exists } \ w^k \in \overline{\partial}\phi(x^k) \ \text{ with } \ \|w^k\| \le b\|x^{k+1} - x^k\|, \quad k \in \mathbb{N}, \tag{4-3}$$

expressed in terms of the convexified subdifferential (2-7). To furnish this, it is assumed in [4] that the functions g and h in the DC decomposition $\phi = g - h$ are strongly convex with g being of class $\mathscr{C}^{1,1}$, and that the basic PŁK condition in (2-4) is replaced by its strong version from Definition 2.7(iii). Under these requirements, [4] establishes the convergence of BDCA iterates to a (C)*larke-stationary point* $0 \in \overline{\partial}\phi(\bar{x})$.

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

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

$$w^k \in \partial\phi(x^k) \ \text{ with } \ \|w^k\| \le b\|x^{k+1} - x^k\| \tag{4-4}$$

expressed in terms of the limiting subdifferential (2-5). The convergence analysis conducted below allows us to achieve, under the basic PŁK property, the global convergence of any algorithm satisfying ($\mathscr{H}1$) and ($\mathscr{H}3$) to an M-stationary point $0 \in \partial\phi(\bar{x})$. This answers in the affirmative the question posted in [4, Remark 4.5], even with the improvement (4-4) of the stronger condition in (4-3) for general l.s.c. functions.

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

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

4.1 Global Convergence of Generic Algorithms

This section studies in parallel the two classes of descent algorithms satisfying the generic conditions in **(a)**: $(\mathscr{H}1)$ and $(\mathscr{H}2)$, and in **(b)**: $(\mathscr{H}1)$ and $(\mathscr{H}3)$, respectively. Similarly to [8], we add to our convergence analysis the following technical assumption:

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

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

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

The major theorem below tells us that the basic PŁK property from Definition 2.7 ensures the global convergence of generic algorithms in both cases (a) and (b). In Remark 3, we discuss the corresponding modifications of the convergence results under the symmetric and strong PŁK conditions formulated in Definition 2.7(ii,iii).

Theorem 4.1 (global convergence of generic algorithms under the basic PŁK) *Let $\phi : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be a proper l.s.c. function bounded from below, and let the sequence $\{x^k\}$ be constructed by the generic algorithms satisfying either $(\mathscr{H}1)$ and $(\mathscr{H}2)$, or $(\mathscr{H}1)$ and $(\mathscr{H}3)$ properties. If the basic PŁK condition holds at some accumulation point $\bar{x} \in \operatorname{dom}\phi$ of $\{x^k\}$, then we have*

$$\sum_{k=0}^{\infty} \left\|x^k - x^{k+1}\right\| < \infty, \tag{4-5}$$

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

Proof. The result under $(\mathscr{H}1)$ and $(\mathscr{H}2)$ is known from [8, Theorem 2.9], and thus we proceed with the proof in the case where $(\mathscr{H}1)$ and $(\mathscr{H}3)$ hold. Since $\bar{x}$ is an accumulation point of $\{x^k\}$, there exists a subsequence $\{x^{k_j}\}$ of $\{x^k\}$ converging to $\bar{x}$ as $j \to \infty$. By the boundedness from below of ϕ and the decreasing of $\{\phi(x^k)\}$, the continuity condition $(\mathscr{H}4)$ implies that $\{\phi(x^k)\}$ converges to $\phi(\bar{x})$ as $k \to \infty$ with $\phi(\bar{x}) < \phi(x^k)$, $k \in \mathbb{N}$. In particular, we get

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

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

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

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

$$\left\|x^{k_0} - \bar{x}\right\| + (a^{-1}(\phi(x^{k_0}) - \phi(\bar{x})))^{\frac{1}{2}} + ba^{-1}\varphi\left(\phi(x^{k_0}) - \phi(\bar{x})\right) < \epsilon, \tag{4-6}$$

$$\phi(\bar{x}) < \phi(x^{k_0}) < \phi(\bar{x}) + \eta \tag{4-7}$$

are satisfied and yield in turn the inclusions

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

Moreover, conditions (4-1) and (4-7) ensure that

$$0 \leq \phi(x^{k_0+1}) - \phi(\bar{x}) \leq \phi(x^{k_0}) - \phi(\bar{x}) < \eta, \tag{4-8}$$

which implies therefore the inequality

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

by the concavity of φ. Combining (4-1) and (4-4) with $k := k_0$, we have

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

The last two inequalities above lead us to

$$\begin{aligned} &\varphi\left(\phi(x^{k_0}) - \phi(\bar{x})\right) - \varphi\left(\phi(x^{k_0+1}) - \phi(\bar{x})\right) \\ &\geq \frac{a}{b}\varphi'\left(\phi(x^{k_0}) - \phi(\bar{x})\right)\left\|w^{k_0}\right\|\left\|x^{k_0} - x^{k_0+1}\right\| \end{aligned} \tag{4-9}$$

with $w^{k_0} \in \partial\phi(x^{k_0})$, and so the basic PŁK property of ϕ at $\bar{x}$ yields $0 \notin \partial\phi(x^{k_0})$ while ensuring together with (2-4) that

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

Combining the latter with (4-9) justifies the estimates

$$ba^{-1}\left(\varphi(\phi(x^k) - \phi(\bar{x})) - \varphi(\phi(x^{k+1}) - \phi(\bar{x}))\right) \geq \left\|x^k - x^{k+1}\right\|, \quad k = k_0. \tag{4-10}$$

Our next step is to verify the inclusion

$$x^k \in B(\bar{x}, \epsilon) \text{ for all } k \geq k_0. \tag{4-11}$$

To proceed by induction, observe that we have already proved that (4-11) holds for $k = k_0$

and now aim at showing that $x^{k_0+1} \in B(\bar{x}, \epsilon)$. It follows from (4-1) and (4-8) that

$$\left\|x^{k_0} - x^{k_0+1}\right\| \le \left(a^{-1}\left(\phi(x^{k_0}) - \phi(x^{k_0+1})\right)\right)^{\frac{1}{2}} < \left(a^{-1}\left(\phi(x^{k_0}) - \phi(\bar{x})\right)\right)^{\frac{1}{2}}. \tag{4-12}$$

The last estimate along with the triangle inequality gives us

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

Combining this with (4-6) implies that $x^{k_0+1} \in B(\bar{x}, \epsilon)$. Take $j > 1$ and assume that (4-11) holds for all $k = k_0 + 1, \ldots, k_0 + j - 1$. In this case, (4-10) is satisfied for $k = k_0 + 1, \ldots, k_0 + j - 1$. Therefore,

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

Using the triangle inequality again tells us that

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

Now we combine the last two inequalities with (4-12) to get the estimates

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

where the second one follows from (4-8) and the fact that φ is increasing. The latter inequality along with (4-6) ensures that $x^{k_0+j} \in B(\bar{x}, \epsilon)$, which verifies (4-11) and shows that (4-10) holds for all $k \ge k_0$. Thus we arrive at

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

which clearly yields (4-5). This tells us that $\{x^k\}$ is a Cauchy sequence and hence converges to $\bar{x}$ since it is an accumulation point of $\{x^k\}$. It follows from (4-4) and (4-5) that the sequence $\{w^k\}$ converges to zero. By definition (2-5) of the limiting subdifferential $\partial\phi$ and the convergence $\phi(x^k) \to \phi(\bar{x})$ as $k \to \infty$ shown above, we get that $0 \in \partial\phi(\bar{x})$, i.e., $\bar{x}$ is an M-stationary point of ϕ, and the proof is complete. $\square$

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

Remark 3 (convergence under symmetric and strong PŁK conditions)

(i) Theorem 4.1 resolves in the affirmative the question posed in [4, Remark 4.5] about the possibility to provide a counterpart of the convergence analysis in [8] under the modified error bound condition (4-4) in ($\mathscr{H}3$).

(ii) It follows from the proof of Theorem 4.1 that our analysis works with replacing (4-4) by (4-3) under the strong PŁK condition from Definition 2.7(iii) and also by using the symmetric subdifferential (2-6) in (4-4) under the symmetric PŁK property from Definition 2.7(ii). However, such replacements lead us to the less informative *C-stationary* $0 \in \overline{\partial}\phi(\bar{x})$ and *symmetric stationary* points $0 \in \partial^0\phi(\bar{x})$, respectively. Similar conclusions concerning the strong and symmetric PŁK conditions can be made for generic algorithms in the case of ($\mathscr{H}2$). Moreover, the same observation applies to the *convergence rate results* presented in Section 4.2 and Section 4.3 below.

4.2 Convergence Rates under Relative Error Condition

In this section, we provide explicit convergence rates in terms of the PŁK exponents of the desingularizing function $\varphi(t) = t^{1-q}$, $q \in [0,1)$, for generic algorithms that satisfy the conditions in ($\mathscr{H}1$) and ($\mathscr{H}2$). While the convergence rate results obtained under the exponent PŁK property for $q = 0$ and $q \in [1/2, 1)$ are generic extensions of those given for particular algorithms (see, e.g., [2–5,29,31,32,35] and the references therein), it is worth highlighting that the main challenge lies in the case where $q \in (0, 1/2)$ for which we provide significant convergence rate improvements.

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

The next lemma is essential to establish the main results of this section on either finite or superlinear convergence of generic algorithms satisfying the conditions in ($\mathscr{H}1$) and ($\mathscr{H}2$) under the PŁK lower exponents.

Lemma 4.2 (convergence rates for monotone sequences) *Let $\sigma > 0$ and $p \in (0,1)$, and let $\{a_k\} \subset [0,\infty)$ be a monotonically decreasing sequence. Assume that*

$$a_{k+1} \le a_k - \sigma a_{k+1}^p \quad k \in \mathbb{N}. \tag{4-13}$$

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

(i) *The sequence $\{a_k\}$ converges superlinearly to zero.*

(ii) *There exist* $k_0 \in \mathbb{N}$ *such that* $a_k = 0$ *for all* $k \geq k_0$.

Proof. Since $a_k \geq 0$ for all $k \in \mathbb{N}$, we have $\bar{a} \geq 0$ be the limit of this sequence. On the other hand, the passage to the limit in (4-13) with taking into account that $\sigma > 0$ yields

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

which implies that $\bar{a} = 0$. Fix $p \in (0,1)$ and observe that if there exists $k_0 \in \mathbb{N}$ such that $a_{k_0} = 0$, then (ii) holds. Suppose now that there exists no $k_0 \in \mathbb{N}$ with $a_{k_0} = 0$, i.e., $a_k > 0$ for all $k \in \mathbb{N}$. Then it follows from (4-13) that

$$a_{k+1} + \sigma a_{k+1}^p \leq a_k, \quad k \in \mathbb{N}. \tag{4-14}$$

Remembering that $a_k > 0$ for all $k \in \mathbb{N}$, we divide both sides of (4-14) by a_{k+1} and take the limit as $k \longrightarrow \infty$ therein. This gives us

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

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

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

Theorem 4.3 **(value convergence rates under exponent PŁK and relative error conditions)** *In the setting of Theorem 4.1, let* $\{x_k\}$ *be a generic sequence of iterates satisfying* $(\mathscr{H}1)$ *and* $(\mathscr{H}2)$*, and let the exponent PŁK property of* ϕ *hold at* $\bar{x}$ *with* $\varphi(t) = Mt^{1-q}$ *for some* $M > 0$ *and* $q \in [0,1)$*. The following convergence rates are guaranteed for the value sequence* $\{\phi(x^k)\}$ *:*

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

$$\phi(x^k) - \phi(\bar{x}) \leq \eta k^{-\frac{1}{2q-1}} \quad \textit{for all large } k \in \mathbb{N}. \tag{4-15}$$

Proof. It follows from Theorem 4.1 that $x^k \to \bar{x}$ as $k \to \infty$ and that $\bar{x}$ is an M-stationary point of ϕ. Defining $\psi(x) := \phi(x) - \phi(\bar{x})$, we see that the sequence $\{\psi(x^k)\} \subset [0,\infty)$ is decreasing. Using (2-4) with $\varphi(t) = Mt^{1-q}$ as $q \in [0,1)$ and $M > 0$, it follows that

$$\begin{aligned} &\|w^{k+1}\| \geq \operatorname{dist}\big(0, \partial\phi(x^{k+1})\big) \geq \frac{1}{M(1-q)}\big(\phi(x^{k+1}) - \phi(\bar{x})\big)^q \\ &\text{with } \ w^{k+1} \in \partial\phi(x^{k+1}), \quad k \in \mathbb{N}. \end{aligned} \tag{4-16}$$

To verify (i), suppose on the contrary that the sequence $\{x^k\}$ is infinitely generated and deduce from (4-16) with $q = 0$ that $\|w^{k+1}\| \geq M^{-1}$ for all large k. This tells us together with (4-2) in ($\mathscr{H}2$) that $(Mb)^{-1} \leq \|x^{k+1} - x^k\|$, which contradicts the fact that $\{x^k\}$ converges to $\bar{x}$ and thus justifies the statement in (i).

