Universidade Federal do Piauí
Centro de Ciências da Natureza
Pós-Graduação em Matemática
Mestrado em Matemática

Error bounds e desigualdade de Kurdyka-Łojasiewicz em complexidade de métodos de descida para funções convexas

Yldenilson Torres Almeida

Teresina - 2017

Yldenilson Torres Almeida

Dissertação de Mestrado:

Error bounds e desigualdade de Kurdyka-Łojasiewicz em complexidade de métodos de descida para funções convexas

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Matemática, da Universidade Federal do Piauí, como requisito parcial para obtenção do grau de Mestre em Matemática.

Orientador:
Prof. Dr. João Carlos de Oliveira Souza

Teresina - 2017

FICHA CATALOGRÁFICA
Serviço de Processamento Técnico da Universidade Federal do Piauí
Biblioteca Setorial do CCN

A447e Almeida, Yldenilson Torres.
Error bounds e desigualdade de Kurdyka-Lojasiewicz em complexidade de métodos de descida para funções convexas / Yldenilson Torres Almeida. – Teresina, 2017.
51f.

Dissertação (Mestrado) – Universidade Federal do Piauí, Centro de Ciências da Natureza, Pós-Graduação em Matemática, 2017.
Orientador: Prof. Dr.João Carlos de Oliveira Souza.

1. Matemática. 2. Error Bourds. 3. Minimização Convexa. I. Título

CDD 510

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PIAUÍ
CENTRO DE CIÊNCIAS DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Error bounds e desigualdade de Kurdyka-Lojasiewicz em complexidade de métodos de descida para funções convexas

YLDENILSON TORRES ALMEIDA

Esta Dissertação foi submetida como parte dos requisitos necessários à obtenção do grau de **Mestre em Matemática**, outorgado pela Universidade Federal do Piauí.

A citação de qualquer trecho deste trabalho é permitida, desde que seja feita em conformidade com as normas da ética científica.

Dissertação aprovada em 09 de Junho de 2017.

Banca Examinadora:

Prof. Dr. João Carlos de Oliveira Souza - orientador

Prof. Dr. João Xavier da Cruz Neto - membro interno

Prof. Dr. Pedro Antônio Soares Júnior - membro externo

Prof. Dr. Jurandir de Oliveira Lopes - membro suplente

Dedico este trabalho aos meus pais, Irisvan e Maria de Fátima, que sempre me inspiraram e me apoiaram incondicionalmente.

Agradecimentos

Agradeço primeiramente a Deus pelo dom da vida, por estar sempre ao meu lado nos momentos mais difíceis e me proporcionar vitórias como essa.

Agradeço aos meus pais Irisvan e Maria de Fátima que sempre me incentivaram e me deram todo o suporte preciso para que eu conseguisse chegar até aqui. Agradeço também aos meus irmãos Yuri e Igor pela convivência e pelo apoio que me dão diariamente.

Agradeço a todos da minha família: tios, tias, padrinhos, que sempre me motivam e me ajudam nos momentos precisos. Agradeço a meus primos e primas pelos vários momentos de diversão, conversas agradáveis e motivação, que sempre me ajudavam a sair do stress do dia-a-dia.

Agradeço a minha noiva Elaila, que sempre esteve do meu lado nessa jornada, me incentivando, me apoiando, me motivando, torcendo por mim e, acima de tudo, nunca deixou de acreditar em mim.

Agradeço ao meu orientador e amigo Prof. Dr. João Carlos pelos ensinamentos, pela motivação que me deu, por ter acreditado em mim e por estar sempre presente quando precisei. Agradeço também ao Prof. Dr. João Xavier que teve parcela fundamental nesse trabalho, sempre prestativo tirando dúvidas e fazendo observações valiosas.

Agradeço a todo o Departamento de Matemática da UFPI. Em especial aos professores: Jurandir, Barnabé, Paulo Alexandre, Newton, Marcos Vinícius, Isaias, Humberto, Mário, Liane, Rondinele, Kelton, Antônio Wilson, Marcondes, José Francisco, Jeferson e a todos os outros que tiveram participação direta ou indiretamente na minha formação.

Agradeço aos incontáveis amigos que ganhei ao longo de toda essa jornada, desde os tempos de graduação. A lista realmente é gigantesca e não citarei muitos nomes pois corro o sério risco de ser traído pela memória, mas não posso deixar de citar aqui em especial

João Augusto, Victor, Samuel e Yan, pelos vários momentos divertidos que passamos juntos, e ainda Edilson, Valéria, Hércules e Raul Kazan que sempre me socorriam nos momentos de dúvidas.

Agradeço ao Prof. Dr. Jurandir e ao Prof. Dr. Pedro Soares por comporem a banca, por lerem este trabalho e pelas valiosas sugestões feitas.

Agradeço a CAPES pelo importante apoio financeiro.

"*Disciplina é liberdade;*

Compaixão é fortaleza;

Ter bondade é ter coragem...".

Renato Russo

Resumo

Neste trabalho mostramos como o conceito de error bounds pode ser usado como uma ferramenta efetiva para se obter resultados de complexidade de métodos de descida de primeira ordem em programação convexa. Para isso, estudamos a relação entre os conceitos de error bounds e desigualdade de Kurdyka-Łojasiewicz. Usando esses conceitos, obtemos a complexidade do método via um algoritmo de ponto proximal unidimensional. Como aplicação, analisamos um método para resolver um problema de viabilidade convexa.
Palavras-chaves: Error bounds, desigualdade KL, minimização convexa.

Abstract

In this work, we show how the concept of error bounds can be used as an effective tool to obtain complexity results of a first order descent method in convex programming. To this end, we study the relationship between the concepts of error bounds and Kurdyka-Łojasiewicz inequality. Using these concepts, we obtain the complexity of the method by means of an unidimensional proximal point algorithm. As an application, we analyze a method for solving a convex feasibility problem.

Keywords: Error bounds, KL inequality, convex minimization.

Sumário

Introdução

Seja $f : D \to \mathbb{R}$ uma função real, onde $D \subset \mathbb{R}^n$ é um subconjunto não-vazio. Consideramos o problema de achar um minimizador de f no conjunto D. Este problema será escrito como

$$\min \{f(x) : x \in D\}.$$

Quando $D = \mathbb{R}^n$ dizemos que temos um *problema de minimização irrestrito*. Quando D é um conjunto convexo e f é uma função convexa dizemos que temos um *problema de minimização convexa* (ver [15]). Em quase todo este trabalho estaremos considerando $f : \mathbb{R}^n \to \mathbb{R}$ uma função própria, convexa, semicontínua inferiormente e assumiremos sempre que o conjunto dos minimizadores de f é diferente do vazio, isto é, $\arg\min f \neq \emptyset$.

Em sua forma mais simples, um *error bounds* é uma desigualdade da forma

$$\omega(f(x) - \min f) \geqslant \operatorname{dist}(x, \arg\min f),$$

onde ω é uma função crescente, nula no 0 - chamada aqui de *função residual* -, e onde x pode se expandir por todo o espaço ou em um conjunto limitado. As primeiras descobertas significativas na teoria de error bounds vieram no fim dos anos 50, com os pioneiros trabalhos de Łojasiewicz [19, 20]. Mostraremos que error bounds podem ser usados como uma ferramenta efetiva para se obter resultados de complexidade de métodos de descida de primeira ordem em programação convexa. Para tal tarefa usamos a *desigualdade de Kurdyka-Łojasiewicz*. Uma função f satisfaz a desigualdade de Kurdyka-Łojasiewicz (ou tem a *propriedade KL*) quando existe uma função φ suave, côncava tal que

$$\|\nabla(\varphi \circ (f - \min f))(x)\| \geqslant 1$$

para todo x em uma vizinhança do conjunto $\arg\min f$. A propriedade KL foi introduzida inicialmente em [6]. Mostraremos que se f é uma função convexa e o error bounds provém de uma função residual com o *comportamento moderado*, então error bounds é equivalente a desigualdade KL.

Uma vez sabendo que error bounds fornece uma desigualdade KL, precisamos fazer uma conexão desse fato com a complexidade dos métodos de primeira ordem. Para isto, consideraremos o *método de descida de primeira ordem*, o qual tem a forma

(i) $f(x) + a\|x_k - x_{k-1}\|^2 \leqslant f(x_{k-1})$,

(ii) $\|\omega_k\| \leqslant b\|x_k - x_{k-1}\|$, onde ω_k pertence ao subdiferencial de f no ponto x_k para todo $k \geqslant 1$

e associamos uma *sequência proximal unidimensional "pior caso"*

$$\alpha_{k+1} = \arg\min\left\{\varphi^{-1}(s) + \frac{1}{2\zeta}(s-\alpha_k)^2 : s \geqslant 0\right\},\ \alpha_0 = \varphi^{-1}(f(x_0)),$$

onde ζ é uma constante que depende explicitamente do terno de números reais positivos (a, b, l) onde $l > 0$ é uma constante Lipschitz de $(\varphi^{-1})'$. Essa idéia é o principal resultado do artigo proposto [7]. Nosso resultado de complexidade afirma que, sob fracas suposições, a sequência proximal unidimensional "pior caso"domina a complexidade do método original através da desigualdade

$$f(x_k) - \min f \ \leqslant\ \varphi^{-1}(\alpha_k),\ k \geqslant 0.$$

Em suma, nossa metodologia será a seguinte: obter um error bounds, calcular a *função desingularizante* sempre que possível, identificar as constantes essenciais no método de descida e, finalmente, obter a complexidade usando a sequência proximal unidimensional "pior caso". Os resultados descritos neste trabalho foram estabelecidos em Bolte et al. [7].

Dividimos este trabalho em quatro capítulos. No primeiro deles abordamos noções básicas de otimização e análise convexa, e o concluímos com um breve comentário sobre *operador proximal*. No segundo capítulo, introduzimos os conceitos de error bounds e desigualdade de Kurdyka-Łojasiewicz. Neste capítulo, definimos ainda *curvas subgradientes* e enunciamos alguns resultados sobre, que serão de fundamental importância na demonstração da equivalência entre error bounds e desigualdade de Kurdyka-Łojasiewicz. No terceiro capítulo apresentamos os métodos de descida de primeira ordem e a sua complexidade. No quarto capítulo, fazemos uma aplicação da teoria no problema de viabilidade convexa com interseção regular.

Capítulo 1

Noções Preliminares

Este capítulo é dedicado a apresentar alguns conceitos básicos, notações e resultados de otimização e análise convexa que serão usados ao longo da dissertação. Grande parte dos resultados aqui contido podem ser encontrados mais detalhadamente em [15, 28].

1.1 Condições de otimalidade e funções convexas

Nesta seção, apresentamos algumas definições e resultados sobre otimização, com enfoque maior àqueles resultados obtidos para funções convexas. O conceito de convexidade é muito importante na teoria de otimização. Veremos que com a noção de convexidade, as condições necessárias de otimalidade passam a ser suficientes. Assim, todo ponto estacionário torna-se solução do problema. Em particular, qualquer minimizador local é global. Todas as demonstrações que forem omitidas nesta seção podem ser facilmente encontradas em [15, 28].

Seja $f : D \subset \mathbb{R}^n \to \mathbb{R}$ uma função. O problema a ser considerado é o de encontrar um minimizador de f em D. Este problema é dito de minimização e será escrito como

$$\min \{f(x) : x \in D\}. \tag{1.1}$$

Dizemos que o conjunto D é o *conjunto viável* do problema. Os pontos de D serão chamados *pontos viáveis* e f será chamada *função objetivo*. Dizemos ainda que $\bar{v} \in [-\infty, +\infty)$ definido por

$$\bar{v} = \inf_{x \in D} f(x)$$

é o *valor ótimo* do problema (1.1).

Definição 1. *Dizemos que um ponto* $\overline{x} \in D$ *é:*

(a) minimizador global de (1.1), se

$$f(\overline{x}) \leqslant f(x), \ \forall x \in D;$$

(b) minimizador local de (1.1), se existe $\epsilon > 0$ *tal que*

$$f(\overline{x}) \leqslant f(x), \ \forall x \in D \cap B(\overline{x}, \epsilon).$$

Notemos que todo minimizador global é também minimizador local, mas o recíproco não é válido.

Chamamos de $\arg\min f$ o conjunto dos minimizadores globais do problema (1.1), isto é,

$$\arg\min f := \{\overline{x} \in D : f(\overline{x}) \leqslant f(x), \ \forall x \in D\} \ = \ \{\overline{x} \in D : f(\overline{x}) = \overline{v}\},$$

onde $\overline{v}$ é o valor ótimo do problema. Uma função pode admitir vários minimizadores globais, mas o valor ótimo do problema sempre é o mesmo. Em alguns casos, mesmo quando $\overline{v}$ é finito, pode não haver um minimizador global para o problema, isto é, $\arg\min f = \emptyset$. Como exemplo disso, citamos o problema:

$$\min \{e^x : \ x \in \mathbb{R}\}.$$

Com base nisso, apresentamos agora alguns critérios que garantem a existência de solução global.

Teorema 1. ***(Teorema de Weierstrass)***
Sejam $D \subset \mathbb{R}^n$ *um conjunto compacto não-vazio e* $f : D \to \mathbb{R}$ *uma função contínua. Então* f *assume máximo e mínimo em* D*. Em particular, o problema (1.1) tem solução global.*

Demonstração. É suficiente provar a existência de um minimizador. Como a imagem de um conjunto compacto por uma função contínua é compacto temos que o conjunto

$$\{y \in \mathbb{R} : y = f(x), \ x \in D\}$$

é um conjunto compacto. Em particular, este conjunto é limitado inferiormente. Seja $-\infty < \bar{v} = \inf_{x \in D} f(x)$. Pela definição de ínfimo, para todo $k \in \mathbb{N}$ existe um $x_k \in D$ tal que

$$\bar{v} \leqslant f(x_k) \leqslant \bar{v} + \frac{1}{k}.$$

Passando ao limite quando $k \to \infty$, concluimos que

$$\lim_{k \to \infty} f(x_k) = \bar{v}.$$

Como $(x_k)_{k \in \mathbb{N}} \subset D$ compacto, segue que $(x_k)_{k \in \mathbb{N}}$ é uma sequência limitada. Logo possui subsequência $(x_{k_j})_{j \in \mathbb{N}}$ que converge para um ponto de D, isto é,

$$\lim_{j \to \infty} x_{k_j} = \bar{x} \in D.$$

Pela continuidade de f,

$$\lim_{j \to \infty} f(x_{k_j}) = f(\bar{x}).$$

Daí,

$$\bar{v} = \lim_{k \to \infty} f(x_k) = \lim_{j \to \infty} f(x_{k_j}) = f(\bar{x}),$$

isto é, f assume o valor mínimo em D no ponto $\bar{x} \in D$. □

Definição 2. *O conjunto de nível da função* $f : D \subset \mathbb{R}^n \to \mathbb{R}$ *associado a* $c \in \mathbb{R}$*, é o conjunto dado por*

$$L_{f,D}(c) = \{x \in D : f(x) \leqslant c\}.$$

Corolário 1. *Sejam* $D \subset \mathbb{R}^n$ *e* $f : D \to \mathbb{R}$ *contínua no conjunto* D*. Suponhamos que existe* $c \in \mathbb{R}$ *tal que o conjunto de nível* $L_{f,D}(c)$ *seja não-vazio e compacto. Então o problema de minimizar* f *em* D *possui uma solução global.*

Demonstração. Pelo Teorema de Weierstrass, o problema

$$\min f(x) \text{ sujeito a } x \in L_{f,D}(c)$$

tem uma solução global, digamos $\bar{x}$. Para todo $x \in D \setminus L_{f,D}(c)$ temos que

$$f(x) > c \geqslant f(\bar{x}).$$

Isso mostra que $\bar{x}$ é um minimizador global de f não só em $L_{f,D}(c)$, mas também em D. □

Definição 3. *Dizemos que a função* $f : D \to \mathbb{R}$ *é coerciva no conjunto D, quando para cada sequência* $(x_k)_{k\in\mathbb{N}} \subset D$ *tal que* $\|x_k\| \to \infty$ *ou* $x_k \to x \in \overline{D} \setminus D$ $(k \to \infty)$*, tem-se* $\limsup_{k\to\infty} f(x_k) = +\infty$.

Corolário 2. *Sejam* $D \subset \mathbb{R}^n$ *e* $f : D \to \mathbb{R}$ *uma função contínua coerciva em* $D \neq \emptyset$. *Então, o problema de minimizar* f *em* D *possui uma solução global.*

Demonstração. Escolha um $\overline{x} \in D$ e defina $c = f(\overline{x})$. Com esta escolha, o conjunto de nível $L_{f,D}(c)$ é não-vazio. Provaremos que ele é compacto, mostrando que ele é limitado e fechado.
Suponhamos que $L_{f,D}(c)$ é ilimitado. Então existe uma sequência $(x_k)_{k\in\mathbb{N}} \subset L_{f,D}(c)$ tal que $\|x_k\| \to +\infty$ $(k \to \infty)$. Como $(x_k)_{k\in\mathbb{N}} \subset D$ e f é coerciva em D,

$$\limsup_{k\to\infty} f(x_k) = +\infty.$$

Por outro lado, $f(x_k) \leqslant c$ para todo k, pela definição de conjunto de nível. Esta contradição implica que $L_{f,D}(c)$ é limitado. Suponhamos agora que $L_{f,D}(c)$ não seja fechado. Então existe uma sequência $(x_k)_{k\in\mathbb{N}} \subset L_{f,D}(c)$ tal que $x_k \to x$ $(k \to \infty)$, com $x \in \overline{L_{f,D}(c)} \setminus L_{f,D}(c)$. Como

$$L_{f,D}(c) = D \cap \{y \in \mathbb{R}^n : f(y) \leqslant c\},$$

e o conjunto $\{y \in \mathbb{R}^n : f(y) \leqslant c\}$ é fechado pela continuidade de f, isto significa que $x_k \to x \in \overline{D} \setminus D$. Agora, a coercividade de f em D implica em

$$\limsup_{k\to\infty} f(x_k) = +\infty.$$

Novamente isso contradiz o fato de que $(x_k)_{k\in\mathbb{N}} \subset L_{f,D}(c)$. Portanto, $L_{f,D}(c)$ é fechado. Concluimos assim que $L_{f,D}(c)$ é não vazio e compacto. A afirmação segue do Corolário 1. $\square$

Nosso próximo teorema nos fornece uma condição necessária para que um ponto $\overline{x} \in D$ seja um minimizador de f em D. Mais a frente veremos que essa condição será também suficiente se a função for convexa e diferenciável.

