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

Análise de convergência do método de descida inexata sob a condição de Kurdyka-Łojasiewicz

Valdinês Leite de Sousa Júnior

Teresina - 2012

Valdinês Leite de Sousa Júnior

Dissertação de Mestrado:

Análise de convergência do método de descida inexata sob a condição de Kurdyka-Łojasiewicz

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 Xavier da Cruz Neto

Teresina - 2012

Sobrenome, Iniciais no Nome.

xxxx Título da Dissertação.

Nome do Aluno – Teresina: ANO.

Orientador: Prof. Dr. XXXXXXXXXX.

Co-Orientador: Prof. Dr. XXXXXXXXXX.

1. Área de Concentração

CDD 516.36

Aos meus pais Maria Pessoa e Valdinês Leite, e ao meu irmão Diôgo.

Agradecimentos

Depois de tantos meses trabalhando nesta dissertação, encerro este trabalho escrevendo estes agradecimentos. São tantas as pessoas a quem sou grato, que este espaço torna-se pequeno e estreito para expor toda a minha gratidão pelos muitos que me ajudaram, me acolheram e me respeitaram. Inicialmente, agradeço por ter tido saúde, determinação e força para completar este mestrado.

Agradeço aos meus pais Maria Pessoa e Valdinês Leite por terem me dado a vida e por terem me amado desde o dia em que nasci. Ao meu irmão Diôgo Macário por seu companheirismo e amizade infinitos.

Aos meus tios Maria das Dores e Benedito Filho por tudo o que fizeram por mim durante os cinco anos em morei em sua casa. Aos meus primos Amaury, Alinson e Hillana por terem me tratado como irmão durante os últimos cinco anos.

Às minhas amigas de infância Gracy Caroline e Greice Kelly, sinto muitas saudades de vocês. Aos meus amigos Anna Thecya e Felipe Marreiros, vocês são incríveis. Aos meus grandes amigos e companheiros de mestrado: Yuri Rafael (mais conhecido como Bacharel, Yhuck, etc,...), Edvaldo Elias (o ED), Cleidinaldo Aguiar (Prof. Naldo), Alex Sandro (Grande Milgran), Antônio Kelson (Kelsinho), Franciane de Brito (Doutora Ciane), Renata Batista (Doutora Tinha), Leandro Pessoa (LPL), Edvalter Senna (Valtim), Italo Dowell, Aílton Campos (HAHAHA...). Aos meus amigos da UESPI pelos quatro anos que convivemos.

Aos professores da UESPI Afonso Norberto, Alessandro Wilk e Arnaldo Silva. Aos Professores da UFPI Paulo Sérgio, Paulo Alexandre, Jurandir Lopes, Juscelino Silva, Liane Mendes, Carlos Humberto, Sissy Sousa, Newton Luis, Halyson Baltazar. Um Agracecimento especial ao professor João Xavier da Cruz Neto, por ter sido meu orientador durante o mestrado, e me ajudado em muitas situações, não só acadêmicas, mas também por ter sido, juntamente com o prodessor Paulo Sérgio, um grande amigo. Ao professor Alfredo Noel Iusem por ter aceitado, gentilmente, ser membro da minha banca de mestrado.

Agradeço a todos que me ajudaram direta e(ou) indiretamente para a conclusão deste trabalho.

"*O que não nos mata, nos fortalece*".

Resumo

Nesta dissertação, será analisada a convergência do método de descida inexata sob a condição de Kurdyka-Łojasiewicz(KL) para funções não suaves. Além disso, faremos uma breve introdução sobre a teoria das funções KL, onde apresentaremos a demonstração da Desigualdade de Kurdyka-Łojasiewicz em uma estrutura o-minimal, além de uma prova da Desigualdade de Łojasiewicz para o caso real. Os resultados contidos neste trabalho foram extraídos dos artigos "Convergence of descent methods for semi-algebric and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods" de Hedy Attouch, Jérôme Bolte e Benar Fux Svaiter, e "On gradients of functions definable in o-minimal structures"de Krzysztof Kurdyka.