For the proof of (ii), subtract $\phi(\bar{x})$ from both sides of (4-1) and get

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

which being combined with (4-2) yields

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

Using the obtained inequality together with (4-16) and the definition of ψ gives us

$$\psi(x^{k+1}) \leq \psi(x^k) - \frac{a}{b^2M^2(1-q)^2}\big(\psi(x^{k+1})\big)^{2q}. \tag{4-17}$$

It follows from Lemma 4.2 with $\sigma = \frac{a}{b^2M^2(1-q)^2}$ and $p = 2q$ that if $q \in (0, 1/2)$, then the sequence $\{\psi(x^k)\}$ is either finite or converges superlinearly, and thus (ii) is verified.

To prove (iii), let $q = 1/2$ and $\alpha = \frac{4a}{b^2M^2}$. By (4-17) we get

$$\psi(x^{k+1}) \leq \frac{1}{1+\alpha}\psi(x^k). \tag{4-18}$$

Since $1/(1+\alpha) \in (0,1)$, assertion (iii) is a consequence of (4-18).

It remains to verify (iv). For $q \in (1/2, 1)$, it follows from (4-17) with $\beta := \dfrac{a}{b^2M^2(1-q)^2}$ that we have

$$\beta \leq \frac{\psi(x^k) - \psi(x^{k+1})}{\psi(x^{k+1})^{2q}} = [\psi(x^k) - \psi(x^{k+1})]\Phi(\psi(x^{k+1})), \tag{4-19}$$

where $\Phi(t) := \dfrac{1}{t^{2q}}$. Suppose first that there exists $Q \in (1, \infty)$ such that $\Phi(\psi(x^{k+1})) \leq$

$Q\Phi(\psi(x^k))$. Combining this fact with (4-19) and Lemma 2.5 brings us to

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

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

$$\frac{\beta(2q-1)}{Q} \le \left(\psi(x^{k+1})\right)^{1-2q} - \left(\psi(x^k)\right)^{1-2q}, \quad k \ge k_0. \tag{4-20}$$

Letting $j-1>k_0$, we deduce from (4-20) that

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

giving us in turn that

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

Therefore, we arrive at the estimate

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

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

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

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

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

and thus verifies the convergence rate estimate in (4-15) in this case.

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

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

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

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

whenever k is sufficiently large. The rest of the proof is similar to the above arguments based on (4-20), and thus we are done with verifying (iv) and the whole theorem. ☐

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

Theorem 4.4 (convergence rates of iterates under exponent PŁK and relative error conditions) *In the setting of Theorem 4.3, the following assertions hold:*

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

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

Proof. By taking into account that

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

assertion (i) of this theorem is an immediate consequence of Theorem 4.3(i). To verify (ii), deduce first from the concavity of φ in (2-4) together with $(\mathscr{H}1)$ and $(\mathscr{H}2)$ that

$$\frac{\|x^{k+1} - x^k\|^2}{\|x^k - x^{k-1}\|} \leq \frac{b}{a}\left[\varphi\left(\phi(x^k) - \phi(\bar{x})\right) - \varphi\left(\phi(x^{k+1}) - \phi(\bar{x})\right)\right]. \tag{4-21}$$

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

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

In case (a), the usage of (4-21) with $\varphi(t) = Mt^{1-q}$ and $\psi(x) = \phi(x) - \phi(\bar{x})$ yields

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

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

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

and therefore arrive at the inequality

$$\|x^{k+1}-x^k\| \leq r\|x^k-x^{k-1}\| + \frac{bM}{ar}\big(\psi(x^k)^{1-q}-\psi(x^{k+1})^{1-q}\big). \tag{4-22}$$

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

$$\sum_{j=k}^{k+l}\|x^{j+1}-x^j\| \leq \frac{r}{(1-r)}\|x^k-x^{k-1}\| + \frac{bM}{ar(1-r)}\big(\psi(x^k)^{1-q}-\psi(x^{k+l+1})^{1-q}\big). \tag{4-23}$$

in fact, let $l \geq 1$, from (4-22) we have

$$\begin{aligned}
\sum_{j=k}^{k+l}\|x^{j+1}-x^j\| &\leq r\sum_{j=k}^{k+l}\|x^j-x^{j-1}\| + \frac{bM}{ar}\big(\Psi(x^k)^{1-q}-\Psi(x^{k+l+1})^{1-q}\big)\\
&= r\|x^{k-1}-x^k\| + r\sum_{j=k}^{k+l-1}\|x^{j+1}-x^j\| + \frac{bM}{ar}\big(\Psi(x^k)^{1-q}-\Psi(x^{k+l+1})^{1-q}\big)\\
&\leq r\|x^{k-1}-x^k\| + r\sum_{j=k}^{k+l}\|x^{j+1}-x^j\| + \frac{bM}{ar}\big(\Psi(x^k)^{1-q}-\Psi(x^{k+l+1})^{1-q}\big),
\end{aligned}$$

thus,

$$(1-r)\sum_{j=k}^{k+l}\|x^{j+1}-x^j\| \leq r\|x^k-x^{k-1}\| + \frac{bM}{ar}\big(\Psi(x^k)^{1-q}-\Psi(x^{k+l+1})^{1-q}\big),$$

as $r \in (0,1)$, follows (4-23). Now, define $s_k := \sum_{j=k}^{\infty}\|x^{j+1}-x^j\|$ and let $l \to \infty$ in (4-23). Remembering that $\{\psi(x^k)\}$ decreases to zero, we get from (4-23) that

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

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

$$s_k \leq \frac{r}{a^{1/2}(1-r)}\big(\psi(x^{k-1})\big)^{1/2} + \frac{bM}{ar(1-r)}\psi(x^k)^{1-q}. \tag{4-24}$$

When $q \in (0,1/2)$, $\psi(x^k)^{1-q} \leq \psi(x^k)^{1/2}$ (because $\psi(x^k) \longrightarrow 0$) and it follows from the above that

$$s_k \leq \frac{r}{a^{1/2}(1-r)}\big(\psi(x^{k-1})\big)^{1/2} + \frac{bM}{ar(1-r)}\psi(x^k)^{1/2}. \tag{4-25}$$

Since the sequence $\{\psi(x^k)\}$ is monotonically decreasing, we deduce from (4-25) that

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

which allows us to verify (ii) by using Theorem 4.3 and the estimate $\|x^k - \bar{x}\| \leq s_k$. Assertions (iii) and (iv) of this theorem follow directly from Theorem 4.3(iii,iv)), respectively. Indeed, in case (iii) we set $q = 1/2$ in (4-24), while (iv) with $q \in (1/2, 1)$ is deduced from the inequality $\left(\psi(x^k)\right)^{1/2} \leq \left(\psi(x^k)\right)^{1-q}$ held for large k. $\square$

The following example presents a family of *univariate smooth functions* for which the classical proximal algorithm, satisfying ($\mathscr{H}1$) and ($\mathscr{H}2$), does *not terminate in finite steps* under the PŁK conditions with lower exponents.

Example 4 (failure of finite termination for the proximal algorithm under PŁK lower exponents) Consider the parametric class of functions $\phi_\alpha(x) := |x|^{1+\alpha}$ with $0 < \alpha < 1$; cf. [40, page 22], where such functions were considered from different prospectives. Each function ϕ_α is continuously differentiable, achieves its global minimum at $\bar{x} = 0$, is C^1, and satisfies the exponent PŁK condition with $q = \frac{\alpha}{1+\alpha} \in (0, 1/2)$. The classical proximal algorithm reads for ϕ_α as follows:

$$x_{k+1} = \text{argmin}_{x\in\mathbb{R}}\left\{|x|^{1+\alpha} + \frac{1}{2\lambda_k}|x - x_k|^2\right\}, \quad \lambda_k > 0,\ k \in \mathbb{N}. \tag{4-26}$$

If we suppose that $\{x_k\}$ is finite, then there exists $k_0 \in \mathbb{N}$ such that $(1+\alpha)\text{sign}(x_{k_0})|x_{k_0}|^\alpha = 0$ and consequently $x_{k_0} = 0$. Applying the stationary rule to the function in (4-26) gives us the equation

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

which yields $x_{k_0-1} = x_{k_0} = 0$. Recursively, we arrive at $x_0 = 0$ and thus conclude that for any starting point $x_0 \neq 0$ the sequence of iterates in (4-26) must be infinite. Theorem 4.4 tells us that $x_k \to 0$ as $k \to \infty$ superlinearly.

To demonstrate the efficiency of the results obtained in Theorem 4.4 and Theorem 4.3, we present a numerical test below, implemented in Python, for a low-exponent PŁK function from Example 4. Taking $\alpha = 1/2$, and consequently $q = \frac{\alpha}{1+\alpha} = 1/3$, with $\lambda_k = 100$ and $x_0 = 100$, we can analyze, after 10 iterations, how close we are to the solution $\bar{x} = 0$.

k	x_k	$f(x_k)$
0	100.00000000000000000000	1000.00000000000000000000
1	0.44053730994470208238	0.29239776137586043214
2	0.00000814768263762068	0.00000002325686417556
3	0.00000045128287240459	0.00000000030316096199
4	0.00000002499560736169	0.00000000000395180532
5	0.00000000138444988806	0.00000000000005151288
6	0.00000000007668887747	0.00000000000000067158
7	0.00000000000424904556	0.00000000000000000876
8	0.00000000000024158453	0.00000000000000000012
9	0.00000000000001421085	0.00000000000000000000
10	0.00000000000000710543	0.00000000000000000000

The python code is available in Appendix A.

The next example illustrates the alternative situation in Theorem 4.4, where the proximal algorithm *terminates in finite steps* in contrast to Example 4.

Example 5 (finite step termination of the proximal algorithm) Let

$$\phi(x) = \begin{cases} -x, & x < 0, \\ x^{3/2}, & x \geq 0. \end{cases} \tag{4-27}$$

We see that the function ϕ in (4-27) achieves its global minimum at $\bar{x} = 0$, where it is nondifferentiable, and bit satisfies the exponent PŁK property with $q = 1/3 \in (0, 1/2)$ at $\bar{x}$. The proximal algorithm for (4-27) is written as

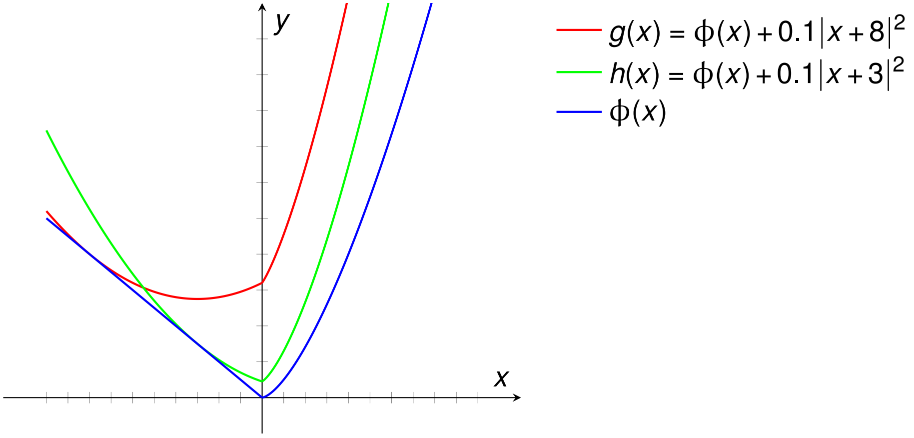
$$x_{k+1} = \text{argmin}_{x \in \mathbb{R}} \left\{ \phi(x) + \frac{1}{2\lambda_k} |x - x_k|^2 \right\}, \quad \lambda_k > 0,\ k \in \mathbb{N}, \tag{4-28}$$

First we claim that if $x_0 < 0$, then $x_k \leq 0$ for all $k \in \mathbb{N}$. On the contrary, suppose that there exists $k_0 \in \mathbb{N}$ for which $x_{k_0} < 0$ while $x_{k_0+1} > 0$. By the smoothness of (4-27) at the points under consideration, we deduce from the stationary rule applied to (4-28) that

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

which is a contradiction that verifies the claim. Letting $\lambda_k := \bar{\lambda} > 0$ for all $k \in \mathbb{N}$ and assuming that the sequence is infinitely generated, we get $x_k < 0$. Applying again the stationary rule to (4-28) in this setting yields $\bar{\lambda} = x_{k+1} - x_k$. which leads us to a contradiction. Therefore, for any fixed $\bar{\lambda} > 0$, the sequence $\{x_k\}$ converges in finite steps if the starting point is selected as $x_0 < 0$.

In the figure below, we can graphically see the example with $x_0 = -8$ and $\lambda_k = 10$ for all natural k.



We finalize this section with the following comments.

Remark 4 (comparison with known results)

(i) The main novelty of Theorem 4.3 and Theorem 4.4 is for the case of lower exponents $q \in (0, 1/2)$, where the obtained results significantly improve the known ones with *linear convergence* for various algorithms satisfying $(\mathscr{H}1)$ and $(\mathscr{H}2)$; see [25, Theorem 3.4] and many other publications in this direction.