Teorema 2. ***(Condição necessária de primeira ordem)***
Suponhamos que $D \subset \mathbb{R}^n$ *é um conjunto aberto e* $f : D \to \mathbb{R}$ *é uma função diferenciável. Se* $\overline{x}$ *é um minimizador local de* f *em* D *então* $\nabla f(\overline{x}) = 0$.

Demonstração. Veja página 17 de [15]. □

Definição 4. *Dizemos que uma função* $f : D \to \mathbb{R}$ *é semicontínua inferiormente no ponto* $x \in D \subset \mathbb{R}^n$*, quando para qualquer sequência* $(x_k)_{k\in\mathbb{N}} \subset D$ *tal que* $x_k \to x$ $(k \to \infty)$*, tem-se*

$$\liminf_{k\to\infty} f(x_k) \geqslant f(x).$$

Uma função f *é dita semi-contínua inferiormente no conjunto D, quando ela é semi-contínua inferiormente em todos os pontos de* D*.*

Definição 5. *Uma função* $f : D \to \mathbb{R}$ *é L-Lipschitz contínua quando existe uma constante* $L > 0$ *tal que*

$$\|f(x) - f(y)\| \leqslant L\|x - y\|, \ \forall x, y \in D.$$

A seguir, apresentamos as definições de conjunto convexo e de funções convexas. Tais conceitos podem ser encontrados com mais detalhes em [15, 28].

Definição 6. *Um conjunto* $D \subset \mathbb{R}^n$ *é chamado conjunto convexo se para quaisquer* $x, y \in D$ *e* $\alpha \in [0, 1]$*, tem-se que* $\alpha x + (1 - \alpha y) \in D$*.*

Exemplo 1. *Dados* $a \in \mathbb{R}^n$ *e* $r > 0$*, a bola* $B(a, r) = \{x \in \mathbb{R}^n : \|x - a\| \leqslant r\}$ *é um conjunto convexo.*
De fato, dados $x, y \in B(a, r)$ *e* $\alpha \in [0, 1]$ *temos que*

$$\begin{aligned}
\|\alpha x + (1-\alpha)y - a\| &= \|\alpha(x-a) + (1-\alpha)(y-a)\| \\
&\leqslant \alpha\|x-a\| + (1-\alpha)\|y-a\| \\
&\leqslant \alpha r + (1-\alpha)r = r.
\end{aligned}$$

Logo $\alpha x + (1 - \alpha)y \in B(a, r)$*.*

Uma projeção (ortogonal) do ponto $x \in \mathbb{R}^n$ sobre um conjunto $D \subset \mathbb{R}^n$ é um ponto de D que está mais próximo de x (onde a distância é medida pela norma euclidiana). Em outras palavras, uma projeção de x sobre D é a solução (global) do problema

$$\min \ \|y - x\| \text{ sujeito a } y \in D.$$

Nosso próximo teorema diz que se D é um conjunto convexo e fechado, então tal projeção é única.

Teorema 3. ***(Teorema da projeção)***

Seja $D \subset \mathbb{R}^n$ *um conjunto convexo e fechado. Então para todo* $x \in \mathbb{R}^n$*, a projeção de* x *sobre* D*, denotada por* $P_D(x)$*, existe e é única. Além disso,* $\overline{x} = P_D(x)$ *se, e somente se,*

$$\overline{x} \in D, \ \langle x - \overline{x}, y - \overline{x} \rangle \leqslant 0, \ \forall y \in D.$$

Demonstração. Ver página 105 de [15]. □

Corolário 3. ***(Operador de projeção é 1-Lipschitz contínuo)***

Seja $D \subset \mathbb{R}^n$ *um conjunto convexo e fechado. Então para* $x \in \mathbb{R}^n$ *e* $y \in \mathbb{R}^n$ *quaisquer,*

$$\|P_D(x) - P_D(y)\| \leqslant \|x - y\|.$$

Em particular, $P_D(\cdot)$ *é contínuo no* $\mathbb{R}^n$*.*

Demonstração. Ver página 109 de [15] □

Definição 7. *Seja* $D \subset \mathbb{R}^n$ *um conjunto convexo. Dizemos que uma função* $f : D \to \mathbb{R}$ *é convexa em* D *quando*

$$f(\alpha x + (1 - \alpha y)) \leqslant \alpha f(x) + (1 - \alpha) f(y), \tag{1.2}$$

para todos $x, y \in D$ *e* $\alpha \in [0, 1]$*.*

Uma função f *é dita estritamente convexa quando a desigualdade (1.2) for estrita para todos* $x \neq y$ *e* $\alpha \in (0, 1)$*.*

Note que toda função estritamente convexa é convexa. Já o recíproco não ocorre, como podemos verificar no exemplo abaixo.

Exemplo 2. *A função* $f : \mathbb{R} \to \mathbb{R}$*, dada por* $f(x) = x^2$ *é estritamente convexa, logo é convexa. Já a função* $g : \mathbb{R} \to \mathbb{R}$*, dada por* $g(x) = x$ *é convexa mas não é estritamente convexa.*

O resultado a seguir ressalta a importância de funções convexas em otimização. Até o fim desta seção, consideraremos $D \subset \mathbb{R}^n$ um conjunto convexo.

Teorema 4. ***(Teorema de minimização convexa)***

Seja $f : D \to \mathbb{R}$ *uma função convexa em* D*. Então todo minimizador local em (1.1) é global. Além disso, o conjunto dos minimizadores é convexo.*

Demonstração. Suponha que $\bar{x} \in D$ seja um minimizador local que não é global. Então existe $y \in D$ tal que $f(y) < f(\bar{x})$. Definimos $x(\alpha) = \alpha y + (1-\alpha)\bar{x}$. Pela convexidade de f, para todo $\alpha \in (0, 1]$, tem-se

$$f(x(\alpha)) = f(\alpha y + (1-\alpha)\bar{x}) \leqslant \alpha f(y) + (1-\alpha) f(\bar{x}) = f(\bar{x}) + \alpha(f(y) - f(\bar{x})) < f(\bar{x}).$$

Tomando $\alpha > 0$ suficientemente pequeno, podemos garantir que o ponto $x(\alpha)$ está arbitrariamente próximo do ponto $\bar{x}$, e ainda tem-se que $f(x(\alpha)) < f(\bar{x})$ com $x(\alpha) \in D$. Isto contradiz o fato de que $\bar{x}$ é um minimizador local. Logo toda solução local é global. Sejam agora $S \subset D$ o conjunto dos minimizadores de f e $\bar{v} \in \mathbb{R}$ o valor ótimo do problema. Para quaisquer $x, \bar{x} \in S$ e $\alpha \in [0, 1]$, pela convexidade de f obtemos

$$f(\alpha x + (1-\alpha)\bar{x}) \leqslant \alpha f(x) + (1-\alpha) f(\bar{x}) = \alpha \bar{v} + (1-\alpha)\bar{v} = \bar{v}.$$

Se $\bar{v} = \inf_{x \in D} f(x)$ e $\alpha x + (1-\alpha)\bar{x} \in D$, então podemos deduzir da desigualdade acima que

$$f(\alpha x + (1-\alpha)\bar{x}) = \bar{v},$$

isto é, $\alpha x + (1-\alpha)\bar{x} \in S$. Logo S é convexo. □

Observação 1. *Se f é estritamente convexa, não pode haver mais de um minimizador. Se existissem $x,\ \bar{x} \in S$ com $x \neq \bar{x}$ podíamos tomar qualquer $\alpha \in (0, 1)$ que, pela convexidade estrita de f, teríamos*

$$f(\alpha x + (1-\alpha)\bar{x}) < \alpha f(x) + (1-\alpha) f(\bar{x}) = \bar{v}.$$

Como o conjunto S é convexo, temos uma solução do problema dada por $\alpha x + (1-\alpha)\bar{x}$ com valor da função objetivo menor do que o valor ótimo $\bar{v}$. Isto é uma contradição. Logo o minimizador do problema deve ser único.

Os teoremas a seguir caracterizam uma função convexa. Suas demonstrações não serão feitas aqui, porém todas podem ser facilmente encontradas em [15, 28].

Teorema 5. *Sejam $f_i : D \to \mathbb{R},\ i = 1, ..., p$, funções convexas em D. Então para quaisquer $\mu_i > 0,\ i = 1, ..., p$, a função*

$$f : D \to \mathbb{R},\ f(x) = \sum_{i=1}^{p} \mu_i f_i(x)$$

é convexa em D.

Demonstração. Ver página 141 de [15] □

Teorema 6. *Sejam $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa e $g : \mathbb{R} \to \mathbb{R}$ uma função convexa e não-decrescente. Então a função*

$$\psi : \mathbb{R}^n \to \mathbb{R}, \ \psi(x) = g(f(x))$$

é convexa.

Demonstração. Ver página 143 de [15]. □

Teorema 7. ***(Caracterização de funções convexas diferenciáveis)***
Sejam $D \subset \mathbb{R}^n$ um conjunto aberto convexo e $f : D \to \mathbb{R}$ uma função diferenciável em D. Então as seguintes propriedades são equivalentes:

(i) A função f é convexa em D;

(ii) Para todo $x \in D$ e todo $y \in D$,

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle;$$

(iii) Para todo $x \in D$ e todo $y \in D$,

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geqslant 0.$$

Além disso, se f é estritamente convexa, estas caracterizações seguem com as desigualdades estritas.

Demonstração. Ver página 159 de [15]. □

Corolário 4. *Seja $f : D \subset \mathbb{R}^n \to \mathbb{R}$ uma função convexa diferenciável no conjunto D aberto. Se $\nabla f(\overline{x}) = 0$, então $\overline{x}$ é minimizador global de f em D.*

Demonstração. Pelo teorema anterior, item (ii), para todo $y \in D$ temos

$$f(y) \geqslant f(\overline{x}) + \langle \nabla f(\overline{x}), y - \overline{x} \rangle = f(\overline{x}).$$

Daí,

$$f(\overline{x}) \leqslant f(y), \ \forall y \in D.$$

Portanto $\overline{x}$ é minimizador global de f em D. □

Definição 8. *Se* $D \subset \mathbb{R}^n$ *é um conjunto convexo, dizemos que* $f : D \to \mathbb{R}$ *é uma função côncava em D, quando a função* $(-f)$ *é convexa em* D.

Notemos que as afirmações do Teorema 4 são verdadeiras se substituimos minimização de uma função convexa num conjunto convexo por maximização de uma função côncava num conjunto convexo.

Nossa próxima definição funciona, em um certo sentido, como uma derivada generalizada para funções convexas não necessariamente diferenciáveis.

Definição 9. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função convexa. Dado* $x \in \mathbb{R}^n$, *o subdiferencial de f em x é definido como sendo*

$$\partial f(x) = \{u \in \mathbb{R}^n : f(y) \geqslant f(x) + \langle u, y - x \rangle, \forall y \in \mathbb{R}^n\}.$$

Exemplo 3. *Seja* $f : \mathbb{R} \to \mathbb{R}$, *dada por* $f(x) = |x|$. *Verifica-se que:*

$$\partial f(x) = \begin{cases} -1 & , \textit{ se } \quad x < 0 \\ [-1, 1] & , \textit{ se } \quad x = 0 \\ 1 & , \textit{ se } \quad x > 0 \end{cases}$$

Neste exemplo, o ponto $\bar{x} = 0$ é um minimizador global. Note ainda que $0 \in \partial f(0)$, porém $0 \notin \partial f(x)$, para todo $x \neq 0$. Mais geralmente, vale a seguinte proposição:

Proposição 1. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função convexa. O ponto* $\bar{x} \in \mathbb{R}^n$ *é mínimo global de* f *se, somente se,* $0 \in \partial f(\bar{x})$.

Demonstração. Segue diretamente da definição de subdiferencial. $\square$

Se uma função f é diferenciável, então vale o seguinte resultado:

Proposição 2. *Uma função convexa* $f : \mathbb{R}^n \to \mathbb{R}$ *é diferenciável no ponto* $x \in \mathbb{R}^n$ *se, somente se, o conjunto* $\partial f(x)$ *contém um elemento só. Neste caso,* $\partial f(x) = \{f'(x)\}$.

Demonstração. Ver página 179 de [15]. $\square$

Proposição 3. *Seja* $(x_k, \omega_k)_{k \in \mathbb{N}}$ *uma sequência em* $\mathbb{R}^n \times \mathbb{R}^n$ *tal que* $(x_k, \omega_k) \in \operatorname{graf} \partial f$ *para todo* $k \in \mathbb{N}$. *Se* (x_k, ω_k) *converge para* (x, ω), *e* $f(x_k)$ *converge para* $f(x)$ *então* $(x, \omega) \in \operatorname{graf} \partial f$.

Demonstração. Veja página 441 de [1]. $\square$

Observação 2. *De acordo com a proposição acima, se mostrarmos que existe uma sequência* $(x_k, \omega_k) \in \text{graf } \partial f$ *tal que* $(x_k, \omega_k) \to (\bar{x}, 0)$ *e* $f(x_k) \to f(\bar{x})$*, então* $\partial f(\bar{x}) = 0$.

A seguinte proposição nos fornece uma regra de cálculo para funções convexas não-diferenciáveis. Podemos encontrar mais propriedades deste tipo de função em [15].

Proposição 4. ***(O subdiferencial da soma de funções convexas)***
Sejam $f_i : \mathbb{R}^n \to \mathbb{R},\ \ i = 1, \ldots, p$ *funções convexas. Então*

$$\partial\left(\sum_{i=1}^{p} f_i(x)\right) = \sum_{i=1}^{p} \partial f_i(x).$$

Demonstração. Ver página 184 de [15]. □

Teorema 8. ***(O subdiferencial de uma função convexa)***
Se $f : \mathbb{R}^n \to \mathbb{R}$ *é uma função convexa então, para todo* $x \in \mathbb{R}^n$*, o conjunto* $\partial f(x)$ *é convexo, compacto e não-vazio.*

Demonstração. Ver página 176 de [15]. □

Observamos que $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ é uma aplicação ponto-conjunto, com domínio

$$\text{dom } \partial f = \{x \in \mathbb{R}^n : \partial f(x) \neq \emptyset\}.$$

Para cada $x \in \text{dom } \partial f$, denotaremos por $\partial^0 f(x)$ o elemento de menor norma de $\partial f(x)$. O vetor $\partial^0 f(x)$ existe e é único, pois é a projeção de $0 \in \mathbb{R}^n$ no conjunto não-vazio, fechado e convexo $\partial f(x)$. Daí temos que

$$\|\partial^0 f(x)\| = \text{dist}(0, \partial f(x)).$$

Quando $x \notin \text{dom } \partial f$ temos $\|\partial^0 f(x)\| = +\infty$. Adotamos por convenção $s \times (+\infty) = +\infty$ para todo $s > 0$.

1.2 Operador Proximal

De agora em diante, assumimos que $f : \mathbb{R}^n \to \mathbb{R}$ é uma função própria, convexa e semicontínua inferiormente. Estamos interessados em algumas propriedades da função f na vizinhança de seus minimizadores. Vamos supor que o conjunto dos minimizadores é diferente do vazio e o denotaremos por S. Assumiremos ainda, sem perda de generalidade, que $\min f = 0$.

Dado $\lambda > 0$, definamos o *operador proximal* como a aplicação ponto-conjunto $\mathrm{prox}_{\lambda f}(x) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ definida através da fórmula

$$\mathrm{prox}_{\lambda f}(x) := \arg\min \left\{ f(y) + \frac{1}{2\lambda}\|y - x\|^2 : y \in \mathbb{R}^n \right\}.$$

Observação 3.

(a) A função $f(y) + \frac{1}{2\lambda}\|y - x\|^2$ é coerciva.

Uma vez que estamos assumindo $S = \arg\min f \neq \emptyset$ temos que f é limitada inferiormente. Assim, existe $L \in \mathbb{R}$ tal que $L < f(x)$ para todo $x \in \mathbb{R}^n$. Daí, tomando uma sequência $(x_k)_{k\in\mathbb{N}} \subset \mathbb{R}^n$ tal que $\|x_k\| \to +\infty$ e usando a Desigualdade de Cauchy-Schwarz, obtemos

$$\begin{aligned}
\lim_{k\to\infty} \left(f(x^k) + \frac{1}{2\lambda}\|x^k - x\|^2 \right) &\geqslant L + \lim_{k\to\infty} \frac{1}{2\lambda}\|x^k - x\|^2 \\
&= L + \frac{1}{2\lambda}\|x\|^2 + \frac{1}{2\lambda} \lim_{k\to\infty} \left(\|x^k\|^2 - 2\langle x^k, x\rangle \right) \\
&\geqslant L + \frac{1}{2\lambda}\|x\|^2 + \frac{1}{2\lambda} \lim_{k\to\infty} \left(\|x^k\|^2 - 2\|x^k\|\|x\| \right) \\
&= +\infty,
\end{aligned}$$

isto é,

$$\lim_{k\to\infty} \left(f(x_k) + \frac{1}{2\lambda}\|x_k - x\|^2 \right) = +\infty.$$

Segue da definição que $f(y) + \frac{1}{2\lambda}\|y - x\|^2$ é coerciva. Logo, tem um minimizador global, garantido pelo Teorema 2.