Abstract

In this dissertation, it will be analysed the convergence of the inexact-descent method over the condition of Kurdyka-Łojasiewicz(KL) for non-smooth functions. Moreover, we will do a brief introduction about the theory of KL function, were we will introduce a demonstration of Kurdyka-Łojasiewicz inequality in o-minimal structure, in addition, a proof of Łojasiewicz inequality for the real case. The results in this work were taken from the article "Convergence of descent methods for semi-algebric and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods" by Hedy Attouch, Jérôme Bolte and Benar Fux Svaiter, and "On gradients of functions definable in o-minimal structures" by Krzysztof Kurdyka.

Sumário

Introdução

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

$$\min f(x), \quad x \in \mathbb{R}^n, \tag{1}$$

é a seguinte. Dada uma aproximação $x^k \in \mathbb{R}^n$ da solução do problema, encontraremos um ponto $x^{k+1} \in \mathbb{R}^n$ tal que

$$f(x^{k+1}) < f(x^k).$$

Isto pode ser feito de várias maneiras. Uma realização desta estratégia básica é tomar uma direção $d^k \in \mathbb{R}^n$ tal que f é decrescente (pelo menos para passos curtos) a partir do ponto x^k nessa direção, e calcular um comprimento de passo $\alpha_k > 0$ que fornece um valor de f menor do que no ponto x^k,

$$f(x^k + \alpha_k d^k) < f(x^k).$$

Assim obtemos o iterando seguinte $x^{k+1} = x^k + \alpha_k d^k$. Repetimos o processo para o novo ponto x^{k+1}, etc. Métodos deste tipo se chamam *métodos de descida*.

Este trabalho abordará um lado bem interessante e talvez pouco explorado desse tipo de método. Com base em [2], objetivaremos minimizar funções $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ próprias, semi-contínuas inferiormente, não-suaves e não-convexas, provando resultados de convergência para o métodos de descida, que garantam a convergência de sequências limitadas para um ponto crítico de f. Para isso, consideraremos o *método de descida* que gera uma sequência $(x^k)_{k\in\mathbb{N}}$ que cumpre as seguintes condições:

$\mathcal{H}1$. (Condição suficiente de decréscimo). Para cada $k \in \mathbb{N}$,

$$f(x^{k+1}) + a\|x^{k+1} - x^k\|^2 \leqslant f(x^k);$$

$\mathcal{H}2$. (Condição relativa de erro). Para cada $k \in \mathbb{N}$, existe $w^{k+1} \in \partial f(x^{k+1})$ tal que

$$\|w^{k+1}\| \leqslant b\|x^{k+1} - x^k\|;$$

onde a e b são constantes positivas fixadas e $\partial f(x^{k+1})$ denota o subdiferencial limite de f em x^{k+1} (veja Definição 2 ítem (b)). A primeira condição caracteriza o modelo do método de descida, que torna a sequência $\{f(x^k)\}_{k\in\mathbb{N}}$ não-crescente. A segunda condição é fundamental para o desenvolvimento da teoria, pois geralmente necessitamos que $\partial f(x^k) \neq \emptyset$. Pedimos ainda que f satisfaça seguinte condição:

$\mathcal{H}3$. (Condição de continuidade). Existe uma subsequência $(x^{k_j})_{j\in\mathbb{N}}$ e $\widetilde{x}$ tal que

$$x^{k_j} \to \widetilde{x} \;\text{ e }\; f(x^{k_j}) \to f(\widetilde{x}), \text{quando } \; j \to \infty.$$

Esta é importante pois nem sempre é necessário que a função objetivo seja contínua em seu domínio efetivo. Veremos que em certos casos que, para garantir a convergência do algoritmo não é necessário nem mesmo supor este último (veja Observação 8).

Considere ainda a seguinte condição:

$\mathcal{H}4$. Para $\delta > 0$ existem $0 < \rho < \delta$ e $\nu > 0$ tal que

$$\left.\begin{array}{c} x \in B(x^*, \rho), f(x) < f(x^*) + \nu \\ y \notin B(x^*, \delta) \end{array}\right\} \Longrightarrow f(x) < f(y) + a\|y - x\|^2.$$

Sob certas hipóteses, esta condição nos permite estabelecer um resultado de convergência para um mínimo local (veja Teorema 11).