(ii) The convergence rate results obtained in Theorems 4.3, 4.4 for the cases where $q = 0$ and $q \in (1/2, 1)$ under the exponent PŁK conditions are closely related to those established in [25, Theorem 3.4] for an implicit abstract model, which encompasses our explicit model defined by $(\mathscr{H}1)$ and $(\mathscr{H}2)$. The proofs of Theorems 4.3, 4.4 are different from those given in [25].

4.3 Convergence Rates under Modified Error Property

In this section, we provide a comprehensive rate convergence analysis of the generic algorithms satisfying $(\mathscr{H}1)$ and $(\mathscr{H}3)$ under the exponent PŁK conditions. It is striking to observe that, in contrast to the results of Section 4.2, the replacement of $(\mathscr{H}2)$ by $(\mathscr{H}3)$ fully *excludes infinite sequences of iterates* in the case of PŁK with lower exponents. Moreover, we show below that, under some additional assumption, the lower exponent PŁK property is *inconsistent* with the conditions in $(\mathscr{H}1)$ and $(\mathscr{H}3)$.

Theorem 4.5 (Convergence rates under exponent PŁK and modified relative error conditions) *In the setting of Theorem 4.1, assume that the exponent PŁK property of ϕ holds at $\bar{x}$ with $\varphi(t) = Mt^{1-q}$ for some $M > 0$ and $q \in [0, 1)$. The following convergence rates are guaranteed for the generic iterative sequences satisfying the conditions in $(\mathscr{H}1)$ and $(\mathscr{H}3)$:*

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

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

Proof. It is shown in the proof of Theorem 4.1 that $\phi(x^k) \to \phi(\bar{x})$ as $k \to \infty$. Defining $\psi(x) := \phi(x) - \phi(\bar{x})$, we see that the sequence $\{\psi(x^k)\} \subset [0, \infty)$ is decreasing, and hence $\psi(x^j) \ge t$ for all $t \in [\psi(x^{j+1}), \psi(x^j)]$, $j \in \mathbb{N}$. Thus

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

which implies in turn that

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

Therefore, we arrive at the inequality

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

which gives us the estimate

$$\sum_{j=k}^{k+l} \frac{\psi(x^j) - \psi(x^{j+1})}{\psi(x^j)^q} \le \frac{1}{1-q}\left(\psi(x^k)^{1-q} - \psi(x^{k+l+1})^{1-q}\right) \ \text{ for all } \ k, l \in \mathbb{N}. \tag{4-29}$$

On the other hand, combining (4-1), the construction of ψ, and (4-4) in ($\mathscr{H}3$) yields

$$\psi(x^{k+1}) \le \psi(x^k) - a\|x^{k+1} - x^k\|^2 \le \psi(x^k) - \frac{a}{b^2}\|w^k\|^2. \tag{4-30}$$

By the PŁK property (2-4) with $\varphi(t) = Mt^{1-q}$, we find $M > 0$ such that

$$\|w^k\| \ge \text{dist}\big(0, \partial\phi(x^k)\big) \ge \frac{1}{M(1-q)}\big(\phi(x^k) - \phi(\bar{x})\big)^q. \tag{4-31}$$

Combining (4-30) with (4-31) tells us that

$$\psi(x^{k+1}) \le \psi(x^k) - \frac{a}{b^2 M^2 (1-q)^2}\big(\psi(x^k)\big)^{2q}. \tag{4-32}$$

To furnish now the proof of assertion (i), consider first the case where $q \in (0, 1/2)$ and suppose, arguing by contradiction, that the sequence $\{x^k\}$ is infinitely generated, which tells us that $\psi(x^k) > 0$ for all $k \in \mathbb{N}$. Recall the convergence $\psi(x^k) \to 0$ as $k \to \infty$ and deduce from $2q - 1 < 0$, (4-32), and the equality

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

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

$$\psi(x^{k_0+1}) \le \psi(x^{k_0})\Big(1 - \frac{a}{b^2M^2(1-q)^2}\big(\psi(x^{k_0})\big)^{2q-1}\Big) < 0,$$

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

To verify (i), it remains to consider the case where $q = 0$. Arguing again by contradiction, suppose that the sequence $\{x^k\}$ is infinitely generated and deduce from (4-31) with $q = 0$ that $\|w^k\| \ge M^{-1}$ for all large k, which tells us together with (4-4) in ($\mathscr{H}3$) that $(Mb)^{-1} \le \|x^{k+1} - x^k\|$. This contradicts the fact that $\{x^k\}$ converges to $\bar{x}$ and thus concludes the proof of assertion (i).

To verify assertion (ii), we get from (4-32) with $q = 1/2$ that

$$\psi(x^{k+1}) \le \Big(1 - \frac{4a}{b^2M}\Big)\psi(x^k). \tag{4-33}$$

Due to $\psi(x^k) \ge 0$ as $k \in \mathbb{N}$, it follows from (4-33) that $4a(b^2M)^{-1} \in (0, 1)$, and hence $\{\phi(x^k)\}$ converges linearly to $\phi(\bar{x})$. On the other hand, we have by (4-1) and (4-4) that

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

which being combined with (4-31) and the construction of ψ yields

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

The obtained inequality readily ensures the estimate

$$\|x^{k+1} - x^k\| \le \frac{Mb}{2a}\Big(\frac{\psi(x^k) - \psi(x^{k+1})}{\psi(x^k)^{\frac{1}{2}}}\Big),$$

which tells us together with (4-29) that

$$\begin{aligned} \sum_{j=k}^{k+l}\|x^{j+1} - x^j\| &\le \frac{Mb}{2a}\sum_{j=k}^{k+l}\frac{\psi(x^j) - \psi(x^{j+1})}{\psi(x^j)^{\frac{1}{2}}} && \text{(4-34)} \\ &\le \frac{Mb}{a}\big(\psi(x^k)^{\frac{1}{2}} - \psi(x^{k+l+1})^{\frac{1}{2}}\big). && \text{(4-35)} \end{aligned}$$

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

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

which being combined with (4-31) and (4-4) brings us to the estimates

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

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

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

from which it clearly follows that

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

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

Theorem 4.6 (Inconsistency of lower exponent PŁK with ($\mathscr{H}1$) and ($\mathscr{H}3$)) *In addition to the general assumption of Theorem 4.5, suppose that $x^{k+1} \ne x^k$ for all $k \in \mathbb{N}$. Then the exponent PŁK property of f at $\bar{x}$ fails whenever $q \in (0, 1/2)$.*

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

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

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

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

Combining the last inequality with the the exponent PŁK condition yields

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

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

$$\big(\phi(x^k)-\phi(\bar{x})\big)^{1-2q}\geq\beta,\ \text{ where }\ \beta:=\frac{a}{(1-q)^2M^2b^2}. \tag{4-36}$$

Hence $\phi(x^k)-\phi(\bar{x})$ as $k\to\infty$, (4-36) fails for large k when $q\in(0,1/2)$. The obtained contradiction verifies the claimed assertion and completes the proof of theorem. □

Remark 5 (novelty under the modified relative error condition)

- **(i)** Similarly to the case of the relative error condition ($\mathscr{H}2$) discussed in Remark 4, the convergence rate results obtained in Theorem 4.5 under the modified version ($\mathscr{H}3$) and the exponent PŁK properties with $q=0$ and $q\in[1/2,1)$ continue to be generic extensions of those given for particular algorithms (see, e.g., [2–5, 25, 28, 31, 35]).
- **(ii)** In the case of *continuous-time* gradient/subgradient systems, a *finite-time* convergence estimate has already appeared in the literature when the exponent PŁK varies within the interval (0,1/2); see [19, Theorem 2.7]. Inspired by the latter result, it is shown in [25, Theorem 3.5(i)] that for certain algorithms including gradient-related methods, this better estimate remains true at least when dealing with the sequence of functional values. Our approach to this issue is novel being also in accordance with the estimates presented in the continuous case.
- **(iii)** To the best of our knowledge, the *inconsistency* result in the setting of Theorem 4.6 has never been observed before. Its analogs in the different setting are discussed in Section 4.5; see Theorem 4.9 and the discussion after its proof.

4.4 Specifications of Generic Descent Algorithms

This section discusses some specifications of the general class of descent methods and convergence analysis for them developed in Sections 4.1 and Section 4.3.

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

4.4.1 Inexact Reduced Graduate Methods

A class of *inexact reduced graduate* (*IRG*) *methods* with various stepsize selections have been recently proposed and developed by Khanh et al. [28] for problems of smooth nonconvex optimization. The model unifying the IRG methods in [28] considers the objective function $\phi:\mathbb{R}^n\to\mathbb{R}$ and the iterative sequence $\{x^k\}$ as $k\in\mathbb{N}$ satisfying

$$\phi(x^k)-\phi(x^{k+1})\geq\frac{\beta}{t_k}\|x^{k+1}-x^k\|^2\ \text{ and }\ \|\nabla\phi(x^k)\|\leq\frac{c}{t_k}\|x^{k+1}-x^k\|, \tag{4-37}$$

where $t_k \subset \mathbb{R}_+$ and $\beta, c > 0$. This can be viewed as a smooth version of the general nonsmooth scheme investigated above. The convergence analysis of IRG algorithms conducted in [28] is based on the observation that the sequences of iterates generated by (4-37) satisfy the following properties:

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

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

- *Complementary descent condition*: we have the implication

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

In fact, the latter conditions, while not (4-37), are used by Absil et al. in [1, page 536] to prove the convergence of the corresponding iterative sequences when the cost function is real analytical. Let us emphasize that the general scheme (4-37) comprises *new linesearch methods* with an *inexact gradient information* for finding stationary points of $\mathscr{C}^1$-smooth functions by using different choices of stepsize rules. To the best of our knowledge, [28] is a first paper dealing with convergence rates for methods satisfying condition (4-37). In particular, it extends some special settings considered, e.g., in [41] and [27], where the authors analyze convergence properties of the exact gradient descent method for functions satisfying the PŁK conditions. More precisely, [28, Theorem 2.5] extends the scope of applicability of the results in the aforementioned papers. The results established above allow us to essentially improve those obtained in [28] for IRG methods and their various linesearch specifications. In particular, we now have the *finite termination* of such algorithms under (4-37) and the exponent PŁK with $q \in [0, 1/2)$ at any *accumulation point* of the iterative sequences *provided* that the corresponding stepsize sequence $\{t_k\}$ is *bounded away* from zero. The latter assumption is required to accommodate ($\mathscr{H}3$) with a fixed $b > 0$ in (4-4).

4.4.2 Boosted DC Algorithm

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

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

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

Algorithm 1 BDCA

```
Step 1: Fix $\alpha > 0, \bar{\lambda} > 0$ and $\beta \in (0,1)$. Let $x_0$ be any initial point and set $k := 0$.
Step 2: Select $u_k \in \partial h(x^k)$ and solve the strongly convex optimization problem
    $(\mathscr{P}'_k) \quad \min_{x \in \mathbb{R}^n} g(x) - \langle u_k, x \rangle$
  to obtain its unique solution $y^k$.
Step 3: Let $d^k := y^k - x^k$. If $d^k = 0$, **stop** and **return** $x^k$. Otherwise, go to Step 4.
Step 4: Choose any $\bar{\lambda}_k \geq 0$. Set $\lambda_k := \bar{\lambda}_k$. **While** $\phi(y^k + \lambda_k d^k) > \phi(y^k) - \alpha\lambda_k^2 \|d^k\|^2$,
  **do** $\lambda_k := \beta\lambda_k$.
Step 5: Let $x^{k+1} := y^k + \lambda_k d^k$. If $x^{k+1} = x^k$, **stop** and **return** $x^k$. Otherwise, set $k := k+1$
  and go to Step 2.
```

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

Next we show that BDCA can be viewed as a specification, for non-Lipschitzian *continuous* functions ϕ, of the generic algorithm developed in Section 4.1 with using the *symmetric subdifferential* (2-6) and the *symmetric PŁK property* of ϕ. This approach allows us to improve and extend the results of [4] about the convergence and convergence rates of BDCA to the case of continuous functions. First we recall some properties of the symmetric subdifferential of continuous functions used below; all of them and much more on (2-6) (including full calculus) can be found in [36, 37].