(b) A função $f(y) + \frac{1}{2\lambda}\|y - x\|^2$ é estritamente convexa.

Como f é convexa e $\|\cdot - x\|^2$ é estritamente convexa, segue do Teorema 5 que a soma $f(y) + \frac{1}{2\lambda}\|y - x\|^2$ é estritamente convexa. Além disso, o Teorema 4 garante que se existe um minimizador, então ele é único.

A observação feita acima nos garante que se f é uma função convexa, então $\mathrm{prox}_{\lambda f}(x)$ tem um único elemento. Agora, se $x^0 \in \mathbb{R}^n$, o *algoritmo proximal* se escreve

$$x^{k+1} \in \mathrm{prox}_{\lambda_k f}(x^k), \tag{1.3}$$

onde λ_k é uma sequência de passos tal que $\lambda_k \in [\lambda^-, \lambda^+] \subset (0, +\infty)$. Afirmamos que a sequência $(x^k)_{k\in\mathbb{N}}$ gerada pelo algoritmo (1.3) satisfaz as seguintes condições:

(i) Para cada $k \in \mathbb{N}$,

$$f(x^{k+1}) + \frac{1}{2\lambda_k}\|x^{k+1} - x^k\|^2 \leqslant f(x^k). \tag{1.4}$$

(ii) Para cada $k \in \mathbb{N}$, existe $w^{k+1} \in \partial f(x^{k+1})$ tal que

$$\lambda_k w^{k+1} + x^{k+1} - x^k = 0. \tag{1.5}$$

De fato, como $x^{k+1} \in \text{prox}_{\lambda_k f}(x^k)$ tem-se que

$$f(x^{k+1}) + \frac{1}{2\lambda_k}\|x^{k+1} - x^k\|^2 \leqslant f(y) + \frac{1}{2\lambda_k}\|y - x^k\|^2 \ , \ \forall y \in \mathbb{R}^n.$$

Em particular se $y = x^k$, obtemos (1.4).

Por outro lado, afirmamos que $w^{k+1} = -\frac{1}{\lambda_k}(x^{k+1} - x^k) \in \partial f(x^{k+1})$. De fato, como $x^{k+1} \in \text{prox}_{\lambda_k f}(x^k)$, segue da Proposição 1 que

$$0 \in \partial\left(f(x^{k+1}) + \frac{1}{2\lambda_k}\|x^{k+1} - x^k\|^2\right) = \partial f(x^{k+1}) + \frac{1}{\lambda_k}(x^{k+1} - x^k),$$

ou seja,

$$-\frac{1}{\lambda_k}(x^{k+1} - x^k) \in \partial f(x^{k+1}).$$

Portanto $w^{k+1} = -\frac{1}{\lambda_k}(x^{k+1} - x^k) \in \partial f(x^{k+1})$, com

$$\lambda_k w^{k+1} + x^{k+1} - x^k = 0.$$

O conceito de operador proximal será útil no decorrer dos próximos capítulos.

Capítulo 2

Error bounds e desigualdade de Kurdyka-Łojasiewicz

Neste capítulo, definiremos *curvas subgradientes, desigualdade de Kurdyka-Łojasiewicz e error bounds*. Veremos ainda um resultado de equivalência geral entre error bounds e desigualdade de Kurdyka-Łojasiewicz (ou abreviadamente, desigualdade KL), que pode ser consultado em Bolte et al. [7]. Nosso principal objetivo é estabelecer um caminho simples e natural de calcular expoentes Łojasiewicz e, mais geralmente, funções desingularizantes. A importância dessa equivalência vem do fato de que a desigualdade KL com constante conhecida é, em geral, mais difícil de se estabelecer enquanto que error bounds são mais maleáveis, isto é, são mais fáceis de se estabelecer (ver, por exemplo, [17]). Além disso, o fato destas duas noções serem equivalentes abre uma grande raio de possibilidades quando vamos analisar complexidade de algoritmos.

2.1 Curvas Subgradientes

Considere a inclusão diferencial

$$\begin{cases} \dot{y}(t) \in -\partial f(y(t)), \text{ para quase todo } t \in (0, +\infty) \\ y(0) = x. \end{cases}$$

onde $x \in \overline{\text{dom } f}$ e $y(\cdot)$ é uma curva absolutamente contínua em $\mathbb{R}^n$. As principais propriedades deste sistema estão resumidas no seguinte:

Teorema 9. *Para cada* $x \in \overline{\text{dom } f}$*, existe uma única curva absolutamente contínua* $\chi_x : [0, \infty) \to \mathbb{R}^n$ *tal que* $\chi_x(0) = x$ *e*

$$\dot{\chi}_x(t) \in -\partial f(\chi_x(t))$$

para quase todo $t > 0$*. Além disso,*

(i) $\frac{d}{dt}\chi_x(t^+) = -\partial^0 f(\chi_x(t^+))$ *para todo* $t > 0$*;*

(ii) $\frac{d}{dt} f(\chi_x(t^+)) = -\|\dot{\chi}_x(t^+)\|^2$ *para todo* $t > 0$*;*

(iii) Para cada $z \in S$*, a função* $t \mapsto \|\chi_x(t) - z\|$ *é decrescente;*

(iv) $\chi_x(t)$ *converge para algum* $\overline{x} \in S$*, quando* $t \to \infty$*.*

Demonstração. Consultar referências [9] e [10]. □

O resultado acima (junto com sua demonstração) foi feito originalmente em espaço de Hilbert e pode ser encontrado em [9], exceto a parte (iv) que foi provado em [10]. A trajetória $t \mapsto \chi_x(t)$ é chamada *curva subgradiente*.

Dados $x \in \overline{\text{dom } f}$, e $0 \leqslant t < s$, escrevemos

$$\text{comp}(\chi_x, t, s) = \int_t^s \|\dot{\chi}_x(\tau)\| d\tau$$

como sendo o *comprimento de* χ_x variando de t a s.

2.2 Desigualdade de Kurdyka-Łojasiewicz

Nesta seção apresentamos a *desigualdade de Kurdyka-Lojasiewicz.* Para simplificar a notação, escreveremos $[f < \mu] = \{x \in \mathbb{R}^n : f(x) < \mu\}$. Seja $r_0 > 0$ e defina

$$\mathcal{K}(0, r_0) = \{\varphi \in C^0[0, r_0) \cap C^1(0, r_0), \varphi(0) = 0, \ \varphi \text{ concava e } \varphi' > 0\}.$$

A função f satistaz a *desigualdade de Kurdyka-Lojasiewicz* (ou tem a propriedade KL) localmente em $\overline{x} \in \text{dom } f$ se existem $r_0 > 0$, $\varphi \in \mathcal{K}(0, r_0)$ e $\epsilon > 0$ tais que

$$\varphi'(f(x) - f(\overline{x}))\, \text{dist}(0, \partial f(x)) \geqslant 1$$

para todo $x \in B(\overline{x}, \epsilon) \cap [f(\overline{x}) < f < f(\overline{x}) + r_0]$. A função φ acima é chamada *função desingularizante* de f em $\overline{x}$. Mostraremos agora que se $\overline{x}$ não é um minimizador de f, a desigualdade KL é satisfeita em $\overline{x}$. Para isto, precisaremos do seguinte:

Lema 1. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função própria e semicontínua inferiormente. Seja* $\overline{x} \in \operatorname{dom} \partial f$ *tal que* $0 \notin \partial f(\overline{x})$. *Se*

$$\|x - \overline{x}\| + |f(x) - f(\overline{x})| < \delta, \text{ com } \delta > 0,$$

então

$$\operatorname{dist}(0, \partial f(x)) \geqslant \delta.$$

Demonstração. Suponha por absurdo que existe uma sequência $(\delta_k)_{k\in\mathbb{N}}$ com $\delta_k > 0$, $\delta_k \to 0$ e uma sequência $(x_k)_{k\in\mathbb{N}}$ tal que

$$\|x_k - \overline{x}\| + |f(x_k) - f(\overline{x})| < \delta_k, \text{ e } \operatorname{dist}(0, \partial f(x_k)) < \delta_k. \tag{2.1}$$

Como $\partial f(x_k)$ é fechado, existe uma sequência $\omega_k \in \partial f(x_k)$ tal que $\|\omega_k\| = \operatorname{dist}(0, \partial f(x_k))$. Segue que $\|\omega_k\| < \delta_k$, e consequentemente, $\omega_k \to 0$. Por outro lado fazendo $k \to +\infty$ em (2.1), obtemos

$$x_k \to \overline{x} \text{ e } f(x_k) \to f(\overline{x}).$$

Logo, segue da Proposição 3 que $0 \in \partial f(\overline{x})$, o que contradiz a hipótese. □

Proposição 5. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função própria, semicontínua inferiormente e* $\overline{x} \in \operatorname{dom} \partial f$ *tal que* $0 \notin \partial f(\overline{x})$. *Então a desigualdade de Kurdyka-Łojasiewicz é válida em* $\overline{x}$.

Demonstração. Se $\overline{x}$ não é um ponto crítico de f e $\partial f(\overline{x})$ é um conjunto fechado, temos que

$$\delta := \operatorname{dist}(0, \partial f(\overline{x})) > 0.$$

Seja $\varphi(s) := s/\delta$ e considere $r_0 := \delta/2$. Daí, para todo $x \in B(\overline{x}, \delta/2) \cap [f(\overline{x}) < f < f(\overline{x}) + r_0]$, temos

$$\|x - \overline{x}\| + |f(x) - f(\overline{x})| < \frac{\delta}{2} + r_0 = \delta.$$

Segue do Lema 1 que para todo $x \in \operatorname{dom} \partial f$

$$\varphi'(f(x) - f(\overline{x})) \operatorname{dist}(0, \partial f(x)) = \frac{1}{\delta} \operatorname{dist}(0, \partial f(x)) < \frac{\delta}{\delta} = 1.$$

□

Sendo assim, focaremos no caso em que $\overline{x} \in S$. Desde que $f(\overline{x}) = 0$, a desigualdade KL torna-se

$$\varphi'(f(x))\|\partial^0 f(x)\| \geqslant 1$$

para $x \in B(\overline{x}, \epsilon) \cap [0 < f < r_0]$. A função f tem a propriedade KL em S, se ela o tem em cada ponto de S.

Exemplo 4. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função duas vezes diferenciável. Dizemos que* f *é uma função de Morse se em cada ponto crítico* $\overline{x}$ *de* f *a matriz hessiana de* f *em* $\overline{x}$ *tem todos os seus autovalores diferentes de zero. Toda função de Morse satisfaz a propriedade KL em cada ponto, com função desingularizante* $\varphi(t) = c\sqrt{t}$ *para algum* $c > 0$ *(ver Teorema 3.8 em [11]). Em particular, segue do Teorema 1.2 em [30], que toda função fortemente convexa que é duas vezes diferenciável satisfaz a propriedade KL em cada ponto.*

O seguinte teorema afirma que, sob certas condições, podemos garantir a existência da propriedade KL numa vizinhança de cada ponto no dom f. Sua demonstração pode ser encontrada em [6].

Teorema 10. *Se* $f : \mathbb{R}^n \to \mathbb{R}$ *é uma função própria, convexa, semicontínua inferiormente e semi-algébrica, então ela tem a propriedade KL numa vizinhança de cada ponto no* dom f.

Demonstração. Consultar [6]. □

2.3 Error Bounds

Consideremos uma função não-decrescente $\omega : [0, +\infty[\to [0, +\infty[$ com $\omega(0) = 0$. A função f satisfaz um *error bounds local* com *função residual* ω se existe $r_0 > 0$ tal que

$$(\omega \circ f)(x) \geqslant \operatorname{dist}(x, S)$$

para todo $x \in [0 \leqslant f \leqslant r_0]$. De importância particular é o caso quando $\omega(s) = \gamma^{-1} s^{\frac{1}{p}}$ com $\gamma > 0$ e $p \geqslant 1$, na qual

$$f(x) \geqslant \gamma \operatorname{dist}(x, S)^p$$

para todo $x \in [0 \leqslant f \leqslant r_0]$.

Desde o celebrado resultado de Hoffman em error bounds para sistemas de desigualdades lineares [14], o estudo de error bounds foi sendo aplicado com sucesso em problemas de sensitividade, estimativa de raio de convergência e problemas de viabilidade. No mundo da otimização, as primeiras extensões naturais foram feitas para funções convexas por

Robinson [29], Mangasarian [25] e Auslender-Crouzeix [3]. Entretanto, a descoberta mais impressionante veio anos antes, nos primeiros trabalhos de Łojasiewicz [19, 20] no fim dos anos 50: sob uma mera condição de compacidade, podíamos garantir a existência de error bounds para funções contínuas semi-algébricas arbitrárias. Apesar da profundidade desse resultado, esses trabalhos permaneceram despercebidos pela comunidade da otimização durante um longo período (ver [22]). No início dos anos 90, motivados pelas numerosas aplicações, muitos pesquisadores começaram a trabalhar nessas linhas, em busca de resultados quantitativos que pudessem produzir mais ferramentas efetivas. As pesquisas de Pang [26] nos fornece um panorama dos resultados obtidos nesse tempo. Os trabalhos de Luo [21, 23] e Dedieu [12] também são marcos importantes na teoria.

Enunciaremos agora o teorema feito por Łojasiewicz em seu trabalho, citado acima, no fim dos anos 50. Sua demonstração pode ser consultada em [19, 20].

Teorema 11. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função própria, convexa, semicontínua inferiormente e semi-algébrica, e assuma que* $\arg\min f$ *é não-vazio e compacto. Então* f *tem um error bounds global*

$$f(x) + f(x)^{\frac{1}{p}} \geqslant \gamma_0 \operatorname{dist}(x, \arg\min f), \forall x \in \mathbb{R}^n,$$

onde $\gamma_0 > 0$ *e* $p \geqslant 1$ *é um número racional.*

Demonstração. Consultar [19]. □

2.4 Error bounds com função residual moderada e desigualdade Łojasiewicz

Error bounds frequentemente ou tem a forma de potência ou são tipo Hölder. Eles podem ser muito simples $s \to as^p$ ou exibir duas somas, por exemplo, $s \to as^p + bs^q$. Uma função $\varphi : [0, r) \to \mathbb{R}$ em $C^1(0, r) \cap C^0[0, r)$ e nula na origem, tem o *comportamento moderado (próximo da origem)* se ela satisfaz a equação diferencial do tipo

$$s\varphi'(s) \geqslant c\varphi(s), \ \forall s \in (0, r),$$

onde c é uma constante positiva. Observemos que pela concavidade de φ temos necessariamente $c \leqslant 1$.

Exemplo 5. *Considere* $\varphi : [0, r) \to \mathbb{R}$ *dada por* $\varphi(s) = \gamma\ s^{\frac{1}{p}}$. *Desde que*

$$\varphi'(s) = \frac{1}{p}\ \gamma\ s^{(\frac{1}{p}-1)}$$

temos que

$$s\ \varphi'(s) = \frac{1}{p}\ \gamma\ s^{\frac{1}{p}} = \frac{1}{p}\ \varphi(s),\ \forall s \in (0, r).$$

Fazendo $c = \frac{1}{p}$, *obtemos que*

$$s\ \varphi'(s) \geqslant c\ \varphi(s),\ \forall s \in (0, r),$$

isto é, φ *tem o comportamento moderado próximo à origem. Além disso, se* $\gamma > 0$ *e* $p \geqslant 1$ *conseguimos mostrar que* φ *é concava com* $c = \frac{1}{p} \leqslant 1$.

Nosso próximo resultado fornece uma estimativa para o comprimento das trajetórias subgradientes, quando f satisfaz a desigualdade KL. A demonstração desse teorema foi feita originalmente em espaço de Hilbert e pode ser encontrada em [7]. Relembramos aqui que $S = \arg\min f$ e que $\min f = 0$.

Teorema 12. ***(KL e limites uniformes de curvas subgradientes)***
Sejam $\overline{x} \in S$, $\rho > 0$ *e* $\varphi \in \mathcal{K}(0, r_0)$. *São equivalentes:*

(i) Para cada $y \in B(\overline{x}, \rho) \cap [0 < f < r_0]$, *temos*

$$\varphi'(f(y))\|\partial^0 f(y)\| \geqslant 1.$$

(ii) Para cada $x \in B(\overline{x}, \rho) \cap [0 < f \leqslant r_0]$ *e* $0 \leqslant t < s$, *temos*

$$\text{comp}(\chi_x, t, s) \leqslant \varphi(f(\chi_x(t))) - \varphi(f(\chi_x(s))).$$

Além disso, sob essas condições, $\chi_x(t)$ *converge para um minimizador quando* $t \to \infty$.