Além das condições descritas acima, a função objetivo deverá atender à Propriedade Kurdyka-Łojasiewicz (veja Definição 25). Essa importante desigualdade foi provada inicialmente para funções analíticas por Łojasiewicz em [13]. Ele provou o seguinte teorema:

Teorema 0.1 (Desigualdade de Łojasiewicz)**.** *Sejam* U *um subconjunto aberto de* $\mathbb{R}^n$*,* $f : U \to \mathbb{R}$ *uma função analítica e* $\overline{x} \in U$ *um ponto crítico de* f*. Então existem* $\theta \in [\frac{1}{2}, 1)$*,* $C > 0$ *e uma vizinhança* W *de* $\overline{x}$ *tal que*

$$\forall x \in W, \;|f(x) - f(\overline{x})|^{\theta} \leqslant C\|\nabla f(x)\|. \tag{2}$$

Mais tarde, em [11], Kurdyka apresentou um prova mais geral dessa desigualdade. Ele mostrou que se f é diferenciável em um domínio limitado, "definable"(veja Definição 22) em alguma estrutura o-minimal, então existe um função $\psi : \mathbb{R}_+ \to \mathbb{R}$ de classe C^1 e uma constante $c > 0$ tal que $\|\nabla(\psi \circ f)\| \geqslant c$ (veja Teorema 7). Em [8], Maurício Graña e Peterzil usam estruturas o-minimais para provar a convergência da trajetória central em programação semidefinida. Até onde temos conhecimento, este é o primeiro trabalho a relacionar Geometria Algébrica com Otimização, mais especificamente, fazem uso de um importante resultado de Geometria Algébrica, conhecido como Lema da Monotonicidade (Veja Lema 2). Em [1], Absil estabeleceu um resultado abstrato de convergência para sequências (que satisfaz uma condição de descida forte) no caso em que a função objetivo é analítica, definida em $\mathbb{R}^n$, e que satisfaz (2).

Este trabalho está dividido em quatro capítulos. O primeiro, engloba todos os resultados a serem utilizados nos capítulos seguintes; com destaque para a teoria dos subdiferenciais Fréchet e limite. No segundo capítulo, faremos uma breve introdução às funções KL (Kurdyka-Łojasiewicz), onde abordaremos resultados interessantes sobre essas funções, e será apresentada uma prova da desigualdade de Łojasiewicz no caso real (veja Teorema 5). O Capítulo 3 tem a missão de expor os principais resultados de convergência para o método de descida inexata, essencialmente teórico, une a teoria das funções KL com resultados de Otimização. Por último, analisaremos a convergência de algoritmos no contexto das funções KL, utilizando os resultados obtidos no Capítulo 3.

Capítulo 1

Preliminares

Neste capítulo abordaremos as definições e os resultados que serão utilizados nos capítulos seguintes.

1.1 Definições e resultados básicos

A definição a seguir reune vários conceitos úteis para este trabalho.

Definição 1. *Sejam* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *uma função,* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *uma aplicação ponto-conjunto e* C *uma subconjunto não-vazio de* $\mathbb{R}^n$.

1. *Os domínios efetivos de* F *e* f *são definidos, respectivamente por*
$$\text{dom } F := \{x \in \mathbb{R}^n : F(x) \neq \emptyset\},$$
e
$$\text{dom } f := \{x \in \mathbb{R}^n : f(x) < +\infty\}.$$

2. *Os gráficos de* F *e* f *são definidos, respectivamente por*
$$\text{Graf } F := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : y \in F(x)\},$$
e
$$\text{Graf } f := \{(x, s) \in \mathbb{R}^n \times \mathbb{R} : s = f(x)\}.$$

3. *Se* $x \in \text{dom } f$*, o* ***epígrafo*** *de* f *é dado por*
$$\text{epi } f := \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leqslant \lambda\}.$$