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

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

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

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

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

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

Theorem 4.7 **(BDCA for continuous functions)** *Consider problem* $(\mathscr{P})$, *where* ϕ *is continuous on its domain. Let the sequence* $\{x^k\}$ *be generated by BDCA, and let* ∇g *be Lipschitz continuous with modulus* L *around an accumulation point* $\bar{x}$ *of* $\{x^k\}$. *Then for all large* $k \in \mathbb{N}$ *we have the conditions:*

$$\phi(x^{k+1}) \le \phi(x^k) - \frac{\alpha\lambda_k^2 + \rho}{(1+\lambda_k)^2}\|x^{k+1} - x^k\|^2, \tag{4-38}$$

$$\textit{there exists } \; w^k \in \partial^0\phi(x^k) \; \textit{ such that } \; \|w^k\| \le L\|x^{k+1} - x^k\|, \tag{4-39}$$

and $\{x^k\}$ *converges to* $\bar{x}$ *as* $k \to \infty$, *which is a symmetric stationary point of* $(\mathscr{P})$ *satisfying* $0 \in \partial^0\phi(\bar{x}) = \nabla g(\bar{x}) - \partial h(\bar{x})$. *If furthermore* ϕ *enjoys the exponent PŁK property with* $\varphi(t) = Mt^{1-q}$ *for some* $M > 0$ *and* $q \in [0,1)$, *then the convergence rates of* $x^k \to \bar{x}$ *are as given in Theorem 4.5.*

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

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

from which we deduce the estimate

$$\|d^k\| = \frac{1}{1+\lambda_k}\|x^{k+1} - x^k\| \le \|x^{k+1} - x^k\|. \tag{4-40}$$

Hence the estimate in (4-38) follows from the conditions

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

obtained for BDCA in [4, Proposition 3.1]. To verify further (4-39), we employ properties $(\mathscr{A}_1) - (\mathscr{A}_4)$ of the symmetric subdifferential and get the relationships

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

Recalling that y^k is the unique solution to the subproblem $(\mathscr{P}'_k)$ in BDCA, i.e., $u^k = \nabla g(y^k)$ ensures that $w^k := \nabla g(x^k) - \nabla g(y^k) \in \partial^0\phi(x^k)$. Take now a neighborhood U of $\bar{x}$ on which ∇g is Lipschitz continuous with modulus L. It follows from (4-40) that

$x^k, y^k \in U$ for all large $k \in \mathbb{N}$, and hence

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

which yields (4-39). Since $\bar{x}$ is an accumulation point of $\{x^k\}$ and ϕ is continuous on its domain, we deduce from (4-38) and (4-39) by the proof of Theorem 4.1 that $x_k \to \bar{x}$ as $k \to \infty$ with $0 \in \partial^0\phi(\bar{x})$. The latter implies by $(\mathscr{A}_1) - (\mathscr{A}_4)$ that $0 \in \nabla g(\bar{x}) - \partial h(\bar{x})$. The convergence rates under the exponent PŁK conditions follows from Theorem 4.5. ☐

Note that the finite termination of BDCA at any accumulation point under the exponent PŁK condition with $q \in (0, 1/2)$ is a significant improvement of [4] even for Lipschitz continuous objectives, where merely linear convergence of $\{x^k\}$ is justified.

4.5 Lower Exponent PŁK for Difference Programs

In this section, we take a close look at the exponent PŁK property from Definition 2.7(iv) with *lower exponents* $q \in (0, 1/2)$ for a class of *difference programs*.

As shown in [24, Theorem 1][2], the classical PL gradient inequality always holds for any *analytic* function $\phi\colon \mathbb{R}^n \to \mathbb{R}$ with some exponent $q \in [1/2, 1)$. Although it does not exclude the fulfillment of the PŁK conditions with lower exponents, we show below that this property *fails* for *local minimizers* of smooth functions with *Lipschitzian gradients*, i.e., of class $\mathscr{C}^{1,1}$, without any connections with decent algorithms satisfying $(\mathscr{H}1)$ and $(\mathscr{H}3)$ as in Theorem 4.6. In fact, our analysis provides such a result for problems of *difference programming* of type $(\mathscr{P})$ (see Subsection 4.4.2) with $\phi = g - h$, where g is of class $\mathscr{C}^{1,1}$ (may not be convex), while $h\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is l.s.c. and convex. Recall that for any $x \in \operatorname{int}(\operatorname{dom} h)$, the directional derivative of ϕ is represented by

$$\phi'(x, d) = \langle \nabla g(x), d\rangle - \max_{v \in \partial h(x)} \langle v, d\rangle \text{ whenever } d \in \mathbb{R}^n. \tag{4-41}$$

Let us introduce the two important *measures of stationarity* at the point $x \in \operatorname{int}(\operatorname{dom} h)$ as follows. Observe first from (4-41) that

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

---

[2] The authors are grateful to Guoyin Li for drawing our attention to paper [24].

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

$$v_1(x) \in \operatorname{argmax}_{v\in\partial h(x)} \|\nabla g(x) - v\|$$

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

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

In other words, we have the equality

$$\langle \nabla g(x) - v_1(x), v_1(x)\rangle = \min_{v\in\partial h(x)} \langle \nabla g(x) - v_1(x), v\rangle. \tag{4-42}$$

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

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

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

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

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

$$L(x_0) = \inf\{L \mid \|\nabla g(x) - \nabla g(y)\| \leq L\|x - y\| \ \text{ over } \ x, y \in \mathscr{L}(x_0)\}. \tag{4-43}$$

The next lemma measures the difference between $\phi(x_0)$ and the global minimum of ϕ with respect to the level set $\mathscr{L}(x_0)$ in terms of the exact Lipschitzian bound (4-43) and the stationarity measure $\sigma_1(x_0)$.

Lemma 4.8 (Distance estimate from global minimum) *Let* $\bar{x} \in \operatorname{int}(\operatorname{dom} h)$ *be a global minimizer of the function* ϕ *in* ($\mathscr{P}$) *with respect to the level set* $\mathscr{L}(x_0)$, *i.e.,*

$$\phi(x) \geq \phi(\bar{x}) \ \textit{ for all } \ x \in \mathscr{L}(x_0).$$

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

$$\phi(x_0) - \phi(\bar{x}) \geq \phi(x_0) - \phi(x_0 - \alpha d(x_0)) \geq \frac{1}{2}\alpha\big(2 - \alpha L(x_0)\big)\sigma_1^2(x_0). \tag{4-44}$$

In particular, one has

$$\phi(x_0) - \phi(\bar{x}) \geq \frac{\sigma_1^2(x_0)}{2L(x_0)}.$$

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

$$\begin{aligned} \xi'(0,1) &= \phi'\big(x_0, -d(x_0)\big) = \langle \nabla g(x_0), -d(x_0)\rangle - \max_{v\in\partial h(x_0)} \langle v, -d(x_0)\rangle \\ &= \langle \nabla g(x_0), -d(x_0)\rangle + \min_{v\in\partial h(x_0)} \langle v, d(x_0)\rangle \overset{(4-42)}{=} -\sigma_1(x_0) < 0. \end{aligned}$$

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

$$\begin{aligned} \xi'(\alpha,1) &= \langle \nabla g(x_\alpha), -d(x_0)\rangle - \max_{v\in\partial h(x_\alpha)} \langle v, -d(x_0)\rangle \\ &\leq \langle \nabla g(x_\alpha), -d(x_0)\rangle - \max_{v\in\partial h(x_0)} \langle v, -d(x_0)\rangle \\ &\overset{(4-42)}{=} \langle \nabla g(x_\alpha), -d(x_0)\rangle + \langle v_1(x_0), d(x_0)\rangle \\ &= \langle -\nabla g(x_\alpha), \nabla g(x_0) - v_1(x_0)\rangle + \langle v_1(x_0), \nabla g(x_0) - v_1(x_0)\rangle \\ &= \langle -\nabla g(x_\alpha) + \nabla g(x_0) - \nabla g(x_0), \nabla g(x_0) - v_1(x_0)\rangle + \langle v_1(x_0), \nabla g(x_0) - v_1(x_0)\rangle \\ &= \langle -\nabla g(x_\alpha) + \nabla g(x_0), \nabla g(x_0) - v_1(x_0)\rangle - \langle \nabla g(x_0) - v_1(x_0), \nabla g(x_0) - v_1(x_0)\rangle \\ &\leq \| -\nabla g(x_\alpha) + \nabla g(x_0)\| \|\nabla g(x_0) - v_1(x_0)\| - \|\nabla g(x_0) - v_1(x_0)\|^2 \\ &= \| -\nabla g(x_\alpha) + \nabla g(x_0)\| \sigma_1(x_0) - \sigma_1^2(x_0) \\ &\leq L(x_0)\|x_0 - x_\alpha\|\sigma_1(x_0) - \sigma_1^2(x_0) \\ &= L(x_0)\|x_0 - (x_0 - \alpha d(x_0))\|\sigma_1(x_0) - \sigma_1^2(x_0) \\ &= L(x_0)\|\alpha(\nabla g(x_0) - v_1(x_0))\|\sigma_1(x_0) - \sigma_1^2(x_0) \\ &= \alpha L(x_0)\sigma_1^2(x_0) - \sigma_1^2(x_0) \\ &= \big(\alpha L(x_0) - 1\big)\sigma_1^2(x_0). \end{aligned}$$

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

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

As $\xi(\alpha) = \phi(x_\alpha)$, $\xi(0) = \phi(x_0)$ and $x_\alpha \in \mathscr{L}(x_0)$, the last inequality inplies

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

hence, we obtain (4-44). Because $x_{\hat{\alpha}} \in \mathscr{L}(x_0)$, taking $\alpha = \hat{\alpha}$ in (4-44), it follows that

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

and thus completes the proof of the lemma. □

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

Theorem 4.9 ***(Inconsistency of the lower exponent PŁK property with Lipschitz continuity of gradients)*** *Let $\bar{x} \in \operatorname{int}(\operatorname{dom} h)$ be a local minimizer of problem ($\mathscr{P}$), where g is of class $\mathscr{C}^{1,1}$ around $\bar{x}$, and where h is convex. Then the exponent PŁK property of ϕ at $\bar{x}$ fails whenever $q \in (0, 1/2)$.*

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

$$\sigma(x) \geq M[\phi(x) - \phi(\bar{x})]^q, \tag{4-45}$$

where $\sigma(x)$ is some estimate for the norm of the corresponding subgradient from (2-4) of ϕ at x. The standard class $\sigma(x)$ is $\sigma(x) = \sigma_0(x)$, but we consider even a broader class where $\sigma(x) = \sigma_1(x)$ in (4-45). Then Lemma 4.8 tells us that for any $x_0 \in U$, the estimates

$$M\left[\frac{1}{2L(x_0)}\sigma_1^2(x_0)\right]^q \leq M\big[\phi(x_0) - \phi(\bar{x})\big]^q \overset{(4\text{-}45)}{\leq} \sigma_1(x_0)$$

are satisfied by definition (4-43). This implies that

$$M\sigma_1^{2q-1}(x_0) \leq [2L(x_0)]^q. \tag{4-46}$$

It follows from (4-46) for $q \in (0, \frac{1}{2})$ that $L(x_0) \to \infty$ as $x_0 \to \bar{x}$, which is inconsistent with the Lipschitz continuity of ∇g around $\bar{x}$ and thus completes the proof. □

As we see, the inconsistency result of Theorem 4.9 affects also problems of *unconstrained minimization of $\mathscr{C}^{1,1}$ functions* at their local minimizers. After a preliminary version of this paper was uploaded to arXiv, we received a message from Nicolas Boumal informing us that about the results and discussions on the failure of the lower exponent PŁK conditions for problems of $\mathscr{C}^{1,1}$-optimization here given in [42, Remark 2.21].

The following example shows that the inconsistency observation of Theorem 4.9 may *fail* if g is *not of class* $\mathscr{C}^{1,1}$. The function below is taken from [31, Example 4.6], where it is used to demonstrate that the tight *quadratic* convergence rate can be achieved for the proximal methods applied to this function.

Example 6 ($\mathscr{C}^{1,1}$ property is essential for inconsistency with lower exponent PŁK) Consider the univariate function $\phi(x) := |x|^{3/2}$, which has the global minimizer $\bar{x} = 0$. It is straightforward to compute that

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

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

CHAPTER 5

Speed of Convergence of Alternating Minimization Algorithms with Costs to Move: Worthwhile to Move Potential Games

The study of descent methods in optimization has witnessed significant advancements, particularly in the context of nonconvex and nonsmooth problems. Despite these achievements, numerous challenges and opportunities for further exploration remain. This chapter aims to outline potential directions for future research in the convergence analysis for the alternating proximal point method proposed and developed by Attouch, Bolte, Redont, and Soubeyran (Math Oper Res. https://doi.org/10.1287/moor.1100.0449,2010), emphasizing emerging trends, open problems, and interdisciplinary applications. By addressing both theoretical gaps and practical implementations, this discussion seeks to inspire novel contributions and foster deeper insights into the behavior of this method under various conditions.

In this chapter, we are interested in revisiting the class of alternating minimization algorithms, which was explored in [7] for nonconvex structured functions of the type $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ given by,

$$L(x,y) = f(x) + Q(x,y) + g(y),$$

where f and g are proper lower semicontinuous functions, defined on Euclidean spaces, and Q is a smooth function that couples the variables x and y. Given $(x^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m$, the dynamics of the method generates iteratively $(x^k, y^k) \to (x^{k+1}, y^k) \to (x^{k+1}, y^{k+1})$ as follows:

$$x^{k+1} \in \arg\min \left\{ L(u, y^k) + \frac{1}{2\lambda_k} \|u - x^k\|^2 : u \in \mathbb{R}^n \right\}, \tag{5-1}$$

$$y^{k+1} \in \arg\min \left\{ L(x^{k+1}, v) + \frac{1}{2\mu_k} \|v - y^k\|^2 : v \in \mathbb{R}^m \right\}, \tag{5-2}$$

with $\{\lambda_k\}$ and $\{\mu_k\}$ being apositive sequences.