Demonstração. Mostraremos primeiramente que $(i) \Rightarrow (ii)$.
Para isto, tome $x \in B(\overline{x}, \rho) \cap [0 < f \leqslant r_0]$ e $0 \leqslant t < s$. Observe primeiro que

$$\begin{aligned}
\varphi(f(\chi_x(t))) - \varphi(f(\chi_x(s))) &= \int_s^t \frac{d}{d\tau}\varphi(f(\chi_x(\tau)))d\tau \\
&= \int_s^t \varphi'(f(\chi_x(\tau)))(-\|\dot{\chi}_x(\tau)\|^2)d\tau \\
&= \int_t^s \varphi'(f(\chi_x(\tau)))\|\dot{\chi}_x(\tau)\|^2 d\tau.
\end{aligned}$$

Desde que $\chi_x(\tau) \in \operatorname{dom} \partial f \cap B(\bar{x}, \rho) \cap [0 < f < r_0]$ para todo $\tau > 0$ e $-\dot{\chi}_x(\tau) \in \partial f(\chi_x(\tau))$ para quase todo $\tau > 0$ (ver Teorema 9), temos que

$$1 \leqslant \|\partial^0(\varphi \circ f)(\chi_x(\tau))\| \leqslant \varphi'(f(\chi_x(\tau)))\|\dot{\chi}_x(\tau)\|$$

para cada τ, onde a primeira desigualdade vem da hipótese (i) e a segunda vem do fato de que $\|\partial^0(\varphi \circ f)(\chi_x(\tau))\| \leqslant \|\partial(\varphi \circ f)(\chi_x(\tau))\|$. Multiplicando por $\|\dot{\chi}_x(\tau)\|$ e integrando de t a s, obtemos

$$\operatorname{comp}(\chi_x, t, s) = \int_t^s \|\dot{\chi}_x(\tau)\| d\tau \leqslant \varphi(f(\chi_x(t))) - \varphi(f(\chi_x(s))).$$

Mostraremos agora que $(ii) \Rightarrow (i)$.

Tome $y \in \operatorname{dom} \partial f \cap B(\bar{x}, \rho) \cap [0 < f < 0]$. Assim, para cada $h > 0$ é válida a desigualdade

$$\begin{aligned} \frac{1}{h}\int_0^h \|\dot{\chi}(\tau)\| d\tau &\leqslant -\frac{\varphi(f(\chi_y(h))) - \varphi(f(\chi_y(0)))}{h} \\ &= -\frac{\varphi(f(\chi_y(h))) - \varphi(f(y))}{h}. \end{aligned} \tag{2.2}$$

Relembre do Teorema do Valor Médio para integrais que, dado $h > 0$, existe $0 \leqslant c \leqslant h$ tal que

$$\int_0^h \|\dot{\chi}_y(\tau)\| d\tau = \|\dot{\chi}_y(c)\|(h - 0) = \|\dot{\chi}_y(c)\|h.$$

Assim, fazendo $h \to 0$ na desigualdade (2.2), obtemos

$$\begin{aligned} \|\dot{\chi}_y(0^+)\| &\leqslant -\varphi'(f(y))(-\|\dot{\chi}_y(0^+)\|^2) \\ &= \varphi'(f(y))\|\partial^0 f(y)\|\|\dot{\chi}_y(0^+)\|, \end{aligned}$$

donde segue que

$$\|\partial^0(\varphi \circ f)(y)\| \geqslant 1.$$

Finalmente, desde que $\|\chi_x(t) - \chi_x(s)\| \leqslant \operatorname{comp}(\chi_x, t, s)$, deduzimos de (ii) que a função $t \mapsto \chi_x(t)$ tem a propriedade de Cauchy quando $t \to \infty$. □

O próximo teorema é o principal resultado desse capítulo. Ele diz que se φ tem um comportamento moderado, f tem a propriedade KL global se, somente se, f tem um error bounds global. Além disso, a função desingularizante na desigualdade KL e a função residual no error bounds são essencialmente as mesmas, a menos de uma constante multiplicativa. Vale observar aqui que esse resultado encontra-se feito na referência [7] e foi feito originalmente em espaço de Hilbert.

Teorema 13. ***(Caracterização de desigualdade de Łojasiewicz para funções convexas)***

Seja $f : \mathbb{R}^n \to \mathbb{R}$ *uma função própria, convexa e semicontínua inferiormente. Seja* $r_0 > 0$, $\varphi \in \mathcal{K}(0, r_0)$, $c > 0$, $\rho > 0$ e $\overline{x} \in \arg\min f$.

(i) (Desigualdade KL implica error bounds) Se $\varphi'(f(x))\|\partial^0 f(x)\| \geqslant 1$ *para todo* $x \in [0 < f < r_0] \cap B(\overline{x}, \rho)$, *então* $\operatorname{dist}(x, S) \leqslant \varphi(f(x))$ *para todo* $x \in [0 < f < r_0] \cap B(\overline{x}, \rho)$.

(ii) (Error bounds implica desigualdade KL) Reciprocamente, se $s\varphi'(s) \geqslant c\varphi(s)$ *para todo* $s \in (0, r_0)$ *(*φ *tem um comportamento moderado), e* $\varphi(f(x)) \geqslant \operatorname{dist}(x, S)$ *para todo* $x \in [0 < f < r_0] \cap B(\overline{x}, \rho)$, *então* $\varphi'(f(x))\|\partial^0 f(x)\| \geqslant c$ *para todo* $x \in [0 < f < r_0] \cap B(\overline{x}, \rho)$.

Demonstração. (i) Relembramos que a aplicação $[0, +\infty) \times \overline{\operatorname{dom} f} \ni (t, x) \to \chi_x(t)$ denota o semifluxo associado a $-\partial f$. Desde que f satisfaz a desigualdade de Kurdyka-Łojasiewicz, podemos aplicar o Teorema 12 para obter

$$\|\chi_x(t) - \chi_x(s)\| \leqslant \operatorname{comp}(\chi_x, t, s) \leqslant \varphi(f(\chi_x(t))) - \varphi(f(\chi_x(s))),$$

para cada $x \in B(\overline{x}, \rho) \cap [0 < f < r_0]$ e $0 \leqslant t < s$. Como estabelecido no Teorema 12, $\chi_x(s)$ converge para $\overline{x} \in S$ quando $s \to \infty$. Tomando então $t = 0$ e fazendo $s \to \infty$ deduzimos que

$$\|\chi_x(0) - \overline{x}\| \leqslant \varphi(f(\chi_x(0))) - \varphi(f(\overline{x})) \;\Rightarrow\; \|x - \overline{x}\| \leqslant \varphi(f(x)).$$

Portanto $\varphi(f(x)) \geqslant \operatorname{dist}(x, S)$ para todo $x \in [0 < f < r_0] \cap B(\overline{x}, \rho)$.

(ii) Tome $x \in [0 < f < r_0] \cap B(x, \rho)$ e escreva $y = P_S(x)$. Por convexidade temos

$$0 = f(y) \geqslant f(x) + \langle \partial^0 f(x), y - x \rangle,$$

onde a igualdade ocorre devido a $y = P_S(x) \in S$ e a desigualdade ocorre porque $\partial^0 f(x)$ é um subdiferencial de f em x. Isto implica que

$$f(x) \leqslant \|\partial^0 f(x)\|\|y - x\| = \operatorname{dist}(x, S)\|\partial^0 f(x)\| \leqslant \varphi(f(x))\|\partial^0 f(x)\|.$$

Como $x \in [0 < f < r_0] \cap B(\overline{x}, \rho)$ segue que $f(x) > 0$. Além disso, $s\varphi'(s) \geqslant c\varphi(s)$ para todo $s \in (0, r_0)$. Daí,

$$1 \leqslant \|\partial^0 f(x)\| \frac{\varphi(f(x))}{f(x)} \leqslant \frac{1}{c}\|\partial^0 f(x)\|\varphi'(f(x)).$$

Assim,

$$\varphi'(f(x))\|\partial^0 f(x)\| \geqslant c,$$

para todo $x \in [0 < f < r_0] \cap B(\overline{x}, \rho)$. □

Em uma demonstração similar ao do Teorema 13, podemos caracterizar a existência global de uma desigualdade gradiente Łojasiewicz.

Corolário 5. ***(Caracterização da desigualdade de Łojasiewicz para funções convexas: caso global)***

Seja $f : \mathbb{R}^n \to \mathbb{R}$ *uma função própria, convexa e semicontínua inferiormente. Seja* $\varphi \in \mathcal{K}(0, +\infty)$ *e* $c > 0$.

(i) Se $\varphi'(f(x))\|\partial^0 f(x)\| \geqslant 1$ *para todo* $x \in [0 < f]$, *então* $\operatorname{dist}(x, S) \leqslant \varphi(f(x))$ *para todo* $x \in [0 < f]$.

(ii) Reciprocamente, se $s\varphi'(s) \geqslant c\varphi(s)$ *para todo* $s \in (0, r_0)$ *(*φ *tem um comportamento moderado), e* $\varphi(f(x)) \geqslant \operatorname{dist}(x, S)$ *para todo* $x \in [0 < f]$, *então* $\varphi'(f(x))\|\partial^0 f(x)\| \geqslant c$ *para todo* $x \in [0 < f]$.

Demonstração. Segue do teorema. □

Observação 4. *Observe uma ligeira dissimetria entre as conclusões de* (i) *e* (ii) *do teorema e o corolário: enquanto uma função desingularizante provê diretamente um error bounds in* (i), *um error bounds (com crescimento moderado) vem de uma função desingularizante depois de um redimensionamento, dado por* $c^{-1}\varphi$.

Capítulo 3

Complexidade para métodos de descida de primeira ordem

No que se segue, assumiremos, assim como antes, que $f : \mathbb{R}^n \to \mathbb{R}$ é uma função própria, convexa, semicontínua inferiormente tal que $S = \arg\min f \neq \emptyset$ e $\min f = 0$. Os conceitos e resultados apresentados podem ser consultados em [7].

3.1 Métodos de descida de primeira ordem

Uma sequência $(x_k)_{k\in\mathbb{N}} \subset \mathbb{R}^n$ é um *método de descida de primeira ordem* ou *algoritmo abstrato* de $f : \mathbb{R}^n \to \mathbb{R}$ se $x_0 \in \text{dom } f$ e existem $a, b > 0$ tais que:

(H1) (Condição suficiente de decrescimento) Para cada $k \geqslant 1$,

$$f(x_k) + a\|x_k - x_{k-1}\|^2 \leqslant f(x_{k-1}).$$

(H2) (Condição de erro relativo) Para cada $k \geqslant 1$, existe $\omega_k \in \partial f(x_k)$ tal que

$$\|\omega_k\| \leqslant b\|x_k - x_{k-1}\|.$$

Observamos que o conceito de algoritmo abstrato foi usado em [2] com uma condição adicional de continuidade. Aqui não será necessário essa condição, visto que f é convexa. Essas condições foram consideradas inicialmente no trabalho de Luo-Tseng [24] e foram usadas para o estudo de raios de convergência para error bounds.

Observação 5. *(Passo explícito para gradientes Lipschitz contínuos)*
Se f *é diferenciável e seu gradiente é L-Lipschitz contínuo, então qualquer sequência satisfazendo:*

(H2') Para cada $k \geqslant 1$,

$$\|\nabla f(x_{k-1})\| \leqslant b\|x_k - x_{k-1}\|,$$

também satisfaz (H2).
De fato, para todo $k \geqslant 1$,

$$\begin{aligned}
\|\nabla f(x_k)\| &= \|\nabla f(x_k) + \nabla f(x_{k-1}) - \nabla f(x_{k-1})\| \\
&\leqslant \|\nabla f(x_{k-1})\| + \|\nabla f(x_k) - \nabla f(x_{k-1})\| \\
&\leqslant b\|x_k - x_{k-1}\| + L\|x_k - x_{k-1}\| \\
&= (b+L)\|x_k - x_{k-1}\|.
\end{aligned}$$

Exemplo 6. *(O método do ponto proximal)*
Seja $f : \mathbb{R}^n \to \mathbb{R}$ *uma função própria, convexa, semicontínua inferiormente e assuma que* $S \neq \emptyset$ *e* $\min f = 0$. *O método do ponto proximal para resolver o problema de otimização*

$$\min_{x \in \mathbb{R}^n} f(x)$$

é definido formalmente do seguinte modo: dados a sequência $(\lambda_k)_{k\in\mathbb{N}}$ *de números reais positivos e* $x_0 \in \mathbb{R}^n$, *defina*

$$x_{k+1} = \arg\min\left\{ f(z) + \frac{\lambda_k}{2}\|z - x_k\|^2 : z \in \mathbb{R}^n \right\}, \quad k = 0, 1, \dots \tag{3.1}$$

Se $(\lambda_k)_{k\in\mathbb{N}}$ *é uma sequência tal que* $0 < \lambda^- \leqslant \lambda_k \leqslant \lambda^+ < +\infty$ *para todo* $k \in \mathbb{N}$ *então*

$$f(x_{k+1}) + \frac{\lambda^-}{2}\|x_{k+1} - x_k\|^2 \leqslant f(x_{k+1}) + \frac{\lambda_k}{2}\|x_{k+1} - x_k\|^2 \leqslant f(x_k), \ \forall k \in \mathbb{N}.$$

Além disso, como $0 = \partial f(x_{k+1}) + \lambda_k(x_{k+1} - x_k)$, *temos que*

$$\begin{aligned}
\partial f(x_{k+1}) = \lambda_k(x_k - x_{k+1}) &\Rightarrow \|\partial f(x_{k+1})\| = \|\lambda_k(x_{k+1} - x_k)\| = \lambda_k\|x_{k+1} - x_k\| \\
&\Rightarrow \|\partial f(x_{k+1})\| \leqslant \lambda^+\|x_{k+1} - x_k\|, \ \forall k \in \mathbb{N}.
\end{aligned}$$

Assim, se tomamos $a = \frac{\lambda^-}{2}$ *e* $b = \lambda^+$, *temos que o método do ponto proximal, com a sequência* $(\lambda_k)_{k\in\mathbb{N}}$ *definida acima, é um exemplo de método de descida de primeira ordem.*

Nosso próximo lema será usado com frequência daqui em diante. Esse resultado pode ser encontrado em [27].

Lema 2. ***(Lema de descida)***

Seja $h : \mathbb{R}^n \to \mathbb{R}$ *uma função continuamente diferenciável na qual o gradiente é L-Lipschitz contínuo. Então*

$$h(u) \leqslant h(v) + \langle \nabla h(v), u - v \rangle + \frac{L}{2}\|u - v\|^2, \ \forall u, v \in \mathbb{R}^n. \tag{3.2}$$

Demonstração. Escreva $\theta = u - v$ e defina $f : [0,1] \to \mathbb{R}$ por $f(t) = h(v + t\theta)$. Então

$$f'(t) = \langle \nabla h(v + t\theta), \theta \rangle$$

para cada $t \in (0,1)$ e, assim,

$$\int_0^1 \langle \nabla h(v + t\theta), \theta \rangle dt = \int_0^1 f'(t)dt = f(1) - f(0) = h(u) - h(v).$$

Segue daí que

$$\begin{aligned}
h(u) - h(v) &= \int_0^1 \langle \nabla h(v + t\theta) + \nabla h(v) - \nabla h(v), \theta \rangle dt \\
&= \int_0^1 \langle \nabla h(v), \theta \rangle dt + \int_0^1 \langle \nabla h(v + t\theta) - \nabla h(v), \theta \rangle dt \\
&\leqslant \langle \nabla h(v), \theta \rangle + \int_0^1 \|\nabla h(v + t\theta) - \nabla h(v)\| \|\theta\| dt \\
&\leqslant \langle \nabla h(v), \theta \rangle + \|\theta\| \int_0^1 L\|(v + t\theta) - v\| dt \\
&= \langle \nabla h(v), \theta \rangle + L\ \|\theta\|^2 \int_0^1 t\ dt \\
&= \langle \nabla h(v), u - v \rangle + \frac{L}{2}\|u - v\|^2
\end{aligned}$$

e então

$$h(u) \leqslant h(v) + \langle \nabla h(v), u - v \rangle + \frac{L}{2}\|u - v\|^2, \ \forall u, v \in \mathbb{R}^n.$$

□

Exemplo 7. *(O método do gradiente)*

Seja $f : \mathbb{R}^n \to \mathbb{R}$ *uma função convexa. Uma das estratégias mais usadas para resolver o problema irrestrito*

$$\min_{x \in \mathbb{R}^n} f(x)$$

é a seguinte: dada uma aproximação $x_k \in \mathbb{R}^n$ *do problema, encontramos um ponto* $x_{k+1} \in \mathbb{R}^n$ *tal que* $f(x_{k+1}) < f(x_k)$. *Obviamente, isto pode ser feito de diversas maneiras. Uma destas maneiras é tomar uma direção* $d_k \in \mathbb{R}^n$ *tal que* f *é decrescente a partir do ponto* x_k *nessa direção, e computar um comprimento de passo* $\lambda_k > 0$ *que fornece um valor de* f *menor do que no ponto* x_k,

$$f(x_k + \lambda_k d_k) < f(x_k).$$

Obtemos assim o seguinte iterando:

$$x_{k+1} = x_k + \lambda_k d_k.$$

Repetimos o processo para um novo ponto x_{k+1}, *etc. Métodos desse tipo são chamados métodos de descida. Um método de descida bastante conhecido na literatura é o método do gradiente. Neste método precisamos assumir ainda que* f *seja diferenciável no* $\mathbb{R}^n$ *e escolhemos a direção de descida* d_k *como sendo o anti-gradiente da função* f *no ponto* x_k, *isto é,* $d_k = -\nabla f(x_k)$. *Se* $\nabla f(x_k) = 0$ *para algum* k, *então o método pára. Em particular, o esquema acima se reduz ao seguinte:*

$$x_{k+1} = x_k - \lambda_k \nabla f(x_k), \; k = 0, 1, \ldots \tag{3.3}$$

Apresentaremos agora alguns resultados que mostram que o método gradiente é um método de descida de primeira ordem.

Lema 3. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função convexa diferenciável. Considere a sequência* $(x_k)_{k\in\mathbb{N}}$ *gerada pela recursão (3.3) e suponha que existem* $\lambda^-, \lambda^+ \in \mathbb{R}$ *tais que*

$$0 < \lambda^- \leqslant \lambda_k \leqslant \lambda^+ < +\infty, \; \forall k \in \mathbb{N}.$$

Então

(i) $\langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{1}{\lambda^+}\|x_{k+1} - x_k\|^2 \leqslant 0, \; \forall k \in \mathbb{N}$;

(ii) $\|\nabla f(x_k)\| \leqslant \frac{1}{\lambda^-}\|x_{k+1} - x_k\|, \; \forall k \in \mathbb{N}$.

Demonstração. Primeiro mostraremos que vale (i). Para isto, é suficiente notar que

$$\langle \nabla f(x_k), x_{k+1} - x_k \rangle = \left\langle -\frac{x_{k+1} - x_k}{\lambda_k}, x_{k+1} - x_k \right\rangle = -\frac{1}{\lambda_k}\|x_{k+1} - x_k\|^2.$$

Assim,

$$0 = \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{1}{\lambda_k}\|x_{k+1} - x_k\|^2 \geqslant \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{1}{\lambda^+}\|x_{k+1} - x_k\|^2,$$

donde segue que

$$\langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{1}{\lambda^+}\|x_{k+1} - x_k\|^2 \leqslant 0, \ \forall k \in \mathbb{N}.$$

Para verificar (ii), basta notarmos que

$$\begin{aligned}
\|\nabla f(x_k)\|^2 = \langle \nabla f(x_k), \nabla f(x_k) \rangle &= \left\langle \nabla f(x_k), -\frac{x_{k+1} - x_k}{\lambda_k} \right\rangle \\
&\leqslant \|\nabla f(x_k)\| \left\| \frac{x_{k+1} - x_k}{\lambda_k} \right\| \\
&= \frac{1}{\lambda_k}\|\nabla f(x_k)\|\|x_{k+1} - x_k\| \\
&\leqslant \frac{1}{\lambda^-}\|\nabla f(x_k)\|\|x_{k+1} - x_k\|, \ \forall k \in \mathbb{N}.
\end{aligned}$$

Segue daí que

$$\|\nabla f(x_k)\| \leqslant \frac{1}{\lambda^-}\|x_{k+1} - x_k\|, \ \forall k \in \mathbb{N}.$$

□

Proposição 6. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa diferenciável. Suponha ainda que ∇f é L-Lipschitz contínua. Considere a sequência $(x_k)_{k \in \mathbb{N}}$ gerada pela recursão (3.3) e suponha que existem $\lambda^-, \lambda^+ \in \mathbb{R}$ tais que*

$$0 < \lambda^- \leqslant \lambda_k \leqslant \lambda^+ < +\infty, \ \forall k \in \mathbb{N}.$$

Então

(a) $f(x_{k+1}) + \left(\frac{1}{\lambda^+} - \frac{L}{2}\right)\|x_{k+1} - x_k\|^2 \leqslant f(x_k), \ \forall k \in \mathbb{N}$;

(b) $\|\nabla f(x_{k+1})\| \leqslant \left(L + \frac{1}{\lambda^-}\right)\|x_{k+1} - x_k\|, \ \forall k \in \mathbb{N}$.

Demonstração. Aplicando o Lema 2 aos pontos $u = x_{k+1}$ e $v = x_k$ e depois usando a desigualdade (i) do Lema 3 obtemos

$$\begin{aligned}
f(x_{k+1}) - f(x_k) &\leqslant \langle \nabla f(x_{k+1}), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\
&\leqslant -\frac{1}{\lambda^+}\|x_{k+1} - x_k\|^2 + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\
&= \left(\frac{L}{2} - \frac{1}{\lambda^+}\right)\|x_{k+1} - x_k\|^2, \ \forall k \in \mathbb{N},
\end{aligned}$$

e assim,

$$f(x_{k+1}) + \left(\frac{1}{\lambda^+} - \frac{L}{2}\right) \|x_{k+1} - x_k\|^2 \leqslant f(x_k), \ \forall k \in \mathbb{N}.$$

Agora, usando a desigualdade (ii) do Lema 3 (e relembrando a Observação 5), temos que

$$\begin{aligned}
\|\nabla f(x_{k+1})\| &= \|\nabla f(x_{k+1}) - \nabla f(x_k) + \nabla f(x_k)\| \\
&\leqslant \|\nabla f(x_{k+1}) - \nabla f(x_k)\| + \|\nabla f(x_k)\| \\
&\leqslant L\|x_{k+1} - x_k\| + \frac{1}{\lambda^-}\|x_{k+1} - x_k\| \\
&= \left(L + \frac{1}{\lambda^-}\right)\|x_{k+1} - x_k\|, \ \forall k \in \mathbb{N}.
\end{aligned}$$

Assim,

$$\|\nabla f(x_{k+1})\| \leqslant \left(L + \frac{1}{\lambda^-}\right) \|x_{k+1} - x_k\|, \ \forall k \in \mathbb{N}.$$

□

Observe que se $\lambda^+ < 2/L$ então $\frac{1}{\lambda^+} > \frac{L}{2}$ e assim teríamos $\frac{1}{\lambda^+} - \frac{L}{2} > 0$. Daí a proposição acima nos garante que o método gradiente é um método de descida de primeira ordem se $\lambda^+ < 2/L$.

Nosso próximo exemplo generaliza os dois últimos.

Exemplo 8. *(O método forward-backward)*
Sejam $g : \mathbb{R}^n \to \mathbb{R}$ *uma função própria, convexa e semicontínua inferiormente e* $h : \mathbb{R}^n \to \mathbb{R}$ *uma função convexa diferenciável na qual o gradiente é L-Lipschitz contínuo. No intuito de minimizar* $f = g + h$ *sobre* $\mathbb{R}^n$*, o método forward-backward gera uma sequência* $(x_k)_{k\in\mathbb{N}}$ *de um dado ponto* $x_0 \in \mathbb{R}^n$*, e usando a recursão*

$$x_{k+1} \in \arg\min \left\{ g(z) + \langle \nabla h(x_k), z - x_k \rangle + \frac{1}{2\lambda_k}\|z - x_k\|^2 : z \in \mathbb{R}^n \right\} \tag{3.4}$$

para $k \geqslant 1$.

Usando a convexidade de g *e combinando (3.4) com a Proposiçao 1 obtemos:*

$$0 \in \partial g(x_{k+1}) + \nabla h(x_k) + \frac{1}{\lambda_k}(x_{k+1} - x_k) = \partial g(x_{k+1}) + \frac{1}{\lambda_k}(\lambda_k \nabla h(x_k) + x_{k+1} - x_k)$$

Assim, se tomamos

$$F(z) = \frac{1}{2\lambda_k}\|z - (x_k - \lambda_k \nabla h(x_k))\|^2,$$

obtemos

$$0 \in \partial g(x_{k+1}) + \nabla F(x_{k+1}),$$

donde segue que

$$x_{k+1} \in \arg\min\Big\{g(z) + \frac{1}{2\lambda_k}\|z - (x_k - \lambda_k \nabla h(x_k))\|^2 : z \in \mathbb{R}^n\Big\} \tag{3.5}$$

Além disso, da definição de operador proximal, podemos reescrever (3.5) como

$$x_{k+1} = \operatorname{prox}_{\lambda_k g}(x_k - \lambda_k \nabla h(x_k)). \tag{3.6}$$

Note que quando $h \equiv 0$*, obtemos o método do ponto proximal para* g*. Por outro lado, se* $g \equiv 0$ *então temos método do gradiente para* h*.*

Lema 4. *Seja* $h : \mathbb{R}^n \to \mathbb{R}$ *uma função continuamente diferenciável na qual o gradiente é L-Lipschitz contínuo e seja* $g : \mathbb{R}^n \to \mathbb{R}$ *uma função própria, semicontínua inferiormente e com* $\inf_{\mathbb{R}^n} g > -\infty$*. Fixe qualquer* $\frac{1}{\lambda} > L$*. Então, para qualquer* $u \in \operatorname{dom} g$ *e qualquer* $v \in \mathbb{R}^n$ *definido por*

$$v \in \operatorname{prox}_{\lambda g}\Big(u - \lambda\ \nabla h(u)\Big) \tag{3.7}$$

temos

$$h(v) + g(v) \leqslant h(u) + g(u) - \frac{1}{2}\Big(\frac{1}{\lambda} - L\Big)\|v - u\|^2. \tag{3.8}$$

Demonstração. Pelo Exemplo 8, a inclusão (3.7) implica em

$$g(v) + \langle \nabla h(u), v - u\rangle + \frac{1}{2\lambda}\|v - u\|^2 \leqslant g(z) + \langle \nabla h(u), z - u\rangle + \frac{1}{2\lambda}\|z - u\|^2$$

para todo $z \in \mathbb{R}^n$. Em particular, para $z = u \in \mathbb{R}^n$, temos

$$g(v) + \langle \nabla h(u), v - u\rangle + \frac{1}{2\lambda}\|v - u\|^2 \leqslant g(u)$$

e assim,

$$g(v) \leqslant g(u) - \langle \nabla h(u), v - u\rangle - \frac{1}{2\lambda}\|v - u\|^2 \tag{3.9}$$

Agora, combinando (3.9) com o Lema 2, obtemos

$$\begin{aligned} h(v) + g(v) &\leqslant h(u) + \langle \nabla h(u), v - u\rangle + \frac{L}{2}\|v - u\|^2 + g(u) - \langle \nabla h(u), v - u\rangle - \frac{1}{2\lambda}\|v - u\|^2 \\ &= h(u) + g(u) - \left(\frac{1}{2\lambda} - \frac{L}{2}\right)\|v - u\|^2 \end{aligned}$$

Segue então o resultado. □

Observação 6. *No caso em que* g *é uma função própria, convexa e semicontínua inferiormente, podemos tomar* $\frac{1}{\lambda} = L$ *(e até mesmo* $\frac{1}{\lambda} > \frac{L}{2}$*) para obter*

$$h(v) + g(v) \leqslant h(u) + g(u) - \left(\frac{1}{\lambda} - \frac{L}{2}\right) \|v - u\|^2. \tag{3.10}$$

Na proposição a seguir, veremos que o método forward-backward gera um método de descida de primeira ordem se o tamanho dos passos forem devidamente escolhidos. Tal demonstração encontra-se feita em [7].

Proposição 7. *Suponha que* $0 < \lambda^- \leqslant \lambda_k \leqslant \lambda^+ < 2/L$ *para todo* $k \in \mathbb{N}$*. Então (H1) e (H2) são satisfeitos para o método forward-backward com*

$$a = \frac{1}{\lambda^+} - \frac{L}{2} \text{ e } b = \frac{1}{\lambda^-} + L.$$

Demonstração. Primeiramente note que

$$\lambda_k < \frac{2}{L} \;\Rightarrow\; \frac{1}{\lambda_k} > \frac{L}{2}.$$

Agora, tomando $k \geqslant 0$ e usando a Observação 6, temos

$$\begin{aligned} g(x_{k+1}) + h(x_{k+1}) &\leqslant g(x_k) + h(x_k) - \left(\frac{1}{\lambda_k} - \frac{L}{2}\right) \|x_{k+1} - x_k\|^2 \\ &\leqslant g(x_k) + h(x_k) - \left(\frac{1}{\lambda^+} - \frac{L}{2}\right) \|x_{k+1} - x_k\|^2. \end{aligned}$$

Assim, para $f = g + h$, temos que

$$f(x_{k+1}) + \left(\frac{1}{\lambda^+} - \frac{L}{2}\right) \|x_{k+1} - x_k\|^2 \leqslant f(x_k),$$

onde, tomando $a = \frac{1}{\lambda^+} - \frac{L}{2} > 0$ temos que (H1) é satisfeita.

Para a constante b, note que a inclusão (3.4) nos indica que

$$0 \in \partial g(x_{k+1}) + \nabla h(x_k) + \frac{1}{\lambda_k}(x_{k+1} - x_k).$$

Daí, existe $\omega_{k+1} \in \partial g(x_{k+1})$ tal que

$$0 = \omega_{k+1} + \nabla h(x_k) + \frac{1}{\lambda_k}(x_{k+1} - x_k) \;\Rightarrow\; \omega_{k+1} + \nabla h(x_k) = \frac{1}{\lambda_k}(x_k - x_{k+1}).$$

Usando agora que ∇h é L-Lipschitz contínua, obtemos

$$\begin{aligned} \|\omega_{k+1} + \nabla h(x_{k+1})\| &= \|\omega_{k+1} + \nabla h(x_k) - \nabla h(x_k) + \nabla h(x_{k+1})\| \\ &\leqslant \frac{1}{\lambda_k}\|x_k - x_{k+1}\| + L\|x_{k+1} - x_k\| \\ &= \left(\frac{1}{\lambda_k} + L\right) \|x_{k+1} - x_k\| \\ &\leqslant \left(\frac{1}{\lambda^-} + L\right) \|x_{k+1} - x_k\|. \end{aligned}$$

Logo, se $f = g + h$, temos que (H2) é satisfeita com $b = \frac{1}{\lambda^-} + L$. □

Se $f = g + h$ tem a propriedade KL, o próximo teorema garante a convergência da sequência gerada pelo método forward-backward. Sua demonstração encontra-se feita em detalhes em [7].

Teorema 14. ***(Convergência do algoritmo de descida)***
Suponha que $f : \mathbb{R}^n \to \mathbb{R}$ *é uma função própria, convexa, semicontínua inferiormente e que tem a propriedade KL em* $[0 < f < \bar{r}]$ *com função desingularizante* $\varphi \in \mathcal{K}(0, \bar{r})$. *Considere o algoritmo abstrato* $(x_k)_{k\in\mathbb{N}}$ *tal que* $f(x_0) \leqslant r_0 < \bar{r}$. *Então* x_k *converge para algum* $x^* \in \arg\min f$ *e*

$$\|x_k - x^*\| \leqslant \frac{b}{a}\varphi(f(x_k)) + \sqrt{\frac{f(x_{k-1})}{a}}, \ \forall k \geqslant 1.$$

Demonstração. Usando (H1), deduzimos que a sequência $(f(x_k))_{k\in\mathbb{N}}$ é não-crescente. Então $0 \leqslant f(x_k) < \bar{r}$, implicando que $x_k \in [0 \leqslant f < \bar{r}]$. Denotamos por i_0 o primeiro índice $i_0 \geqslant 1$ tal que $\|x_{i_0} - x_{i_0-1}\| = 0$ sempre que ele existir. Se i_0 existe, então

$$0 \leqslant \|\partial f(x_{i_0})\| \leqslant b\|x_{i_0} - x_{i_0-1}\| = 0 \ \Rightarrow \ \partial f(x_{i_0}) = 0.$$

Daí, x_{i_0} é minimizador de f e, como estamos assumindo $\min f = 0$, temos que $f(x_{i_0}) = 0$. Como f é não-crescente, isto implica ainda que $f(x_{i_0+1}) = 0$ e portanto $x_{i_0+1} = x_{i_0}$. Logo a sequência é estacionária. Daí o limite foi estabelecido para todo k tal que $i_0 \geqslant k + 1$, isto é, para todo $k \leqslant i_0 - 1$. Um argumento similar se aplica ao caso em que $f(x_{i_0}) = 0$. Suponha agora que $f(x_k) > 0$ e $\|x_k - x_{k-1}\| > 0$ para todo $k \geqslant 1$.
Lembramos que se f tem a propriedade KL então

$$\varphi'(f(x_k))\|\partial^0 f(x_k)\| \geqslant 1 \ \Rightarrow \ \varphi'(f(x_k)) \geqslant \frac{1}{\|\partial^0 f(x_k)\|}.$$

Lembramos ainda que se φ é côncava, vale a seguinte desigualdade:

$$-\varphi(f(x_{k+1})) \ \geqslant \ -\varphi(f(x_k)) - \varphi'(f(x_k))(f(x_{k+1}) - f(x_k))$$

Agora, combinando (H1), (H2) e usando a concavidade de φ obtemos

$$\begin{aligned}
\varphi(f(x_k)) - \varphi(f(x_{k+1})) &\geqslant \varphi'(f(x_k))(f(x_k) - f(x_{k+1})) \\
&\geqslant \frac{(f(x_k) - f(x_{k+1}))}{\|\partial^0 f(x_k)\|} \\
&\geqslant \frac{a\|x_k - x_{k+1}\|^2}{b\|x_k - x_{k-1}\|} \\
&\geqslant \frac{a}{b}\frac{(2\|x_k - x_{k+1}\|\|x_k - x_{k-1}\| - \|x_{k-1} - x_k\|^2)}{\|x_k - x_{k-1}\|} \\
&= \frac{a}{b}(2\|x_k - x_{k+1}\| - \|x_{k-1} - x_k\|), \ \forall k \geqslant 1. \qquad (3.11)
\end{aligned}$$

Assim sendo,

$$\begin{aligned}
\varphi(f(x_1)) - \varphi(f(x_{k+1})) &= \varphi(f(x_1)) - \varphi(f(x_2)) + \varphi(f(x_2)) - \varphi(f(x_3)) + \varphi(f(x_3)) + \dots \\
&\quad \dots - \varphi(f(x_k)) + \varphi(f(x_k)) - \varphi(f(x_{k+1})) \\
&\geqslant \frac{a}{b}\Big[(2\|x_1 - x_2\| - \|x_0 - x_1\|) + (2\|x_2 + x_3\| - \|x_1 - x_2\|) + \dots \\
&\quad \dots + (2\|x_k - x_{k+1}\| - \|x_{k-1} - x_k\|)\Big] \\
&= \frac{a}{b}\Big[\|x_1 - x_2\| - \|x_0 - x_1\| + \|x_2 + x_3\| + \dots + \|x_k - x_{k+1}\|\Big]
\end{aligned}$$

implicando que

$$\frac{b}{a}(\varphi(f(x_1)) - \varphi(f(x_{k+1}))) + \|x_0 - x_1\| \geqslant \sum_{i=1}^{k} \|x_i - x_{i+1}\|, \ \forall k \in \mathbb{N}. \tag{3.12}$$

Como $\varphi > 0$ segue que

$$\frac{b}{a}(\varphi(f(x_1))) + \|x_0 - x_1\| \geqslant \sum_{i=1}^{k} \|x_i - x_{i+1}\|, \ \forall k \in \mathbb{N}.$$

Portanto, a série $\sum_{i=1}^{\infty} \|x_i - x_{i+1}\|$ é convergente, implicando que a sequência $(x_k)_{k\in\mathbb{N}}$ converge para algum ponto $x^* \in \mathbb{R}^n$.