The convergence speed of a non-cooperative game is a central topic in game theory, yet results in this area remain scarce. This paper presents a valuable refinement concerning the speed of convergence for a new class of alternating games, featuring symmetric or asymmetric costs to move considered [6, 7, 9]. More precisely, we have established a finite or superlinear convergence rate in the case where the bifunction L satisfies the PŁK property with lower exponents. This has been done under the same assumptions assumed in [7], namely,

$$(\mathscr{H})\begin{cases} L(x,y)=f(x)+Q(x,y)+g(y), & \\ f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}, & \text{is proper lower semicontinuous,}\\ g:\mathbb{R}^m\to\mathbb{R}\cup\{+\infty\}, & \text{is proper lower semicontinuous,}\\ Q:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}, & \text{is a } C^1 \text{ function,}\\ \nabla Q & \text{is Lipschitz continuous on bounded subsets of } \mathbb{R}^n\times\mathbb{R}^m. \end{cases}$$

The following permanent assumption is made about equations (5-1) and (5-2):

$$(\mathscr{H}_1)\begin{cases} \inf_{\mathbb{R}^n\times\mathbb{R}^m} L > -\infty,\\ \text{the function } L(\cdot,y^0) \text{ is proper,}\\ \text{for some positive } 0<r_-<r_+, \text{ the sequences of stepsizes } \lambda_k,\mu_k \text{ belong to } (r_-,r_+). \end{cases}$$

with respect to which the authors in [7, Lemma 3.1] proved both the good definition of the method and decrease conditions that will be useful throughout the paper:

Lemma 5.1 *Under assumptions* $(\mathscr{H})$ *and* $(\mathscr{H}_1)$*, the sequences* $(x_k),(y_k)$ *are correctly defined. Moreover, the following hold:*

(i) For all $k\geq 1$*,*

$$L(x_k,y_k)+\frac{1}{2\lambda_{k-1}}\|x_k-x_{k-1}\|^2+\frac{1}{2\mu_{k-1}}\|y_k-y_{k-1}\|^2\leq L(x_{k-1},y_{k-1}). \tag{5-3}$$

Hence $L(x_k,y_k)$ *does not increase.*

(ii)

$$\sum_{k=1}^{\infty}\left(\|x_k-x_{k-1}\|^2+\|y_k-y_{k-1}\|^2\right)<+\infty,$$

and, in particular,

$$\lim_{k\to\infty}\left(\|x_k-x_{k-1}\|+\|y_k-y_{k-1}\|\right)=0.$$

(iii) Define

$$(x_k^*, y_k^*) = \left(\nabla_x Q(x_k, y_k) - \nabla_x Q(x_k, y_{k-1}), 0\right) - \left(\frac{1}{\lambda_{k-1}}(x_k - x_{k-1}), \frac{1}{\mu_{k-1}}(y_k - y_{k-1})\right),$$

for $k \geq 1$; we have

$$(x_k^*, y_k^*) \in \partial L(x_k, y_k).$$

For any bounded subsequence (x_{k_j}, y_{k_j}) of (x_k, y_k), we have $(x_{k_j}^*, y_{k_j}^*) \to 0 \quad$ as $\quad k_j \to +\infty$. Hence, $\operatorname{dist}(0, \partial L(x_{k_j}, y_{k_j})) \to 0 \quad$ *as* $\quad k_j \to +\infty$.

Given the structure of the function L, it is crucial to determine the exponent of a desingularizing function associated with L, provided that the exponents of the desingularizing functions associated with f, g, and Q are already known. The calculus of the PŁK exponent and its applications to the linear convergence of first-order methods has been a subject of study in recent years; see, for instance, [32], where the authors obtain explicit convergence rates for various first-order methods applied to a wide range of optimization models.

It is worth noting that for specific choices of Q and g, the structure of L recovers an important class of bifunctions used in the convergence analysis of several first-order methods whose iterates involve momentum terms. For further details, see, for example, [17, 38] for the convergence analysis of certain inertial proximal algorithms and [18, 26] for the convergence analysis of the proximal gradient algorithm with extrapolation. When it is known that f has a PŁK exponent $\alpha \in [1/2, 1)$, $g \equiv 0$, and $Q(x, y) = \|x - y\|^2$, the authors in [32, Theorem 3.6 and Theorem 5.1] respectively, determine the PŁK exponent of L and establish the convergence rate of an inertial proximal algorithm with constant step sizes. In this paper, we will also open this discussion about the PŁK exponent calculation.

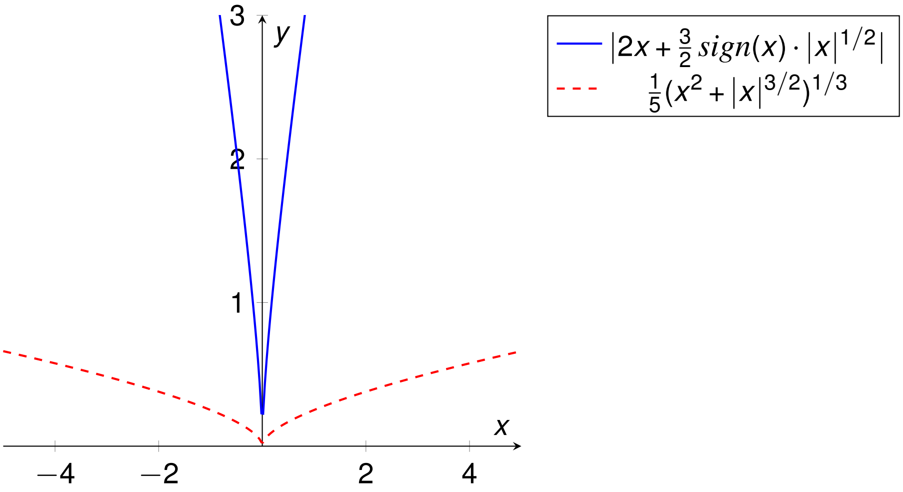
5.1 Preliminaries on the Calculus of the PŁK Exponent

In this section we present some facts concerning the calculation of the PŁK exponent. Given the structure of the function L, it is essential to understand how to compute the exponent of a desingularizing function associated with L, provided that the exponents of the desingularizing functions associated with f, g, and Q are known.

Due to the significance of Theorem [12, Theorem 7], which deals with inconsistency of the Polyak-Łojasiewicz-Kurdyka property for L with a low exponent in the case where $h = h_1 - h_2$, where h_1 has continuous Lipschitz gradient, it is natural to consider the case of $L = Q + (f + g)$, where both functions are convex and Q admits a Lipschitz continuous gradient. We will now illustrate some situations that show that, in this case, there are some situations in which L admits a low exponent. Let's start with the following one-dimensional case:

Example 7 *Let us consider* $Q, f : \mathbb{R} \longrightarrow \mathbb{R}$, *given by* $Q(x) = x^2$, $f(x) = |x|^{3/2}$ *and* $g \equiv 0$. *Note that* $\bar{x} = 0$, f *has minimum PŁK exponent* $q = 1/3$, Q *has continuous Lipschitz derivative and satisfies the PŁK inequality with* $q = 1/2$. *We affirm that* L *has PLK exponent equal to* $1/3$, *i.e.,*

$$|3/2sign(x)|x|^{1/2} + 2x| \geq 1/5[|x|^{3/2} + x^2]^{1/3}.$$



Indeed, because the function $|\frac{3}{2}sign(x)|x|^{1/2} + 2x|$ *and* $\frac{1}{5}(x^2 + |x|^{3/2})^{1/3}$ *is even, it is sufficient to consider the case where* $x > 0$. *Let*

$$f_1(x) = \left(\frac{3}{2}x^{1/2} + 2x\right)^3, \qquad \text{and } g_1(x) = x^{3/2} + x^2.$$

Using Newton's binomial, we have

$$f_1(x) = \frac{27}{8}x^{3/2} + \frac{27}{2}x^2 + 9x^{5/2} + 8x^3.$$

Thus, in a neighborhood of the origin, we have:

$$\left(\left|\frac{3}{2}sign(x)|x|^{1/2} + 2x\right|\right)^3 \geq |x|^{3/2} + x^2.$$

which concludes the statement.

Next, we recall [32, Theorem 3.3 and Theorem 3.6], where the authors establish the computation of the exponent, respectively, for block-separable sums of PŁK functions and for a potential function used in the convergence analysis of the inertial proximal algorithm in [38]).

Theorem 5.2 (Exponent for block separable sums of PŁK functions) *Let* $n_i, n \in \mathbb{N}$, *for* $i = 1, \dots, m$, *such that* $\sum_{i=1}^{m} n_i = n$. *Let*

$$f(x) = \sum_{i=1}^{m} f_i(x_i),$$

where each f_i, $1 \leq i \leq m$, *is a proper closed function on* $\mathbb{R}^{n_i}$ *with* $x = (x_1, \dots, x_m) \in \mathbb{R}^n$. *Suppose further that each* f_i *is a PŁK function with an exponent* $\alpha_i \in (0,1)$, *and that each* f_i *is continuous on* $\operatorname{dom} \partial f_i$, *for* $i = 1, \dots, m$. *Then* f *is a PŁK function with an exponent*

$$\alpha = \max\{\alpha_i : 1 \leq i \leq m\}.$$

Theorem 5.3 (Exponent for a potential function for iPiano) *Suppose that* f *is a proper closed function that has the PŁK property at* $\bar{x} \in \operatorname{dom} \partial f$ *with an exponent* $\alpha \in \left[\frac{1}{2}, 1\right)$ *and let* $\beta > 0$. *Consider the function*

$$F(x,y) := f(x) + \frac{\beta}{2}\|x - y\|^2.$$

Then the function F *has the PŁK property at* $(\bar{x}, \bar{x})$ *with an exponent of* α.

Remark 6

a) *Take* $\bar{x} = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_m) \in \mathbb{R}^n$. *Referring to Theorem 5.2, let us assume that* $\nabla f_j(\bar{x}_j) = 0$ *and that* $\alpha_j \in [0,1)$ *is the smallest PŁK exponent of* f_j *in* $\bar{x}_j$ *for all* $j = 1, ..., m$. *We may approximate around the point* $\bar{x} = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_m) \in \mathbb{R}^n$, *by* $(\bar{x}_1, \bar{x}_2, ..., x_j, ..., \bar{x}_m)$, *and in this case, for all* $j \in \{1, ..., m\}$, *we have*

$$\begin{aligned} \|\nabla f(\bar{x}_1, ..., x_j, ..., \bar{x}_m)\| = \|\nabla f_j(x_j)\| &\geq (f_j(x_j) - f_j(\bar{x}_j))^{\alpha_j} \\ &= (f(\bar{x}_1, ..., x_j, ..., \bar{x}_m) - f(\bar{x}_1, ..., \bar{x}_j, ..., \bar{x}_m))^{\alpha_j}. \end{aligned}$$

Therefore, the smallest PŁK exponent of f *in* $\bar{x} = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_m)$ *is the* $\max\{\alpha_1, ..., \alpha_m\}$. *By way of illustration, let us consider* $f(x,y) = g(x) + h(y)$, *with* $g(x) = |x|^{3/2}$ *and* $h(y) = y^2$. *It is known that* g, h *satisfy the PŁK inequality at* $\bar{x} = 0$ *and* $\bar{y} = 0$ *with exponent* $1/3$ *and* $1/2$, *respectively. From Theorem 5.2, we have that* f *satisfies the PŁK condition at* $(\bar{x}, \bar{y}) = (0,0)$ *with exponent* $q_0 = \max\{1/3, 1/2\} = 1/2$. *We can consider the approximation by the* y *axis of* $(\bar{x}, \bar{y})$, *obtaining*

$$|2y| = \|\nabla f(0,y)\| \geq M(|0|^{3/2} + y^2)^q = My^{2q}, \quad M > 0.$$

This allows us to conclude that in fact, the smallest possible exponent for f, *when analyzing the property in a neighborhood of* $(0,0)$, *is* $q = 1/2$;

b) *Without the assumption that $\bar{x}_j$ is a critical point of f_j, Theorem 5.2 does not yield the best possible result for the PŁK exponent. Indeed, let us consider the function $f(x,y) = x^2 + y$. Since $\nabla f(x,y) = (2x, 1)$, then [32, Lemma 2.1] implies that f satisfies the PŁK property for every $q \in [0,1)$ for all point $(\bar{x},\bar{y}) \in \mathbb{R}^2$.*

Note that Theorems 5.2 and 5.3 does not apply, for example, to bifunctions of the form $L_1(x,y) = |x|^{3/2} + |y|^{3/2} + |x-y|^2$ or $L_2(x,y) = |x| + |y| + |x-y|^2$. For the case where $L = L_1$, we conjecture that L satisfies the PŁK inequality at $(\bar{x},\bar{y})$ with $q = 1/3$. Let's see:

$$\nabla L(x,y) = \left(\frac{3}{2}\operatorname{sgn}(x)|x|^{1/2} + 2(x-y),\ \frac{3}{2}\operatorname{sgn}(y)|y|^{1/2} - 2(x-y)\right),$$

and, for (x,y) in a neighborhood of the $(\bar{x},\bar{y}) = (0,0)$, it follows that

$$\left\|\left(\frac{3}{2}\operatorname{sgn}(x)|x|^{1/2} + 2(x-y),\ \frac{3}{2}\operatorname{sgn}(y)|y|^{1/2} - 2(x-y)\right)\right\|^3 \geq |x|^{3/2} + |y|^{3/2} + |x-y|^2,$$

where

$$\|\nabla L(x,y)\|^3 = \left[\frac{9}{4}(|x|+|y|) + 8(x-y)^2 + 6(x-y)\left(\operatorname{sgn}(x)|x|^{1/2} - \operatorname{sgn}(y)|y|^{1/2}\right)\right]^{3/2}.$$

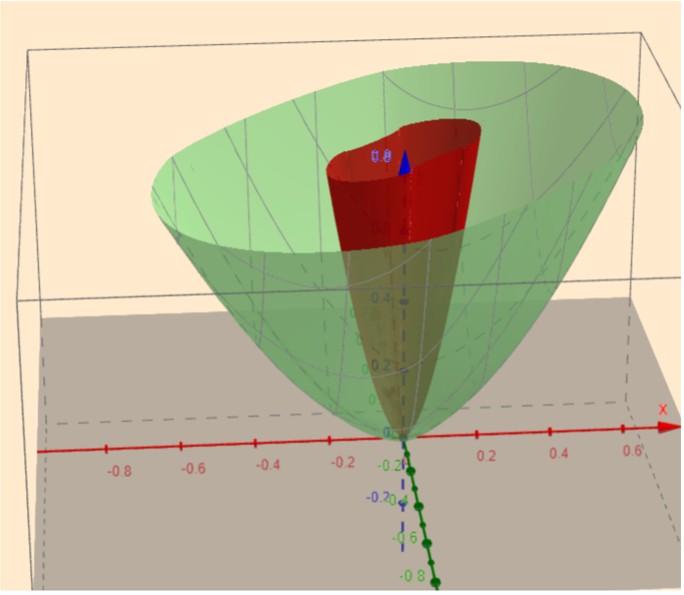


Figure 5.1: — $\|\nabla L(x,y)\|^3$ — $L(x,y)$.

In the case where $L = L_2$, we have that $(\bar{x},\bar{y}) = (0,0)$ is a critical point of L. For $x \neq 0$ and $y \neq 0$, we have

$$\nabla L(x,y) = (sign(x) + 2(x-y), sign(y) - 2(x-y)).$$

Hence, $\|\nabla L(x,y)\|^2 = 2+8(x-y)^2+4(x-y)(sign(x)-sign(y)) \geq c > 0$.

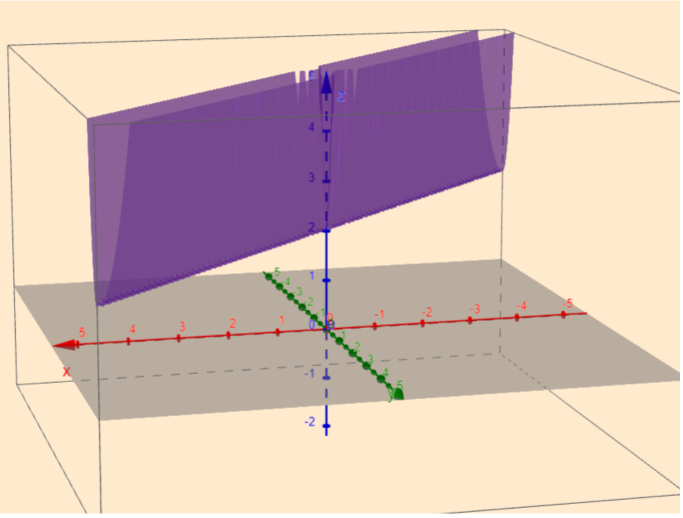


Figure 5.2: — $\|\nabla L(x,y)\|^2$

Thus, we have that for all $q \in [0,1)$, exist $U_{(0,0)}$ e $M > 0$ such that

$$\|\nabla L(x,y)\| \geq c \geq M(|x|+|y|+(x-y)^2)^q, \quad \text{for all } (x,y) \in U_{(0,0)}.$$

5.2 Convergence Rate

Theorem 5.4 (Rate of Convergence) *Assume that L satisfies conditions $\mathscr{H}$ and $\mathscr{H}_1$. Further, assume that the sequence (x^k,y^k) converges to (x^*,y^*) and that L has the PŁK property at (x^*,y^*) with*

$$\varphi(t) = ct^{1-\theta}, \quad \theta \in [0,1), \quad c > 0.$$

Then, the following estimates hold:

(i) If $\theta = 0$, then the sequence $L(x^k,y^k)$ converges in a finite number of steps

(ii) If $\theta \in \left(0,\frac{1}{2}\right)$, then $L(x^k,y^k)$ converges finite or superlinear to $L(x^*,y^*)$.

(iii) If $\theta = 1/2$, then $L(x^k,y^k)$ converges linear to $L(x^*,y^*)$.

(iv) If $\theta \in \left(\frac{1}{2},1\right)$, then there exists $c > 0$ such that

$$L(x^k,y^k) - L(x^*,y^*) \leq ck^{-\frac{1}{2q-1}}.$$

Proof. As emphasized in the introduction of the paper, the sequence $\{(x^k,y^k)\}$ generated from (5-1) and (5-2) converges to $(\bar{x},\bar{y})$. From the decreasing condition in (5-3), we have the sequence $\{L(x^k,y^k)\}$ is monotone decreasing. Because $(\bar{x},\bar{y})$ is an accumulation point of the sequence $\{(x^k,y^k)\}$, then $\{L(x^k,y^k)\}$ converges to $L(x^*,y^*)$ as k goes to infinity.

Defining $\Psi(x,y) := L(x,y) - L(x^*,y^*)$, it follows that $\{\Psi(x^k,y^k)\} \subset [0,+\infty)$ is monotone decreasing.

Now, combining (5-3) and the definition of the Ψ from the fact that

$$\|(x_k^*,y_k^*)\| \leq (C+\frac{1}{r_-})\|(x^k,y^k)-(x^{k-1},y^{k-1})\|,$$

we obtain

$$\begin{aligned}\Psi(x^{k-1},y^{k-1})-\Psi(x^k,y^k) &\geq \frac{1}{2\lambda_{k-1}}\|x^k-x^{k-1}\|^2+\frac{1}{2\mu_{k-1}}\|y^k-y^{k-1}\|^2\\ &\geq \frac{1}{2r_+}\|z^k-z^{k-1}\|^2 \geq \frac{1}{2r_+(C+1/r_-)^2}\|(x_k^*,y_k^*)\|^2,\end{aligned}$$

as $(x_k^*,y_k^*) \in \partial L(x^k,y^k)$, combining the last inequality as the condition PŁK, follows that

$$\Psi(x^{k-1},y^{k-1})-\Psi(x^k,y^k) \geq C_0\left(\Psi(x^k,y^k)\right)^{2\theta}. \tag{5-4}$$

where $C_0 := \frac{1}{2r_+(C+1/r_-)^2c^2(1-\theta)^2}$.

(i) Let $\theta = 0$, suppose the sequence is infinitely generated, in (5-4) obtain

$$\Psi(x^{k-1},y^{k-1})-\Psi(x^k,y^k) \geq C_0,$$

generating a contradiction.

(ii) Let $\theta \in (0,1/2)$. Suppose the sequence is infinitely generated, In (5-4) obtain

$$1+C_0\left(\Psi(x^k,y^k)\right)^{2\theta-1} \leq \frac{\Psi(x^{k-1},y^{k-1})}{\Psi(x^k,y^k)},$$

As $\theta \in (0,1/2)$, we have $2\theta-1<0$, by the fact that $\{\Psi(x^k,y^k)\}$ converges a zero, from the last inequality it follows that $\lim_{k\to+\infty}\frac{\Psi(x^{k-1},y^{k-1})}{\Psi(x^k,y^k)} = +\infty$. Thus,

$$\lim_{k\to+\infty}\frac{\Psi(x^k,y^k)}{\Psi(x^{k-1},y^{k-1})} = 0,$$

and the result follows.

(iii) If $\theta = 1/2$, from (5-4) it follows that

$$\Psi(x^k,y^k) \leq \frac{1}{1+C_0}\Psi(x^{k-1},y^{k-1})$$

Since $1/(1+C_0) \in (0,1)$, assertion (iii) is a consequence of the last inequality.

(iv) For $q \in (1/2, 1)$, follows of (5-4) that

$$C_0 \leq \frac{\Psi(x^{k-1}, y^{k-1}) - \Psi(x^k, y^k)}{\big(\Psi(x^k, y^k)\big)^{2\theta}}.$$

Let $\Phi(t) := \dfrac{1}{t^{2q}}$. Suppose first that there exists $P \in (1, \infty)$ such that $\Phi(\Psi(x^k, y^k)) \leq P\Phi(\Psi(x^{k-1}, y^{k-1}))$. Combining this fact with the last inequality and Lemma 2.5 we have that

$$\begin{aligned}
C_0 &\leq \frac{\Psi(x^{k-1}, y^{k-1}) - \Psi(x^k, y^k)}{\big(\Psi(x^k, y^k)\big)^{2\theta}} = \int_{\Psi(x^k,y^k)}^{\Psi(x^{k-1},y^{k-1})} \Phi(\Psi(x^k, y^k))dt \\
&\leq P\int_{\Psi(x^k,y^k)}^{\Psi(x^{k-1},y^{k-1})} \Phi(\Psi(x^{k-1}, y^{k-1}))dt \leq \frac{P}{1-2\theta}\Big[\big(\Psi(x^{k-1}, y^{k-1})\big)^{1-2\theta} - \big(\Psi(x^k, y^k)\big)^{1-2\theta}\Big] \\
&= \frac{P}{2\theta - 1}\Big[\big(\Psi(x^k, y^k)\big)^{1-2\theta} - \big(\Psi(x^{k-1}, y^{k-1})\big)^{1-2\theta}\Big],
\end{aligned}$$

which implies by $2\theta - 1 > 0$ that

$$\frac{C_0(2\theta - 1)}{P} \leq \big(\Psi(x^k, y^k)\big)^{1-2\theta} - \big(\Psi(x^{k-1}, y^{k-1})\big)^{1-2\theta}, \quad k \geq k_0. \tag{5-5}$$

Letting $j > k_0$, we deduce of the last inequality that

$$\sum_{k=k_0}^{j} \frac{C_0(2\theta - 1)}{P} \leq \sum_{k=k_0}^{j} \big(\Psi(x^k, y^k)\big)^{1-2\theta} - \big(\Psi(x^{k-1}, y^{k-1})\big)^{1-2\theta}$$

which in turn implies that

$$\frac{(j+1-k_0)C_0(2\theta - 1)}{P} \leq \big(\Psi(x^j, y^j)\big)^{1-2\theta} - \big(\Psi(x^{k_0}, y^{k_0})\big)^{1-2\theta}.$$

Consequently, we establish the estimate

$$\big(\Psi(x^j, y^j)\big)^{1-2\theta} \geq \frac{(j+1-k_0)C_0(2\theta - 1)}{P} + \big(\Psi(x^{k_0}, y^{k_0})\big)^{1-2\theta}.$$

Since the function $t \mapsto t^{-\frac{1}{2\theta-1}}$ is monotonically decreasing for $\theta \in (1/2, 1)$, we obtain that

$$\Psi(x^j, y^j) \leq \left(\frac{(j+1-k_0)C_0(2\theta - 1)}{P} + \big(\Psi(x^{k_0}, y^{k_0})\big)^{1-2\theta}\right)^{-\frac{1}{2\theta-1}},$$

Thus, there exists $\eta > 0$ such that

$$\Psi(x^j, y^j) \leq \eta j^{-\frac{1}{2\theta-1}}.$$

and therefore confirms the convergence rate estimate in (iv) for this case.