De (H2), existe uma sequência $\omega_k \in \partial f(x_k)$ tal que

$$0 \leqslant \|\omega_k\| \leqslant b\|x_k - x_{k-1}\| \to b\|x^* - x^*\| = 0 \quad (k \to \infty),$$

isto é, existe $\omega_k \in \partial f(x_k)$ tal que $\omega_k \to 0$ $(k \to \infty)$. Desde que f é convexa e semicontínua inferiormente, o gráfico de ∂f é fechado em $\mathbb{R}^n \times \mathbb{R}^n$. Segue daí que,

$$(x_k, \omega_k) \to (x^*, 0) \ \Rightarrow \ 0 \in \partial f(x^*) \ \Rightarrow \ x^* \in \arg\min f,$$

provando então que $(x_k)_{k\in\mathbb{N}}$ converge para um minimizador de f. Voltando à desigualdade (3.11), podemos deduzir também que

$$\frac{b}{a}(\varphi(f(x_k)) - \varphi(f(x_{k+m}))) + \|x_{k-1} - x_k\| \geqslant \sum_{i=1}^{k+m} \|x_i - x_{i+1}\|, \ \forall k, m \in \mathbb{N}. \tag{3.13}$$

Sabendo de (H1) que

$$\|x_k - x_{k-1}\| \leqslant \sqrt{\frac{f(x_{k-1}) - f(x_k)}{a}}, \ \forall k \geqslant 1,$$

podemos escrever ainda a desigualdade (3.13) como sendo

$$\frac{b}{a}(\varphi(f(x_k)) - \varphi(f(x_{k+m}))) + \sqrt{\frac{f(x_{k-1}) - f(x_k)}{a}} \geqslant \sum_{i=1}^{k+m} \|x_i - x_{i+1}\|, \ \forall k, m \in \mathbb{N}.$$

Observamos agora que

$$\begin{aligned} \|x_k - x_{k+m+1}\| &= \|x_k - x_{k+1} + x_{k+1} - x_{k+2} + \ldots - x_{k+m} + x_{k+m} - x_{k+m+1}\| \\ &\leqslant \|x_k - x_{k+1}\| + \|x_{k+1} - x_{k+2}\| + \ldots + \|x_{k+m} - x_{k+m+1}\| \\ &= \sum_{i=1}^{k+m} \|x_i - x_{i+1}\|. \end{aligned}$$

Segue daí que

$$\frac{b}{a}(\varphi(f(x_k)) - \varphi(f(x_{k+m}))) + \sqrt{\frac{f(x_{k-1}) - f(x_k)}{a}} \geqslant \|x_k - x_{k+m+1}\|, \ \forall k, m \in \mathbb{N}.$$

Fazendo $m \to \infty$ na desigualdade acima obtemos

$$\frac{b}{a}(\varphi(f(x_k)) - \varphi(f(x^*))) + \sqrt{\frac{f(x_{k-1}) - f(x_k)}{a}} \geqslant \|x_k - x^*\|, \ \forall k \in \mathbb{N}.$$

Como $f(x_k) \geqslant 0$ para todo $k \in \mathbb{N}$ e $\varphi(f(x^*)) = 0$ ($x^* \in \arg\min f$) obtemos que

$$\frac{b}{a}(\varphi(f(x_k))) + \sqrt{\frac{f(x_{k-1})}{a}} \geqslant \|x_k - x^*\|, \ \forall k \in \mathbb{N}.$$

O caso em que $\|x_k - x_{k-1}\|$ ou $f(x_k)$ é nula para algum k segue facilmente usando o argumento do começo da prova. □

3.2 Complexidade do método de descida de primeira ordem

Esta seção é dedicada ao estudo de complexidade dos métodos de descida de primeira ordem para funções convexas com a propriedade KL. Os resultados obtidos aqui encontram-se em [7].

Seja $0 < r_0 < \bar{r}$. Assumiremos que f tem a propriedade KL em $[0 < f < \bar{r}]$ com função desingularizante $\varphi \in \mathcal{K}(0, \bar{r})$. Daí, lembrando que $\arg\min f \neq \emptyset$ e $\min f = 0$, temos

$$\varphi'(f(x))\|\partial^0 f(x)\| \geqslant 1$$

para todo $x \in [0 < f < \bar{r}]$. Seja $\alpha_0 = \varphi(r_0)$ e considere a função

$$\psi = (\varphi|_{[0,r_0]})^{-1} : [0, \alpha_0] \to [0, r_0] \tag{3.14}$$

a qual é crescente e convexa. A seguinte suposição será útil no que se segue:

(A) A função ψ' é l-Lipschitz contínua (em $[0, \alpha_0]$) e $\psi'(0) = 0$.

Focamos nos métodos gerados pelo método de descida de primeira ordem, ou seja, nos métodos que satisfazem (H1) e (H2). Seja

$$\zeta = \frac{\sqrt{1 + 2lab^{-2}} - 1}{l} > 0 \tag{3.15}$$

onde $a > 0, b > 0$ e $l > 0$ são dados em (H1), (H2) e (A), respectivamente. Começando de α_0, definimos a *sequência proximal unidimensional "pior caso"* indutivamente por

$$\alpha_{k+1} = \arg\min \left\{ \psi(u) + \frac{1}{2\zeta}(u - \alpha_k)^2 : u \geqslant 0 \right\} \tag{3.16}$$

para $k \geqslant 0$. Observamos que α_k está bem definido e é positivo para cada $k \geqslant 0$. Note que a definição de operador proximal nos dá

$$\alpha_{k+1} = \mathtt{prox}_{\zeta\psi}(\alpha_k).$$

Além disso, como $\psi(u) + \frac{1}{2\zeta}(u - \alpha_k)^2$ é coerciva e estritamente convexa, com ψ diferenciável, vale o seguinte:

$$\begin{aligned} 0 = \psi'(\alpha_{k+1}) + \frac{1}{\zeta}(\alpha_{k+1} - \alpha_k) \quad &\Rightarrow \quad 0 = \zeta\, \psi'(\alpha_{k+1}) + (\alpha_{k+1} - \alpha_k) \\ &\Rightarrow \quad \alpha_k = \zeta\, \psi'(\alpha_{k+1}) + \alpha_{k+1} \\ &\Rightarrow \quad \alpha_k = (I + \zeta\, \psi')(\alpha_{k+1}) \\ &\Rightarrow \quad (I + \zeta\, \psi')^{-1}(\alpha_k) = \alpha_{k+1}, \end{aligned}$$

ou seja,

$$\alpha_{k+1} = (I + \zeta\, \psi')^{-1}(\alpha_k) \tag{3.17}$$

para todo $k \geqslant 0$ e onde I é a identidade em $\mathbb{R}$.

Afirmação 1. *A sequência $(\alpha_k)_{k\in\mathbb{N}}$ é decrescente e converge para zero.*

A primeira parte da afirmação segue diretamente da desigualdade

$$\psi(\alpha_{k+1}) \leqslant \psi(\alpha_{k+1}) + \frac{1}{2\zeta}(\alpha_{k+1} - \alpha_k)^2 \leqslant \psi(\alpha_k)$$

e do fato de que ψ é uma função crescente. Daí,

$$\psi(\alpha_{k+1}) \leqslant \psi(\alpha_k) \Rightarrow \alpha_{k+1} \leqslant \alpha_k,\ \forall k \geqslant 0 \Rightarrow (\alpha_k)_{k\in\mathbb{N}} \text{ é decrescente.}$$

Para a segunda parte da afirmação, notemos que $(\alpha_k)_{k\in\mathbb{N}}$ *é decrescente e está contida no compacto* $[0, \alpha_0]$. *Logo* $(\alpha_k)_{k\in\mathbb{N}}$ *converge para algum* $\overline{\alpha} \geqslant 0$. *Suponha que* $\overline{\alpha} > 0$. *Então*

$$0 = \psi'(\alpha_{k+1}) + \frac{1}{\zeta}(\alpha_{k+1} - \alpha_k) \to \psi'(\overline{\alpha}) \text{ quando } k \to \infty.$$

Como estamos supondo (A), a igualdade acima não ocorre. Daí, $\overline{\alpha} = 0$ *e segue então que* $(\alpha_k)_{k\in\mathbb{N}}$ *converge para zero.*

Faremos agora um resultado que será essencial na demonstração do teorema que vem a seguir. Esse resultado encontra-se feito em [7].

Lema 5. *Considere a* ψ *dada em (3.14) e assuma (A).*

(i) Se $\lambda^0 > \lambda^1$ *e* $\gamma > 0$ *então*

$$(I + \lambda^0\psi')^{-1}(\gamma) < (I + \lambda^1\psi')^{-1}(\gamma).$$

(ii) Se $(\lambda_k^0)_{k\in\mathbb{N}}$ *e* $(\lambda_k^1)_{k\in\mathbb{N}}$ *são duas sequências positivas tais que* $\lambda_k^0 \geqslant \lambda_k^1$ *para todo* $k \geqslant 0$, *e*

$$\beta_{k+1}^0 = (I + \lambda_k^0\psi')^{-1}(\beta_k^0) \quad , \quad \beta_{k+1}^1 = (I + \lambda_k^1\psi')^{-1}(\beta_k^1)$$

são duas sequências proximais com $\beta_0^0 = \beta_0^1 \in (0, r_0]$. *Então* $\beta_k^0 \leqslant \beta_k^1$ *para todo* $k \geqslant 0$.

Demonstração. (i) Seja $\delta = (I + \lambda^1\psi')^{-1}(\gamma)$. Então $(I + \lambda^1\psi')(\delta) = \gamma$ e assim

$$(I + \lambda^0\psi')(\delta) = (I + \lambda^1\psi' - \lambda^1\psi' + \lambda^0\psi')(\delta) = (I + \lambda^1\psi')(\delta) + (\lambda^0 - \lambda^1)\psi'(\delta) > \gamma.$$

Usando a monotonicidade de $I + \lambda^0\psi'$ obtemos

$$(I + \lambda^1\psi')^{-1}(\gamma) = \delta > (I + \lambda^0\psi')^{-1}(\gamma).$$

(ii) Usaremos indução sobre k. Para $k = 0$ já temos que $\beta_0^0 \leqslant \beta_0^1$ (por hipótese, vale a igualdade). Suponhamos então que a desigualdade é válida para k, ou seja, $\beta_k^0 \leqslant \beta_k^1$, e mostraremos que a desigualdade vai continuar valendo para $k+1$. Usando a monotonicidade de $(I + \lambda_k^0\psi')^{-1}$ junto à hipótese de indução obtemos:

$$(I + \lambda_k^0\psi')^{-1}(\beta_k^0) \leqslant (I + \lambda_k^0\psi')^{-1}(\beta_k^1). \tag{3.18}$$

Usando agora a primeira parte do lema para $\lambda_k^0, \lambda_k^1 > 0$ e $\beta_k^1 > 0$ temos:

$$(I + \lambda_k^0\psi')^{-1}(\beta_k^1) \leqslant (I + \lambda_k^1\psi')^{-1}(\beta_k^1). \tag{3.19}$$

Combinando as desigualdades (3.18) e (3.19) obtemos

$$\beta^0_{k+1} = (I + \lambda^0_k \psi')^{-1}(\beta^0_k) \leqslant (I + \lambda^0_k \psi')^{-1}(\beta^1_k) \leqslant (I + \lambda^1_k \psi')^{-1}(\beta^1_k) = \beta^1_{k+1},$$

ou seja, $\beta^0_{k+1} \leqslant \beta^1_{k+1}$. □

Teorema 15. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função própria, convexa, semicontínua inferiormente com* $\arg\min f \neq \emptyset$ *e* $\min f = 0$. *Suponha ainda que* f *tem a propriedade KL em* $[0 < f < \bar{r}]$. *Seja* $(x_k)_{k\in\mathbb{N}}$ *um algoritmo abstrato com* $f(x_0) = r_0 \in (0, \bar{r})$ *e suponha que vale a hipótese (A) (no intervalo* $[0, \alpha_0]$ *com* $\psi(\alpha_0) = r_0$*). Defina a sequência proximal unidimensional "pior caso"*$(\alpha_k)_{k\in\mathbb{N}}$ *como em (3.16). Então,* $(x_k)_{k\in\mathbb{N}}$ *converge para algum minimizador* x^* *e, mais ainda,*

$$f(x_k) \leqslant \psi(\alpha_k), \ \forall k \geqslant 0, \tag{3.20}$$

$$\|x_k - x^*\| \leqslant \frac{b}{a}\alpha_k + \sqrt{\frac{\psi(\alpha_{k-1})}{a}}, \ \forall k \geqslant 1. \tag{3.21}$$

Demonstração. Primeiramente observamos que a sequência $(x_k)_{k\in\mathbb{N}}$ converge para algum minimizador x^* segue diretamente do Teorema 14. Agora, para cada $k \geqslant 1$, definimos $r_k := f(x_k)$. Se $r_k = 0$ então

$$f(x_k) = r_k = 0 \leqslant \psi(\alpha_k)$$

e a desigualdade é satisfeita. Se $r_k > 0$ então também temos $r_j > 0$, para $j = 1, \ldots, k$. Definimos $\beta_k = \psi^{-1}(r_k) > 0$, $s_k = \frac{\beta_{k-1} - \beta_k}{\psi'(\beta_k)} > 0$ e então

$$\begin{aligned} s_k \ \psi'(\beta_k) = \beta_{k-1} - \beta_k \ &\Rightarrow \ \beta_{k-1} = \beta_k + s_k \ \psi'(\beta_k) = (I + s_k\psi')(\beta_k) \\ &\Rightarrow \ \beta_k = (I + s_k\psi')^{-1}(\beta_{k-1}), \end{aligned}$$

ou seja, β_k é gerada pela sequência proximal

$$\beta_k = (I + s_k\psi')^{-1}(\beta_{k-1}).$$

Provaremos que $s_k \geqslant \zeta$, onde ζ é a constante dada em (3.15). Combinando a desigualdade KL e (H2), obtemos que

$$b^2\varphi'(r_k)^2\|x_k - x_{k-1}\|^2 \geqslant \varphi'(r_k)^2\|\omega_k\|^2 \geqslant 1 \tag{3.22}$$

onde ω_k é como em (H2). Usando (H1) temos que

$$1 \leqslant b^2\varphi'(r_k)^2\|x_k - x_{k-1}\|^2 \leqslant b^2\varphi'(r_k)^2\left(\frac{f(x_{k-1}) - f(x_k)}{a}\right),$$

isto é,

$$\frac{a}{b^2} \leqslant \varphi'(r_k)^2(f(x_{k-1}) - f(x_k)).$$

Agora (tendo em mente que $\psi = (\varphi|_{[0,r_0]})^{-1} : [0, \alpha_0] \to [0, r_0]$) a fórmula da derivada da função inversa nos dá

$$(\varphi^{-1})'(\beta_k) = \frac{1}{\varphi'(\varphi^{-1}(\beta_k))} \;\Rightarrow\; \psi'(\beta_k) = \frac{1}{\varphi'(r_k)} \;\Rightarrow\; \varphi'(r_k) = \frac{1}{\psi'(\beta_k)}$$

implicando em

$$\varphi'(r_k)^2 = \frac{1}{\psi'(\beta_k)^2}.$$

Segue daí, e das desigualdades obtidas, que

$$\frac{a}{b^2} \leqslant \varphi'(r_k)^2(r_{k-1} - r_k) = \frac{(\psi(\beta_{k-1}) - \psi(\beta_k))}{\psi'(\beta_k)^2}.$$

Como estamos supondo que ψ é l-Lipschitz contínua, podemos usar o Lema 2 para obter

$$\frac{a}{b^2} \leqslant \frac{(\beta_{k-1} - \beta_k)\psi'(\beta_k)}{\psi'(\beta_k)^2} + \frac{l}{2}\frac{(\beta_{k-1} - \beta_k)^2}{\psi'(\beta_k)^2} = \frac{(\beta_{k-1} - \beta_k)}{\psi'(\beta_k)} + \frac{l}{2}\frac{(\beta_{k-1} - \beta_k)^2}{\psi'(\beta_k)^2} = s_k + \frac{l}{2}s_k^2.$$

Daí,

$$\begin{aligned}
2ab^{-2} \leqslant 2s_k + ls_k^2 \;&\Rightarrow\; 2lab^{-2} \leqslant 2ls_k + l^2s_k^2 \\
&\Rightarrow\; 2lab^{-2} + 1 \leqslant 2ls_k + l^2s_k^2 + 1 \\
&\Rightarrow\; 2lab^{-2} + 1 \leqslant (ls_k + 1)^2 \\
&\Rightarrow\; \frac{\sqrt{2lab^{-2} + 1} - 1}{l} \leqslant s_k,
\end{aligned}$$

isto é, $s_k \geqslant \zeta$ para qualquer $k \geqslant 1$ tal que $r_k > 0$.

Observando agora que α_k e β_k estão definidas pelas sequências proximais

$$\alpha_{k+1} = (I + \zeta\, \psi')^{-1}(\alpha_k), \quad \beta_{k+1} = (I + s_k\psi')^{-1}(\beta_k)$$

e sabendo que $s_k \geqslant \zeta$, podemos usar o Lema 5 para obter que $\alpha_k \geqslant \beta_k$. Como ψ é crescente segue que

$$\psi(\alpha_k) \geqslant \psi(\beta_k) = r_k = f(x_k),$$

como queríamos.