Now, let us examine a different case where the number P referenced above is nonexistent, i. e., for any $P \in (1,\infty)$ we have $\Phi(\Psi(x^k,y^k)) > P\Phi(\Psi(x^{k-1},y^{k-1}))$ for all k large enough. fixed $P \in (1,\infty)$ and define $P_1 := \frac{1}{P^{\frac{1}{2\theta}}}$. Then it follows from the definitions that $\Psi(x^k,y^k) \leq P_1\Psi(x^{k-1},y^{k-1})$ and therefore, $\Psi(x^k,y^k)^{1-2q} \geq P_1^{1-2\theta}\Psi(x^{k-1},y^{k-1})^{1-2\theta}$ because $1-2\theta < 0$. Subtracting $\Psi(x^{k-1},y^{k-1})^{1-2\theta}$ on both sides of the last inequality results in

$$\Psi(x^k,y^k)^{1-2\theta} - \Psi(x^{k-1},y^{k-1})^{1-2\theta} \geq \left(P_1^{1-2\theta} - 1\right)\Psi(x^{k-1},y^{k-1})^{1-2q}.$$

Note that $P_1 \in (0,1)$ and therefore $P_1^{1-2q} > 1$ for all $q \in (1/2,1)$. The convergence $\Psi(x^k,y^k) \to 0$ yields $\psi(x^k)^{1-2q} \to \infty$ as $k \to \infty$, and hence

$$\Psi(x^k,y^k)^{1-2\theta} - \Psi(x^{k-1},y^{k-1})^{1-2\theta} \geq \frac{C_0(2\theta-1)}{P}$$

whenever k is sufficiently large. The rest of the proof is similar to the above arguments based on (5-5), and thus we are done with verifying (iv) and the whole theorem. □

For the next result, we will demonstrate only item (ii), as it is the only variation of θ that provides improved precision in the convergence rate of the presented model.

Theorem 5.5 *Assume that L satisfies conditions $\mathscr{H}$ and $\mathscr{H}_1$. Further, assume that the sequence (x^k,y^k) converges to (x^*,y^*) and that L has the PŁK property at (x^*,y^*) with*

$$\varphi(s) = cs^{1-\theta}, \quad \theta \in [0,1), \quad c > 0.$$

Then, the following estimates hold:

(i) If $\theta = 0$, then the sequence (x^k,y^k) converges in a finite number of steps

(ii) If $\theta \in \left(0,\frac{1}{2}\right)$, then (x^k,y^k) converges finite or superlinear to (x^*,y^*).

(iii) If $\theta = 1/2$, then (x^k,y^k) converges linear to (x^*,y^*).

(iv) If $\theta \in \left(\frac{1}{2},1\right)$, then there exists $c > 0$ such that

$$(x^k,y^k) - (x^*,y^*) \leq ck^{-\frac{1-q}{2q-1}}.$$

Proof. In [7, Theorem 3.1], we have

$$\varphi\left(L(x^k,y^k) - L(x^*,y^*)\right) - \varphi\left(L(x^{k+1},y^{k+1}) - L(x^*,y^*)\right) \geq \frac{1}{M}\frac{\|z^{k+1}-z^k\|^2}{\|z^k-z^{k-1}\|}, \tag{5-6}$$

Here $M = 2r_+(C+1/r_-)$ and $z^k = (x^k,y^k)$. Given $r \in (0,1)$, we have the two possibilities for such k:

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

In case (a), the usage of (5-6) with $\varphi(s)=cs^{1-\theta}$ and $\Psi(x,y):=L(x,y)-L(x^*,y^*)$ yields

$$\|z^{k+1}-z^k\| \leq \frac{Mc}{r}\Psi(z^k)^{1-\theta}-\frac{Mc}{r}\Psi(z^{k+1})^{1-\theta} \leq r\|z^k-z^{k-1}\|+\frac{cM}{r}\Psi(z^k)^{1-\theta}-\frac{Mc}{r}\Psi(z^{k+1})^{1-\theta}.$$

Conversely, in case (b), we immediately obtain

$$\|z^{k+1}-z^k\| \leq r\|z^k-z^{k-1}\| \leq r\|z^k-z^{k-1}\|+\frac{cM}{r}\Psi(z^k)^{1-\theta}-\frac{Mc}{r}\Psi(z^{k+1})^{1-\theta}.$$

So, in both cases we have

$$\|z^{k+1}-z^k\| \leq r\|z^k-z^{k-1}\|+\frac{cM}{r}\Psi(z^k)^{1-\theta}-\frac{Mc}{r}\Psi(z^{k+1})^{1-\theta}$$

This last inequality gives the following estimate for any natural $l \geq 1$.

$$\sum_{j=k}^{k+l}\|z^{k+1}-z^k\| \leq \frac{r}{(1-r)}\|z^k-z^{k-1}\|+\frac{cM}{r(1-r)}\left(\Psi(z^k)^{1-\theta}-\Psi(z^{k+l+1})^{1-\theta}\right). \tag{5-7}$$

Define $s_k := \sum_{j=k}^{\infty}\|z^{k+1}-z^k\|$ and let $l \to \infty$ in (5-7) we have

$$s_k \leq \frac{r}{(1-r)}\|z^k-z^{k-1}\|+\frac{cM}{r(1-r)}\Psi(z^k)^{1-\theta}$$

Using the last inequality with (5-3) and $L(z^{k-1})-L(z^k) \leq L(z^{k-1})-L(x^*,y^*)=\Psi(z^{k-1})$ bring us to the estimate

$$s_k \leq \frac{r}{(1-r)}\left(\Psi(z^{k-1})\right)^{1/2}+\frac{cM}{r(1-r)}\Psi(z^k)^{1-\theta}.$$

When $\theta \in (0,1/2)$, we obtain from the last inequality that

$$s_k \leq \frac{r}{(1-r)}\left(\Psi(z^{k-1})\right)^{1/2}+\frac{cM}{r(1-r)}\Psi(z^k)^{1/2}.$$

Since the sequence $\{\psi(x^k)\}$ is monotonically decreasing, follows that

$$s_k \leq \frac{r^2+cM}{(1-r)}\left(\Psi(z^{k-1})\right)^{1/2},$$

and the result follows of the Theorem (5.4), item (ii). □

5.3 The Importance of Worthwhile to Move Potential Games in Behavioral Sciences

The convergence speed of a non-cooperative game is of central importance in game theory, and results do not abound because, in the long run, we are all dead. This note provides a valuable refinement about the speed of convergence of a new class of alternating games with symmetric or asymmetric costs to move initiated by Attouch et al. [6, 7, 9]. They represent examples of worthwhile-to-move potential games played in asymmetric metric spaces (Cruz Neto et al. [20], Soubeyran et al., [46]), where each agent plays, in alternation, a worthwhile move, given the action done by the other. We show that finite termination occurs when the payoff functions of two players are sharp enough concerning the worthwhile change in their own actions. This is a striking result. The central concept of such a behavioral game is that of an individual worthwhile move that improves a player's payoff without requiring a too high cost to move; if other players stay at the status quo. That is, it better satisfies the needs and desires of this individual without too many sacrifices for himself. This concept is borrowed from the novel Variational rationality approach of individual and social stay and change human dynamics; see, for instance [43, 44] and also a huge list of applications of different optimization algorithms in behavioral sciences in https://sites.google.com/view/antoine-soubeyran.

5.3.1 Worthwhile to Move Games Within the Variational Rationality Approach of Human Dynamics

At a higher level of generality (Soubeyran, [45], submitted, and Soubeyran, 2025, in preparation), for each individual, such a move $m = (u, \omega_{u,v}, v)$ starts from having done a thing u (= a task, an activity, a bundle of daily activities), continues, after stopping to do u, with $\omega_{u,v}$ = becoming able to do v, to end in doing v. Then, an individual move has two stages; becoming able to do, then, do a thing. This move must be contrasted with the stay $\sigma = (u, \omega_{u,u^{=}}, u^{=})$, where $u^{=}$ is doing u again. This stay can be a habit where we repeat doing u again and again. Then, for an agent, the net value (= benefits - costs in terms of satisfaction - dissatisfaction feelings) of doing a move is the sum of, i) the net value of doing v and; ii) the net value of becoming able to do it. This move starting from having done u is worthwhile when its net value is higher than the net value of staying at u. A succession of worthwhile moves represents a worthwhile transition or a way to our goals. If these worthwhile moves are also satisficing (= improve enough, make enough progress, go fast enough), this transition can approach at a sufficient speed, and a cost low enough, and, finally reach a desired end (a goal), where a selected list of our needs and desires are satisfied. In this way, the VR approach gives, in the context of optimizing algorithms and variational principles, a broad outline of how to solve the most general problem addressed to humans: how to find cheap and quick enough ways of motion to satisfice our recurrent and changing needs/desires. Soubeyran (2025, in preparation) presents a variational/recursive (algorithmic) reconstruction of psychology via the net value of choosing, doing a move, and moving, i.e., becoming able to do, do a thing. Hopefully, the reason why psychology admits such a powerful algorithmic reconstruction is simple: a perturbation term in an algorithm (for example, a proximal algorithm) represents, most of the time, the costs of becoming able to do a thing. Then, what is a trick in an optimization algorithm, becomes an essential part of the description of a move (behavior) defined as becoming able to do something and doing it. That is, if we use the VR approach to understand psychology, the proof (= way) giving the resolution of an algorithm is as important as the result! Notice that the (VRRP) Variationality rationality research program, which proposes for the first time (Soubeyran, [45]), an algorithmic reconstruction of psychology in a dynamical system setting, will, we hope, greatly help to reinforce the link between Artificial Intelligence (A.I,, i.e., how machines can function as well, then, better than intelligent humans) and its genitors, including Psychology/Behavioral Sciences, Optimizing algorithms/Variational analysis/Game theory, and Computer Sciences in different asymmetric distances spaces because costs to move from x to y differ from the opposite.

The present note on the speed of convergence of a worthwhile to move game goes in this broad direction.

5.3.2 Example: An Educational Family Game

References on non-cooperative games, potential games, and several examples of worthwhile-to -move potential games, are given in Soubeyran et al.(24). Among potential games, tragedy of the common games plays a major role. For example, it concerns two parents when one parent does not participate enough in educating his(her) children, leading to some possible future drawbacks for children, i.e., a bad education level. Let us consider an instructive mini-model (Soubeyran et al., 24) where two parents $j \in \{f, g\}$ share their daily time $T > 0$ between two tasks: spending the time $x_j \in \mathbb{R}, 0 \leq x_j \leq T$, in educating their children, and spending the remaining time $T - x_j \geq 0$ in practicing their hobbies. Let us go a little further. The level of education $Q(x_f, x_g) = x_f^{\alpha_f} x_g^{\alpha_g}$ given to their children increases with the time they spend, both, in the education task, at a decreasing rate, i.e., $\alpha_\alpha, \alpha_g \in]0, 1]$. In this context, $x_f^{\alpha_f}$and $x_g^{\alpha_g}$ must be concave increasing functions, zero at zero on $\mathbb{R}_+$. For example, $Q(x_f, x_g) = x_f^{1/2} x_g^{1/3}$. Then, i) the educational level $Q(x_f, x_g)$ is zero if one of the parents does not participate in the education task (say, $x_f = 0$), and ii) the more a parent participates in the education task, the more the education level increases if the other parent participates more (a positive spillover). Each parent has two conflicting goals: to experience the pleasant feeling $E_j(x_j, x_{\diagdown j}) = \beta_j Q(x_j, x_{\diagdown j})$, $\beta_j > 0, j \in \{f, g\}$ of helping children in their personal development (= in improving their level of education), and to experience a fun and accomplishment feeling $H_j(x_j) = \gamma_j \left[T - x_j\right], \lambda_j > 0$, in practicing some hobbies to improve his(her) personal development. Then, because their daily time $T < +\infty$ is limited, these two goals (education and hobbies) conflict for each parent. The degree of conflicts between these two goals is different if $\beta_f \not\neq \beta_g$. For simplification, we will suppose that the two parents experience the same pleasant feeling to help for the education of their children, i.e., $\beta_f = \beta_g = \beta$. Then, the payoff (in terms of net satisfaction feeling) of parent $j = f, g$ are,

$$P_f(x_f, x_g) = H_f(x_f) + E_f(x_f, x_g) = \gamma_f [T - x_f] + \beta Q(x_f, x_g),$$

$$P_g(x_g, x_j) = H_g(x_g) + E_g(x_f, x_g) = \gamma_g \left[T - x_g\right] + \beta Q(x_f, x_g).$$

The potential function of this game is $P(x_f, x_g) = \gamma_f [T - x_f] + \beta Q(x_f, x_g) + \gamma_g \left[T - x_g\right]$. The striking fact is that this potential function is algebraic if, for example, the level of education of children is $Q(x_f, x_g) = x_f^{\alpha_f} x_g^{\alpha_g}$, where $\alpha_\alpha, \alpha_g \in]0, 1]$ are rational numbers.

In the variational rationality approach, costs to move refer to capability costs, i.e., costs to acquire capabilities. In this educational context, they are the costs of becoming able (= ready) to spend more time on the education task. They include the costs of becoming prepared to sacrifice hobbies, the costs of accepting to stop making progress in spending less time to play tennis, costs of penalizing our team if we do not participate in a local competition, the emotional costs of refusing others to participate to some social activity useful for poor people, the costs of breaking social relationships, due to a lack of time, ...In this example, costs to move are supposed to be quadratic, i.e., they increase more and more, the more we reduce hobby time. To meet the mathematical part, we make two strong hypotheses: we take them symmetric, and we suppose that costs to stay are zero. The general behavioral case is the opposite. In this simplified context, the costs of moving of parents f and g are $C(x_f^k, u) = (1/2\lambda_f^k) \left|u - x_f^k\right|^2$ and $C(x_g^k, v) = (1/2\lambda_g^k) \left|v - x_g^k\right|^2$, with $\lambda_j^k > 0, j = f, g$. They appear in the formulas (5), (6). The negative $\left\{-L(u, y_k) - (1/2\lambda_k) \left\|\left|u - x_k\right|^2\right\|\right\}$ of (5) and the negative $\left\{-L(x_{k+1}, v) - (1/2\mu_k) \left\|v - y_k\right\|^2\right\}$ of (6) define, in this behavioral example, the net value, i.e., the satisfaction feelings $V_j(u_j, x_{\diagdown j})$ that parent j derives from sharing his(her) time between education and personal hobbies. It is $V_j(u_j, x_{\diagdown j}) = \gamma_j \left[T - u_j\right] + \beta u_j^{\alpha_j} x_{\diagdown j}^{\alpha_{\diagdown j}} - (1/2)(u_j - x_j)^2$, $j = f, g$. This is the payoff that parent j tries to maximize when he (her) chooses to spend the educational time u_j, given the educational time x_{-j} chosen by parent $-j$.

Future Prospects

The topic under discussion in Chapter 5 presents a wide range of research opportunities with the potential to advance scientific knowledge. In what follows, we highlight some of these directions.

- To investigate the possibility of improving the convergence result by removing the assumption that the function $Q(x,y)$ has a Lipschitz continuous gradient.
- To introduce a linear resource constraint (working on the simplex), forming the Lagrangian, using the Euclidean distance for application in behavioral sciences. The model alternately minimizes $L(x,y) = f(x) + Q(x,y) + g(y)$, under the linear constraint $Ax + By = r$.
- Study of the minimum exponent of the PŁK property for the function $L(x,y) = f(x) + Q(x,y) + g(y)$, through the minimum exponents of f, g, and Q at a given point.

Another natural direction for future research is to extend the results developed in this thesis to problems formulated in the Riemannian setting. Such an approach would enable the investigation of whether the results established in the Euclidean framework can be generalized to curved spaces, where the geometry plays a crucial role. The extension of theories and optimization methods from the Euclidean to the Riemannian context has been extensively explored in the literature; in particular, the PŁK inequality in this setting has been investigated, for example,in [13, 21–23].

Bibliography

[1] ABSIL, P.-A.; MAHONY, R.; ANDREWS, B. **Convergence of the iterates of descent methods for analytic cost functions**. *SIAM Journal on Optimization*, 16(2):531–547, 2005.

[2] ARAGÓN ARTACHO, F. J.; FLEMING, R. M.; VUONG, P. T. **Accelerating the dc algorithm for smooth functions**. *Mathematical programming*, 169:95–118, 2018.

[3] ARAGÓN-ARTACHO, F. J.; MORDUKHOVICH, B. S.; PÉREZ-AROS, P. **Coderivative-based semi-newton method in nonsmooth difference programming**. *Mathematical Programming*, p. 1–48, 2024.

[4] ARAGOÓN ARTACHO, F. J.; VUONG, P. T. **The boosted difference of convex functions algorithm for nonsmooth functions**. *SIAM Journal on Optimization*, 30(1):980–1006, 2020.

[5] ATTOUCH, H.; BOLTE, J. **On the convergence of the proximal algorithm for nonsmooth functions involving analytic features**. *Mathematical Programming*, 116:5–16, 2009.

[6] ATTOUCH, H.; BOLTE, J.; REDONT, P.; SOUBEYRAN, A. **Alternating proximal algorithms for weakly coupled convex minimization problems. applications to dynamical games and pde's**. *Journal of Convex Analysis*, 15(3):485, 2008.

[7] ATTOUCH, H.; BOLTE, J.; REDONT, P.; SOUBEYRAN, A. **Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-łojasiewicz inequality**. *Mathematics of operations research*, 35(2):438–457, 2010.

[8] ATTOUCH, H.; BOLTE, J.; SVAITER, B. F. **Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized gauss–seidel methods**. *Mathematical Programming*, 137(1):91–129, 2013.

[9] ATTOUCH, H.; REDONT, P.; SOUBEYRAN, A. **A new class of alternating proximal minimization algorithms with costs-to-move**. *SIAM Journal on Optimization*, 18(3):1061–1081, 2007.

[10] AUSLENDER, A. **Optimisation**. *Méthodes numériques*, 1976.

[11] BENEDETTI, R.; RISLER, J.-J.; OTHERS. **Real algebraic and semialgebraic sets**. Hermann, 1991.

[12] BENTO, G.; MORDUKHOVICH, B.; MOTA, T.; NESTEROV, Y. **Convergence of descent optimization algorithms under polyak-lojasiewicz-kurdyka conditions**. *J Optim Theory Appl*, 207:41, 2025.

[13] BENTO, G.; NETO, J.; OLIVEIRA, P. **Convergence of inexact descent methods for nonconvex optimization on riemannian manifolds**. *arXiv preprint arXiv:1103.4828*, 2011.

[14] BOCHNAK, J.; COSTE, M.; ROY, M.-F. **Real algebraic geometry**, volume 36. Springer Science & Business Media, 2013.

[15] BOLTE, J.; DANIILIDIS, A.; LEWIS, A. **The łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems**. *SIAM Journal on Optimization*, 17(4):1205–1223, 2007.

[16] BOLTE, J.; DANIILIDIS, A.; LEWIS, A.; SHIOTA, M. **Clarke subgradients of stratifiable functions**. *SIAM Journal on Optimization*, 18(2):556–572, 2007.

[17] BOŢ, R. I.; CSETNEK, E. R. **An inertial tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems**. *Journal of Optimization Theory and Applications*, 171(2):600–616, 2016.

[18] CHAMBOLLE, A.; DOSSAL, C. **On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm"**. *Journal of Optimization theory and Applications*, 166:968–982, 2015.

[19] CHILL, R.; FIORENZA, A. **Convergence and decay rate to equilibrium of bounded solutions of quasilinear parabolic equations**. *Journal of Differential Equations*, 228(2):611–632, 2006.

[20] CRUZ NETO, J.; OLIVEIRA, P.; SOARES JR, P.; SOUBEYRAN, A. **Learning how to play nash, potential games and alternating minimization method for structured nonconvex problems on riemannian manifolds**. *J. Convex Anal*, 20(2):395–438, 2013.

[21] DA CRUZ NETO, J. X.; MELO, Í. D. L.; SOUSA, P. A.; DE OLIVEIRA SOUZA, J. C. **On the relationship between the kurdyka–łojasiewicz property and error bounds on hadamard manifolds**. *Journal of Optimization Theory and Applications*, 200(3):1255–1285, 2024.

[22] DE CARVALHO BENTO, G.; BITAR, S. D. B.; DA CRUZ NETO, J. X.; OLIVEIRA, P. R.; DE OLIVEIRA SOUZA, J. C. **Computing riemannian center of mass on hadamard manifolds**. *Journal of Optimization Theory and Applications*, 183(3):977–992, 2019.

[23] DE CARVALHO BENTO, G.; DA CRUZ NETO, J. X.; OLIVEIRA, P. R. **A new approach to the proximal point method: convergence on general riemannian manifolds**. *Journal of Optimization Theory and Applications*, 168(3):743–755, 2016.

[24] FEEHAN, P. **Resolution of singularities and geometric proofs of the łojasiewicz inequalities**. *Geometry & Topology*, 23(7):3273–3313, 2019.

[25] FRANKEL, P.; GARRIGOS, G.; PEYPOUQUET, J. **Splitting methods with variable metric for kurdyka–łojasiewicz functions and general convergence rates**. *Journal of Optimization Theory and Applications*, 165:874–900, 2015.

[26] JOHNSTONE, P. R.; MOULIN, P. **Local and global convergence of an inertial version of forward-backward splitting**. *arXiv preprint arXiv:1502.02281*, 2015.

[27] KARIMI, H.; NUTINI, J.; SCHMIDT, M. **Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition**. In: *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part I 16*, p. 795–811. Springer, 2016.

[28] KHANH, P. D.; MORDUKHOVICH, B. S.; TRAN, D. B. **Inexact reduced gradient methods in nonconvex optimization**. *Journal of Optimization Theory and Applications*, p. 1–41, 2023.

[29] KHANH, P. D.; MORDUKHOVICH, B. S.; TRAN, D. B. **Inexact reduced gradient methods in nonconvex optimization**. *Journal of Optimization Theory and Applications*, p. 1–41, 2023.

[30] KURDYKA, K. **On gradients of functions definable in o-minimal structures**. In: *Annales de l'institut Fourier*, volume 48, p. 769–783, 1998.

[31] LI, G.; MORDUKHOVICH, B.; ZHU, J. **Generalized metric subregularity with applications to high-order regularized newton methods**. *arXiv preprint arXiv:2406.13207*, 2024.

[32] Li, G.; Pong, T. K. **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.

[33] Lojasiewicz, S. **Une propriété topologique des sous-ensembles analytiques réels**. *Les équations aux dérivées partielles*, 117:87–89, 1963.

[34] Lojasiewicz, S. **Une propriété topologique des sous-ensembles analytiques réels**. *Les équations aux dérivées partielles*, 117:87–89, 1963.

[35] Mordukhovich, B. S. **Second-order variational analysis in optimization, variational stability, and control: theory, algorithms, applications**. Springer Nature, 2024.

[36] Mordukhovich, B. S. **Variational analysis and applications**, volume 30. Springer, 2018.

[37] Mordukhovich, B. **Variational Analysis and Generalized Differentiation I: Basic Theory**. Springer, Berlin, 2006.

[38] Ochs, P.; Chen, Y.; Brox, T.; Pock, T. **ipiano: Inertial proximal algorithm for nonconvex optimization**. *SIAM Journal on Imaging Sciences*, 7(2):1388–1419, 2014.

[39] Ortega, J. M.; Rheinboldt, W. C. **Iterative solution of nonlinear equations in several variables**. SIAM, 2000.

[40] Polyak, B. T. **Introduction to optimization**. *Optimization Software Inc*, 1987.

[41] Polyak, B. T. **Gradient methods for the minimisation of functionals**. *USSR Computational Mathematics and Mathematical Physics*, 3(4):864–878, 1963.

[42] Rebjock, Q.; Boumal, N. **Fast convergence to non-isolated minima: four equivalent conditions for c 2 functions**. *Mathematical Programming*, p. 1–49, 2024.

[43] Soubeyran, A. **Variational rationality, a theory of individual stability and change: worthwhile and ambidextry behaviors**. *Pre-print. GREQAM, Aix Marseillle University*, 2009.

[44] Soubeyran, A. **Variational rationality and the unsatisfied man: routines and the course pursuit between aspirations, capabilities and beliefs**. *Preprint. GREQAM, Aix Marseillle University*, 2010.

[45] SOUBEYRAN, A. **Variational rationality: Finding cheap and quick enough ways of motion to satisfice our recurrent and changing needs/desires**. *AMSE, Aix-Marseille University, submitted*, 2024.

[46] SOUBEYRAN, A.; SOUZA, J. C.; NETO, J. X. C. **How variational rational agents would play nash: A generalized proximal alternating linearized method**. In: *Variational Analysis and Set Optimization*, p. 25–43. CRC Press, 2019.

[47] VAN DEN DRIES, L.; MILLER, C. **Geometric categories and o-minimal structures**. *Duke Mathematical*, 1996.

[48] VON NEUMANN, J. **Functional operators: Measures and integrals**, volume 1. Princeton University Press, 1950.

APPENDIX A

```
import numpy as np
from scipy.optimize import minimize_scalar

# Função f(x) = |x|^{3/2}
def f(x):
    return np.abs(x)**1.5

# Passo proximal: minimiza f(x) + (1/(2*lam))(x - xk)^2
def prox_f(xk, lam):
    obj = lambda x: f(x) + (1 / (2 * lam)) * (x - xk)**2
    res = minimize_scalar(obj, bounds=(-200, 200), method=’bounded’)
    return res.x

# Parâmetros
lam = 100.0         # valor de lambda_k
xk = 100.0          # ponto inicial
num_iter = 10       # número de iterações

# Armazenar valores
x_vals = [xk]
f_vals = [f(xk)]

# Iterações do método
print(f"Iteração 0: x = {xk:.20f} \tf(x) = {f(xk):.20f}")
for k in range(1, num_iter + 1):
    xk = prox_f(xk, lam)
    fxk = f(xk)
    x_vals.append(xk)
    f_vals.append(fxk)
    print(f"Iteração {k}: x = {xk:.20f} \tf(x) = {fxk:.20f}")
```