Falta mostrar que vale a desigualdade (3.21). Para isto, usaremos o Teorema 14 juntamente com os resultados obtidos acima. Assim,

$$\begin{aligned}\|x_k - x^*\| &\leqslant \frac{b}{a}\varphi(f(x_k)) + \sqrt{\frac{f(x_{k-1})}{a}}\\ &= \frac{b}{a}\beta_k + \sqrt{\frac{f(x_{k-1})}{a}}\\ &\leqslant \frac{b}{a}\alpha_k + \sqrt{\frac{\psi(\alpha_{k-1})}{a}},\ \forall k \geqslant 1.\end{aligned}$$

□

Um importante caso a ser estudado é o caso em que $\psi(s) = \frac{l}{2}s^2$. Neste caso, a suposição (A) vale e obtemos o seguinte resultado particular do Teorema 15:

Corolário 6. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função própria, convexa, semicontínua inferiormente com* $\arg\min f \neq \emptyset$ *e* $\min f = 0$. *Assuma ainda que* f *tem a propriedade KL e que* $\psi(s) = \frac{l}{2}s^2$ *em* $[0 < f < \bar{r}]$. *Seja* $(x_k)_{k\in\mathbb{N}}$ *um algoritmo abstrato com* $f(x_0) = r_0 \in (0, \bar{r})$. *Defina a sequência proximal unidimensional "pior caso"* $(\alpha_k)_{k\in\mathbb{N}}$. *Seja* $\sigma = lb^{-2}$. *Então,* $(x_k)_{k\in\mathbb{N}}$ *converge para algum minimizador* x^* *e, mais ainda,*

$$f(x_k) \leqslant \frac{f(x_0)}{(1+2a\sigma)^k},\ \forall k \geqslant 0, \tag{3.23}$$

$$\|x_k - x^*\| \leqslant \left[1 + \frac{1}{a\sigma\sqrt{1+\frac{1}{2a\sigma}}}\right]\frac{\sqrt{\frac{1}{a}f(x_0)}}{(1+2a\sigma)^{\frac{k-1}{2}}},\ \forall k \geqslant 1. \tag{3.24}$$

Demonstração. Relembramos primeiro que a sequência unidimensional "pior caso" $(\alpha_k)_{k\in\mathbb{N}}$ será dada por $\alpha_0 = \varphi(r_0)$, e

$$\alpha_{k+1} = \arg\min\left\{\frac{l}{2}s^2 + \frac{1}{2\zeta}(s-\alpha_k)^2 : s \geqslant 0\right\}$$

para todo $k \geqslant 0$, onde

$$\zeta = \frac{\sqrt{1+2lab^{-2}}-1}{l}.$$

Como ψ é convexa temos

$$\begin{aligned}0 = l\,\alpha_{k+1} + \frac{1}{\zeta}(\alpha_{k+1}-\alpha_k) &\Rightarrow \alpha_k = \zeta\, l\,\alpha_{k+1} + \alpha_{k+1}\\ &\Rightarrow \alpha_k = (l\,\zeta+1)\alpha_{k+1}\\ &\Rightarrow \alpha_{k+1} = \frac{\alpha_k}{(1+l\,\zeta)}\end{aligned}$$

e então

$$\alpha_k = \frac{1}{(1+l\zeta)} \cdot \alpha_{k-1} = \frac{1}{(1+l\zeta)} \cdot \frac{\alpha_{k-2}}{(1+l\zeta)} = \frac{\alpha_{k-2}}{(1+l\zeta)^2} = \cdots = \frac{\alpha_0}{(1+l\zeta)^k}, \ \forall k \geqslant 0.$$

Usando a desigualdade (3.20) obtemos

$$f(x_k) \leqslant \frac{l}{2}\alpha_k^2 = \frac{l}{2}\left(\frac{\alpha_0}{(1+l\zeta)^k}\right)^2 \tag{3.25}$$

Como $\varphi = \psi^{-1}$ no domínio conveniente temos que

$$\varphi(t) = \frac{\sqrt{2t}}{\sqrt{l}}$$

nesse domínio e assim,

$$\alpha_0 = \psi^{-1}(r_0) = \varphi(r_0) = \frac{\sqrt{2r_0}}{\sqrt{l}} = \frac{\sqrt{2f(x_0)}}{\sqrt{l}}.$$

Além disso,

$$\zeta = \frac{\sqrt{1+2lab^{-2}}-1}{l} \quad \Rightarrow \quad 1+l\zeta = \sqrt{1+2lab^{-2}} = \sqrt{1+2a\sigma}.$$

Voltando a desigualdade (3.25) obtemos

$$f(x_k) \leqslant \frac{l}{2}\left[\left(\frac{\sqrt{2f(x_0)}}{\sqrt{l}}\right)\left(\frac{1}{(1+l\zeta)^k}\right)\right]^2 = \frac{f(x_0)}{(1+2a\sigma)^k}, \forall k \geqslant 0.$$

Portanto vale a desigualdade (3.23).

Para mostrar a segunda desigualdade note que

$$\frac{b}{a}\alpha_k = \frac{b}{a}\frac{\alpha_0}{(1+l\zeta)^k} = \frac{b}{a}\frac{1}{(1+l\zeta)^k}\frac{\sqrt{2f(x_0)}}{\sqrt{l}} = \frac{b}{a\sqrt{l}}\frac{\sqrt{2f(x_0)}}{(1+2lab^{-2})^{\frac{k}{2}}} \tag{3.26}$$

e

$$\sqrt{\frac{\psi(\alpha_{k-1})}{a}} = \sqrt{\frac{l}{2a}\alpha_{k-1}{}^2} = \sqrt{\frac{l}{2a}\left(\frac{\alpha_0}{(1+l\zeta)^{k-1}}\right)^2} = \sqrt{\frac{1+2lab^{-2}}{2a}}\frac{\sqrt{2f(x_0)}}{(1+2lab^{-2})^{\frac{k}{2}}}. \tag{3.27}$$

Aplicando em (3.21) temos que

$$\|x_k - x^*\| \leqslant \left[\frac{b}{a\sqrt{l}} + \sqrt{\frac{1}{2a}+\frac{l}{b^2}}\right]\frac{\sqrt{2f(x_0)}}{(1+2lab^{-2})^{\frac{k}{2}}} = \left[\frac{b}{a\sqrt{l}} + \sqrt{\frac{1}{2a}+\frac{l}{b^2}}\right]\frac{\sqrt{2f(x_0)}}{(1+2a\sigma)^{\frac{k}{2}}},$$

para todo $k \geqslant 1$. Para concluir, observe que

$$\left[\frac{b}{a\sqrt{l}} + \sqrt{\frac{1}{2a}+\frac{l}{b^2}}\right] = \sqrt{\frac{1}{2a}+\frac{l}{b^2}}\left[1+\frac{1}{\sqrt{\frac{a\sigma}{2}+a^2\sigma^2}}\right] = \sqrt{\frac{1+2a\sigma}{2a}}\left[1+\frac{1}{a\sigma\sqrt{1+\frac{1}{2a\sigma}}}\right]$$

donde segue que

$$\|x_k - x^*\| \leqslant \sqrt{\frac{1+2a\sigma}{2a}}\left[1+\frac{1}{a\sigma\sqrt{1+\frac{1}{2a\sigma}}}\right]\frac{\sqrt{2f(x_0)}}{(1+2a\sigma)^{\frac{k}{2}}} = \left[1+\frac{1}{a\sigma\sqrt{1+\frac{1}{2a\sigma}}}\right]\frac{\sqrt{\frac{1}{a}f(x_0)}}{(1+2a\sigma)^{\frac{k-1}{2}}},$$

para todo $k \geqslant 1$. Segue o resultado. $\square$

Capítulo 4

Problema de viabilidade convexa

Neste capítulo apresentamos o problema de viabilidade convexa com interseção regular. Faremos aqui uma aplicação de basicamente tudo o que foi visto nos capítulos anteriores. Assim como no capítulo anterior, praticamente todas as demonstrações desse capítulo podem ser encontradas em [7].

Para $m \geqslant 2$, consideremos os subconjuntos convexos fechados $C_1, \ldots, C_m$ de $\mathbb{R}^n$ os quais a interseção contém uma bola aberta não-vazia. A proposição que faremos agora será essencial para o que vem a seguir.

Proposição 8. *Suponha que existe* $\bar{x} \in \mathbb{R}^n$ *e* $R > 0$ *tal que*

$$B(\bar{x}, R) \subset C := \bigcap_{i=1}^{m} C_i.$$

Então,

$$\operatorname{dist}(x, \cap_{i=1}^m C_i) \leqslant \left(1 + \frac{2\|x - \bar{x}\|}{R}\right)^{m-1} \max\{\operatorname{dist}(x, C_i), i = 1, \ldots, m\}, \ \forall x \in \mathbb{R}^n. \quad (4.1)$$

Demonstração. Em um primeiro passo assumimos $m = 2$.
Ponha $C = C_1 \cap C_2$, fixe $x \in \mathbb{R}^n$ e defina $d = 2\max\{\operatorname{dist}(x, C_1), \operatorname{dist}(x, C_2)\}$. Sabemos que a função $\operatorname{dist}(\cdot, C_2)$ é 1-Lipschitz contínua. Assim,

$$|\operatorname{dist}(P_{C_1}(x), C_2) - \operatorname{dist}(x, C_2)| \leqslant \|P_{C_1}(x) - x\|$$

e então

$$\operatorname{dist}(P_{C_1}(x), C_2) \leqslant \|P_{C_1}(x) - x\| + \operatorname{dist}(x, C_2) = \operatorname{dist}(x, C_1) + \operatorname{dist}(x, C_2) \leqslant d.$$

Tomando

$$y = \overline{x} + \frac{R}{d}(P_{C_1}(x) - P_{C_2}(P_{C_1}(x)))$$

obtemos

$$\|y - \overline{x}\| = \frac{R}{d}\ \|P_{C_1}(x) - P_{C_2}(P_{C_1}(x))\| = \frac{R}{d}\ \mathrm{dist}(P_{C_1}(x), C_2) \leqslant \frac{R}{d}\ d = R,$$

isto é, $y \in B(\overline{x}, R) \subset C_1 \cap C_2$.

Construimos agora um ponto específico $z \in C$ como se segue

$$z = \frac{d}{R+d}y + \frac{R}{R+d}P_{C_2}(P_{C_1}(x)).$$

Como $\frac{d}{R+d} + \frac{R}{R+d} = 1$ e $y,\ P_{C_2}(P_{C_1}(x)) \in C_2$ temos que z é uma combinação convexa de dois elementos de C_2 donde segue que $z \in C_2$. Substituindo em z, o ponto y por $\overline{x} + \frac{R}{d}(P_{C_1}(x) - P_{C_2}(P_{C_1}(x)))$ obtemos

$$\begin{aligned}
z &= \frac{d}{R+d}\left(\overline{x} + \frac{R}{d}(P_{C_1}(x) - P_{C_2}(P_{C_1}(x)))\right) + \frac{R}{R+d}P_{C_2}(P_{C_1}(x)) \\
&= \frac{d}{R+d}\ \overline{x} + \frac{d}{R+d}\left(\frac{R}{d}\ P_{C_1}(x)\right) - \frac{d}{R+d}\left(\frac{R}{d}\ P_{C_2}(P_{C_1}(x))\right) + \frac{R}{R+d}\ P_{C_2}(P_{C_1}(x)) \\
&= \frac{d}{R+d}\ \overline{x} + \frac{R}{R+d}P_{C_1}(x) \in C_1.
\end{aligned}$$

Isso implica que $z \in C_1 \cap C_2$. Assim sendo,

$$\mathrm{dist}(x, C) \leqslant \|x - z\| = \|x - P_{C_1}(x) + P_{C_1}(x) - z\| \leqslant \|x - P_{C_1}(x)\| + \|z - P_{C_1}(x)\|.$$

Desde que $\overline{x} \in C_1 \cap C_2$,

$$\begin{aligned}
\|z - P_{C_1}(x)\| &= \left\|\frac{d}{R+d}\ \overline{x} + \frac{R}{R+d}P_{C_1}(x) - P_{C_1}(x)\right\| \\
&= \left\|\frac{d}{R+d}\ \overline{x} + \frac{R}{R+d}P_{C_1}(x) - \left(\frac{R+d}{R+d}\right) P_{C_1}(x)\right\| \\
&= \frac{d}{R+d}\|\overline{x} - P_{C_1}(x)\| \\
&= \frac{d}{R+d}\|P_{C_1}(\overline{x}) - P_{C_1}(x)\| \\
&\leqslant \frac{d}{R+d}\|\overline{x} - x\|,
\end{aligned}$$

donde a última desigualdade vem do fato de que o operador de projeção é não-expansivo.

Sabendo que

$$\|x - P_{C_1}(x)\| = \mathrm{dist}(x, C_1) \leqslant \frac{d}{2},$$

podemos combinar com os resultados acima para obter

$$\begin{aligned}
\operatorname{dist}(x, C) &\leqslant \frac{d}{2} + \frac{d}{R+d}\ \|x - \overline{x}\| \\
&= \frac{d}{2} + \frac{2d}{2(R+d)}\ \|x - \overline{x}\| \\
&= \frac{d}{2} + \frac{d}{2}\ \frac{2}{R+d}\ \|x - \overline{x}\| \\
&= \left(1 + \frac{2}{R+d}\ \|x - \overline{x}\|\right) \max\{\operatorname{dist}(x, C_1), \operatorname{dist}(x, C_2)\} \\
&\leqslant \left(1 + \frac{2\|x - \overline{x}\|}{R}\right) \max\{\operatorname{dist}(x, C_1), \operatorname{dist}(x, C_2)\} \qquad (4.2)
\end{aligned}$$

Agora, sabendo que a interseção de conjuntos convexos é um conjunto convexo, podemos tomar $m \geqslant 2$ arbitrário e aplicar a desigualdade acima para os dois conjuntos convexos C_1 e $\cap_{i=2}^m C_i$, obtendo então

$$\operatorname{dist}(x, \cap_{i=1}^m C_i) \leqslant \left(1 + \frac{2\|x - \overline{x}\|}{R}\right) \max\{\operatorname{dist}(x, C_1), \operatorname{dist}(x, \cap_{i=2}^m C_i)\}. \qquad (4.3)$$

Afirmação 2. *Se* $\lambda > 1$ *e* $a, b \geqslant 0$ *então* $\max\{a, \lambda b\} \leqslant \lambda \max\{a, b\}$.

Com efeito, se supomos $a \geqslant \lambda b$ *então, como* $\lambda > 1$*, teremos*

$$a \geqslant \lambda b > b.$$

Logo $\max\{a, b\} = a$ *implicando que* $a \leqslant \lambda a$*, o que é verdade pois* $\lambda > 1$*.*
Se $\lambda b \geqslant a$ *então devemos analisar os dois seguintes casos:* $b > a > \frac{a}{\lambda}$ *e* $a > b > \frac{a}{\lambda}$*. Se ocorre* $b > a > \frac{a}{\lambda}$ *então* $\max\{a, b\} = b$ *e a desiguadade se torna* $\lambda b \leqslant \lambda b$*. Por outro lado, se ocorre* $a > b > \frac{a}{\lambda}$ *então* $\max\{a, b\} = a$ *e assim teremos* $\lambda b \leqslant \lambda a$*, que por sua vez é verdade pois* $b < a$*. Portanto vale a afirmação.*

Agora, sabendo que

$$\operatorname{dist}(x, \cap_{i=2}^m C_i) \leqslant \left(1 + \frac{2\|x - \overline{x}\|}{R}\right) \max\{\operatorname{dist}(x, C_2), \operatorname{dist}(x, \cap_{i=3}^m C_i)\}$$

e tendo em mente a Afirmação 2, segue da desigualdade (4.3) que

$$\operatorname{dist}(x, \cap_{i=1}^m C_i) \leqslant$$

$$\leqslant \left(1 + \frac{2\|x - \overline{x}\|}{R}\right)\left(1 + \frac{2\|x - \overline{x}\|}{R}\right) \max\{\operatorname{dist}(x, C_1), \operatorname{dist}(x, C_2), \operatorname{dist}(x, \cap_{i=3}^m C_i)\} =$$

$$= \left(1 + \frac{2\|x - \overline{x}\|}{R}\right)^2 \max\{\operatorname{dist}(x, C_1), \operatorname{dist}(x, C_2), \operatorname{dist}(x, \cap_{i=3}^m C_i)\}.$$

Repetindo esse processo, agora $(m-2)$ vezes, obtemos

$$\mathrm{dist}(x, \cap_{i=1}^{m} C_i) \leqslant \left(1 + \frac{2\|x - \overline{x}\|}{R}\right)^{m-1} \max\{\mathrm{dist}(x, C_i), i = 1, \dots, m\}, \ \forall x \in \mathbb{R}^n.$$

□

Seja $C := \bigcap_{i=1}^{m} C_i$. Se $C \neq \emptyset$, encontrar um ponto em C é equivalente a minimizar a seguinte função convexa sobre $\mathbb{R}^n$

$$f(x) = \frac{1}{2}\sum_{i=1}^{m} \alpha_i \mathrm{dist}^2(x, C_i), \tag{4.4}$$

onde $\alpha_i > 0$ para todo $i = 1, \dots, m$ e $\sum_{i=1}^{m} \alpha_i = 1$. Agora, com o auxílio da Proposição 8, obteremos um error bounds para f. É claro que $C = \arg\min f = \{x \in \mathbb{R}^n : f(x) = 0\}$. Observamos ainda que

$$\max\{\mathrm{dist}(x, C_i), i = 1, \dots, m\} \leqslant \sqrt{\mathrm{dist}^2(x, C_1) + \cdots + \mathrm{dist}^2(x, C_m)}.$$

Assim, fixando $x_0 \in \mathbb{R}^n$ e usando a Proposição 8, temos que

$$\begin{aligned}
&\mathrm{dist}(x, C) \leqslant \\
&\leqslant \left(1 + \frac{2\|x_0 - \overline{x}\|}{R}\right)^{m-1} \sqrt{\mathrm{dist}^2(x, C_1) + \cdots + \mathrm{dist}^2(x, C_m)} = \\
&= \left(1 + \frac{2\|x_0 - \overline{x}\|}{R}\right)^{m-1} \frac{\left(\frac{2}{\min\limits_{i=1,\dots,m} \alpha_i}\right)^{1/2}}{\left(\frac{2}{\min\limits_{i=1,\dots,m} \alpha_i}\right)^{1/2}} \sqrt{\mathrm{dist}^2(x, C_1) + \cdots + \mathrm{dist}^2(x, C_m)} = \\
&= \left(1 + \frac{2\|x_0 - \overline{x}\|}{R}\right)^{m-1} \left(\frac{2}{\min\limits_{i=1,\dots,m} \alpha_i}\right)^{1/2} \sqrt{\frac{\left(\min\limits_{i=1,\dots,m} \alpha_i\right)\left(d^2(x, C_1) + \cdots + d^2(x, C_m)\right)}{2}} \leqslant \\
&\leqslant \left(1 + \frac{2\|x_0 - \overline{x}\|}{R}\right)^{m-1} \left(\frac{2}{\min\limits_{i=1,\dots,m} \alpha_i}\right)^{1/2} \sqrt{\frac{\alpha_1 \mathrm{dist}^2(x, C_1) + \cdots + \alpha_m \mathrm{dist}^2(x, C_m)}{2}} = \\
&= \left(1 + \frac{2\|x_0 - \overline{x}\|}{R}\right)^{m-1} \left(\frac{2}{\min\limits_{i=1,\dots,m} \alpha_i}\right)^{1/2} \sqrt{\frac{1}{2}\sum_{i=1}^{m} \alpha_i \mathrm{dist}^2(x, C_i)} = \\
&= \left(1 + \frac{2\|x_0 - \overline{x}\|}{R}\right)^{m-1} \left(\frac{2}{\min\limits_{i=1,\dots,m} \alpha_i}\right)^{1/2} \sqrt{f(x)},
\end{aligned}$$

para todo $x \in B(\bar{x}, \|x_0 - \bar{x}\|)$. Se tomamos

$$M = \frac{1}{4}\left(1 + \frac{2\|x_0 - \bar{x}\|}{R}\right)^{2-2m}\left(\min_{i=1,\dots,m} \alpha_i\right) \tag{4.5}$$

então a desigualdade acima pode ser reescrita como sendo

$$\operatorname{dist}(x, C) \leqslant \sqrt{\frac{f(x)}{2M}}. \tag{4.6}$$

Assim, a função f satisfaz a condição de error bounds local com função residual

$$\omega(s) = \sqrt{\frac{s}{2M}}.$$

Além disso, a função $\omega(s) = \sqrt{\frac{s}{2M}}$ tem um comportamento moderado. Para provar isto, note que

$$\omega'(s) = \frac{1}{2M}\frac{1}{2}\left(\frac{s}{2M}\right)^{-1/2} = \frac{1}{4M}\left(\frac{s}{2M}\right)^{-1/2}$$

implicando em

$$s\omega'(s) = \frac{s}{4M}\left(\frac{s}{2M}\right)^{-1/2} = \frac{1}{2}\left(\frac{s}{2M}\right)^{1/2} = \frac{1}{2}\omega(s) = c\omega(s), \text{ onde } c = \frac{1}{2}.$$

Agora, usando o Teorema 13 juntamente com as informações obtidas, deduzimos que f satisfaz a desigualdade KL em $B(\bar{x}, \|x_0 - \bar{x}\|) \cap [0 < f]$ com função desingularizante

$$\varphi(s) = \sqrt{\frac{2}{M}s}.$$

Começando de $x_0 \in \mathbb{R}^n$, chamaremos de *método de projeção baricêntrica* ao método que gera a sequência $(x_k)_{k\in\mathbb{N}}$ dada pela seguinte recursão

$$x_{k+1} = \sum_{i=1}^{m} \alpha_i P_{C_i}(x_k),$$

onde $\alpha_i > 0$ e $\sum_{i=1}^m \alpha_i = 1$. Usando a função $f = \frac{1}{2}\sum_{i=1}^{m} \alpha_i \operatorname{dist}^2(\cdot, C_i)$ e sabendo que $\operatorname{dist}^2(x, C_i) = \|x - P_{C_i}(x)\|^2$ verificamos que

$$\begin{aligned}
\nabla f(x) &= \Big(\alpha_1(x - P_{C_1}(x)) + \alpha_2(x - P_{C_2}(x)) + \cdots + \alpha_m(x - P_{C_m}(x))\Big) \\
&= \sum_{i=1}^{m} \alpha_i (x - P_{C_i}(x)) \\
&= x\sum_{i=1}^{m} \alpha_i - \sum_{i=1}^{m} \alpha_i P_{C_i}(x) \\
&= x - \sum_{i=1}^{m} \alpha_i P_{C_i}(x),
\end{aligned}$$

onde $x \in \mathbb{R}^n$. Logo a sequência $(x_k)_{k\in\mathbb{N}}$ pode ser descrita pela recursão

$$x_{k+1} = x_k - \nabla f(x_k), \; k \geqslant 0.$$

Afirmação 3. ∇f *é Lipschitz contínuo com constante* $L = 2$.

De fato, dados $x, y \in \mathbb{R}^n$*, temos que*

$$\begin{aligned}
\|\nabla f(x) - \nabla f(y)\| &= \left\| x - \sum_{i=1}^{m} \alpha_i P_{C_i}(x) - y + \sum_{i=1}^{m} \alpha_i P_{C_i}(y) \right\| \\
&\leqslant \|x - y\| + \left\| \sum_{i=1}^{m} \alpha_i P_{C_i}(x) - \sum_{i=1}^{m} \alpha_i P_{C_i}(y) \right\| \\
&\leqslant \|x - y\| + \sum_{i=1}^{m} \alpha_i \|x - y\| \\
&= \left(1 + \sum_{i=1}^{m} \alpha_i\right) \|x - y\| \\
&= 2\, \|x - y\|,
\end{aligned}$$

ou seja,

$$\|\nabla f(x) - \nabla f(y)\| \leqslant 2\, \|x - y\|, \; \forall x, y \in \mathbb{R}^n.$$

Como f é uma função convexa, diferenciável e ∇f é Lipschitz contínuo com constante 2 temos que $f = 0 + f$ satisfaz todas as condições do Exemplo 8. Assim, fazendo $\lambda^- = \frac{1}{2} = \lambda^+$ na Proposição 7, temos que

$$0 < \lambda_k = \frac{1}{2} < 1, \; \forall k \in \mathbb{N}$$

implicando que o método forward-backward para f gera uma sequência que satisfaz (H1) e (H2) com

$$a = \frac{1}{1/2} - \frac{2}{2} = 1 \quad \text{e} \quad b = \frac{1}{1/2} + 2 = 4,$$

ou seja, $(x_k)_{k\in\mathbb{N}} \subset \mathbb{R}^n$ é uma sequência que satisfaz:

$$f(x_k) + \|x_k - x_{k-1}\|^2 \leqslant f(x_{k-1}), \; \forall k \geqslant 1. \tag{4.7}$$

$$\|\nabla f(x_k)\| \leqslant 4\|x_k - x_{k-1}\|, \; \forall k \geqslant 1. \tag{4.8}$$

Afirmação 4. *Para qualquer* $\hat{x} \in C$*, a sequência* $(\|x_k - \hat{x}\|)_{k\in\mathbb{N}}$ *é decrescente.*

Para provar a afirmação, observamos primeiro que a convexidade de f *implica em*

$$f(\hat{x}) - f(x_k) \geqslant \langle \nabla f(x_k), \hat{x} - x_k \rangle, \; \forall k \in \mathbb{N}.$$

Daí,

$$\begin{aligned}\|x_{k+1}-\hat{x}\|^2 &= \|x_{k+1}-x_k\|^2+\|x_k-\hat{x}\|^2+2\langle x_{k+1}-x_k, x_k-\hat{x}\rangle\\ &= \|x_{k+1}-x_k\|^2+\|x_k-\hat{x}\|^2+2\langle \nabla f(x_k), \hat{x}-x_k\rangle\\ &\leqslant f(x_k)-f(x_{k+1})+\|x_k-\hat{x}\|^2+2[f(\hat{x})-f(x_k)]\\ &= 2f(\hat{x})-f(x_k)-f(x_{k+1})+\|x_k-\hat{x}\|^2\end{aligned}$$

Como $\hat{x}\in C$ *temos que* $f(\hat{x})=0$. *Temos ainda que* $f(x_k)\geqslant 0$ *para todo* $k\in\mathbb{N}$. *Segue daí que*

$$2f(\hat{x})-f(x_k)-f(x_{k+1})=-f(x_k)-f(x_{k+1})\leqslant 0,\ \forall k\in\mathbb{N}.$$

Logo,

$$\|x_{k+1}-\hat{x}\|^2 \leqslant \|x_k-\hat{x}\|^2,\ \forall k\in\mathbb{N}.$$

Isso implica que

$$\|x_{k+1}-\hat{x}\|\leqslant\|x_k-\hat{x}\|,\ \forall k\in\mathbb{N},$$

isto é, a sequência $(\|x_k-\hat{x}\|)_{k\in\mathbb{N}}$ *é decrescente para cada* $\hat{x}\in C$.

Como, para qualquer $\hat{x}\in C$, a sequência $(\|x_k-\hat{x}\|)_{k\in\mathbb{N}}$ é decrescente temos que $x_k\in B(\overline{x},\|x_0-\overline{x}\|)$ para todo $k\geqslant 0$. Como consequência, f tem uma função desingularizante global φ em $B(\overline{x},\|x_0-\overline{x}\|)$, dada por

$$\varphi(s)=\sqrt{\frac{2}{M}s}$$

a qual tem inversa

$$\psi(s)=\frac{M}{2}s^2,\ s\geqslant 0,$$

onde M é dada por (4.5). Usando o Corolário 6 com $a=1$, $b=4$ e $l=M$, obtemos:

Teorema 16. ***(Complexidade do método de projeção baricêntrica para interseção regular)***

A sequência $(x_k)_{k\in\mathbb{N}}$ *gerada pela projeção baricêntrica converge para um ponto* $x^*\in C$ *e*

$$f(x_k)\leqslant\frac{f(x_0)}{\left(1+\frac{M}{8}\right)^k},\ \forall k\geqslant 0,$$

$$\|x_k-x^*\|\leqslant\left[1+\frac{16}{M\sqrt{1+\frac{8}{M}}}\right]\frac{\sqrt{f(x_0)}}{\left(1+\frac{M}{8}\right)^{\frac{k-1}{2}}},\ \forall k\geqslant 1,$$

onde M *é dado em (4.5).*

Demonstração. Pelo Corolário 6, a sequência $(x_k)_{k\in\mathbb{N}}$ converge para um ponto x^* que está em $\arg\min f = C$. Além disso,

$$f(x_k) \leqslant \frac{f(x_0)}{(1+2lab^{-2})^k} = \frac{f(x_0)}{(1+2\cdot M\cdot 1\cdot \frac{1}{4^2})^k} = \frac{f(x_0)}{\left(1+\frac{M}{8}\right)^k},\ \forall k \geqslant 0.$$

e

$$\begin{aligned}
\|x_k - x^*\| &\leqslant \left[1 + \frac{1}{lab^{-2}\sqrt{1+\frac{1}{2lab^{-2}}}}\right] \frac{\sqrt{\frac{1}{a}f(x_0)}}{(1+2lab^{-2})^{\frac{k-1}{2}}} \\
&= \left[1 + \frac{1}{M\cdot 1\cdot \frac{1}{4^2}\sqrt{1+\frac{1}{2\cdot M\cdot 1\cdot \frac{1}{4^2}}}}\right] \frac{\sqrt{\frac{1}{1}f(x_0)}}{(1+2\cdot M\cdot 1\cdot \frac{1}{4^2})^{\frac{k-1}{2}}} \\
&= \left[1 + \frac{1}{\frac{M}{16}\sqrt{1+\frac{1}{\frac{M}{8}}}}\right] \frac{\sqrt{f(x_0)}}{\left(1+\frac{M}{8}\right)^{\frac{k-1}{2}}} \\
&= \left[1 + \frac{16}{M\sqrt{1+\frac{8}{M}}}\right] \frac{\sqrt{f(x_0)}}{\left(1+\frac{M}{8}\right)^{\frac{k-1}{2}}},\ \forall k \geqslant 1.
\end{aligned}$$

□

Referências Bibliográficas

[1] 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, no.2, (2010), 438-457.

[2] Attouch, H., Bolte, J., Svaiter, B. F. - *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward backward splitting, and regularized Gauss-Seidel methods*. Math. Program. Ser.A, 137(1-2), 91-129 (2013)

[3] Auslender, A., Crouzeix, J.-P - *Global regularity theorems*. Math. Oper. Res. 13, 243-253 (1988)

[4] Beck, A., Shtern, S. - *Linearly Convergent Away-Step Conditional Gradient for Non-strongly Convex Functions*, http://arxiv.org/abs/1504.05002

[5] Bocknak, J., Coste, M., Roy, M.-F. - *Real Algebraic Geometry*. Springer, Berlin (1998)

[6] Bolte, J., Daniilidis, A., Lewis, A. - *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*. SIAM J. Optim. 17, 1205-1223 (2006)

[7] Bolte, J., Nguyen, T.P., Peypouquet, J., Suter, B.W. - *From error bounds to the complexity of first-order descent methods for convex functions*. Math. Program., 1–37, 2016. doi:10.1007/s10107-016-1091-6

[8] Bolte, J., Sabach, S., Teboulle, M. - *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*. Math. Program. Ser.A 146, 1-16 (2013)

[9] Brézis H. - *Opérateurs maximaux monotones et semi-groupes de contractions dans les espace Hilbert*, North-Holland Mathematics studies 5, North-Holland Publishing Co., (1973)

[10] Bruck, R.E. - *Asymptotic convergence of nonlinear contraction semigroups in Hilbert space*. J. Funct. Anal. 18, 15-26 (1975)

[11] Cruz Neto, J.X., Oliveira, P.R., Soares Jr., P.A., Soubeyran, A. - *Learning how to play Nash, potential games and alternating minimization method for structured nonconvex problems on Riemannian manifolds*. J. Convex Anal. 20, 395-438 (2013)

[12] Dedieu, J.-P. - *Penalty functions in subanalytic optimization*. Optimization 26, 27-32 (1992)

[13] Hirsch, M. W. - *Differential Topology*. Spring Verlag, New York (1976)

[14] Hoffman, A.J. - *On approximate solutions of systems of linear inequalities*. J. Res. Natl. Bur. Stand. 49(4), 263-265 (1952)

[15] Ismailov, A., Solodov, M. *Otimização - volume 1 - Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade*. 3. ed. Rio de Janeiro: IMPA, 2014. 274p.

[16] Ismailov, A., Solodov, M. *Otimização - volume 2 - Métodos computacionais*. 2. ed. Rio de Janeiro: IMPA, 2012. 458p.

[17] Li, G. - *Global error bound for piecewise convex polynomials*. Math. Program. 137 (1-2), 37-64 (2013)

[18] Li, G., Mordukhovic B.S., Pham, T.S. - *New fractional error bound for polynomial systems with applications to Hölderian stability in optimization and spectral theory of tensors, to appear in Math. Program.*, Ser.A.

[19] Łojasiewicz, S. - *Division d'une distribution par une fonction analytique de variables réelles*. C. R. Acad. Sci., Paris 246, 683-686 (1958)

[20] Łojasiewicz, S. - *Sur la problème de la division*. Studia Mathematica 18, 87-136 (1959)

[21] Luo, X.D., Luo, Z.Q - *Extensions of Hoffman's error bound to polynomial systems.* SIAM J. Optim. 4, 383-392 (1994)

[22] Luo, Z.-Q., Pang, J.S. - *Error bounds for analytic systems and their application.* Math. Program. 67, 1-28 (1994)

[23] Luo, Z.-Q., Sturm, J.F. - *Error bound for quadratic systems.* Appl. Optim. 33, 383-404 (2000)

[24] Luo, Z.-Q., Tseng, P. - *Error bounds and convergence analysis of feasible descent methods: a general approach.* Ann. Oper. Res. 46-47(1), 157-178 (1993)

[25] Mangasarian, O.L. - *A condition number for differentiable convex inequalities.* Math. Oper. Res. 10, 175-179 (1985)

[26] Pang, J.S.-*Error bounds in mathematical programming.* Math. Program. 79, 299-332 (1997)

[27] Peypouquet, J. - *Convex Optimization in Normed Spaces: Theory, Methods and Examples.* Springer, Cham (2015)

[28] Ribeiro, A.A, Karas, E.W. - *Otimização contínua: aspectos teóricos e computacionais.* 1. ed. São Paulo: Cengage Learning, 2013. 272p.

[29] Robinson, S.M. - *An application of error bounds for convex programming in a linear space.* SIAM J. Control 13, 271-273 (1975)

[30] Udriste, C. - *Convex functions and optimization Algorithms on Riemannian manifolds, in: Mathematics and its Applications.* 297, Kluwer Academic Publishers, Dordrecht, Holland (1994)

[31] Vui, H.H. - *Global Hölderian error bound for non degenerate polynomials.* SIAM J. Optim. 23(2), 917-933 (2003)