4. *Se* $\text{dom } f \neq \emptyset$*, o conjunto dos seus minimizadores globais, possivelmente vazio, é denotado por*
$$\text{argmin } f := \{x \in \mathbb{R}^n : f(x) = \inf f\}.$$

5. f *é dita* ***semi-contínua inferiormente*** *no ponto* $\overline{x} \in \mathbb{R}^n$*, quando para qualquer sequência* $x^k \to \overline{x}$ $(k \to \infty)$*, tem-se*
$$\liminf_{k\to\infty} f(x^k) \geqslant f(\overline{x}).$$
A função f *é semi-contínua inferiormente em* $\mathbb{R}^n$*, quando ela é semi-contínua inferiormente em todos os pontos de* $\mathbb{R}^n$*.*
Pode-se verificar que $f : \mathbb{R}^n \to \mathbb{R}$ *é semi-contínua inferiormente no conjunto fechado* D *se, e somente se, os seus conjuntos de nível* $L_{f,D}(c) = \{x \in D : f(x) \leqslant c\}$ *são fechados para todo* $c \in \mathbb{R}$*. Veja* [9].

6. *Se* C *é um conjunto convexo e* $x \in \operatorname{dom} f$*, então* f *é convexa em* C *quando para quaisquer* $x \in C$*,* $y \in C$ *e* $\alpha \in [0, 1]$*, tem-se*
$$f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha) f(y).$$

7. *f é* ***côncava*** *se, e somente se,* $-f$ *é convexa.*

8. *A função indicador* i_C *é definida por*
$$i_C(x) = \begin{cases} 0, & \textit{se} \quad x \in C \\ +\infty, & \textit{se} \quad x \notin C. \end{cases}$$
É fácil ver que i_C *é convexa se, e somente se,* C *é um conjunto convexo, e que* i_C *é semi-contínua inferiormente se, e somente se,* C *é fechado.*

9. *f é* ***coerciva*** *se, e somente se,*
$$\|x^k\| \to +\infty \Rightarrow \limsup_{k\to+\infty} f(x^k) = +\infty.$$
Note que se f *é uma função contínua e coerciva, então o conjunto* $\operatorname{argmin} f \neq \emptyset$*.*

10. *Uma aplicação* $f : C \to \mathbb{R}^m$ *diz-se* L*-Lipschitz contínua, quando existe uma constante* $L > 0$ *tal que*
$$\|f(y) - f(x)\| \leqslant L\|y - x\|, \ \forall x, y \in C.$$

Em grande parte deste trabalho, consideraremos funções não necessariamente convexas nem diferenciáveis, sendo assim, torna-se necessário estabelecer generalizações dos subdiferenciais. Estes fatos podem ser encontrados em Rockafellar e Wets [15]. Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.

Definição 2. (a) *Para cada* $\overline{x} \in \operatorname{dom} f$*, o* ***subdiferencial de Fréchet*** *de* f *em* $\overline{x}$*, denotado por* $\widehat{\partial} f(\overline{x})$*, é o conjunto dos vetores* $v \in \mathbb{R}^n$ *tais que*

$$\liminf_{y \to \overline{x}} \frac{1}{\|\overline{x} - y\|} \Big[f(y) - f(\overline{x}) - \langle v, y - \overline{x} \rangle \Big] \geqslant 0, \quad y \neq \overline{x}.$$

Quando $\overline{x} \notin \operatorname{dom} f$*, definimos* $\widehat{\partial} f(\overline{x}) = \emptyset$*. Em outras palavras,* $v \in \widehat{\partial} f(\overline{x})$ *se, e somente se,*

$$f(y) \geqslant f(\overline{x}) + \langle v, y - \overline{x} \rangle + o(\|y - \overline{x}\|),$$

onde $\lim_{y \to \overline{x}} \frac{o(\|y - \overline{x}\|)}{\|y - \overline{x}\|} = 0$.

Em algumas situações é conveniente considerar o subdiferencial de Fréchet como segue

$$\widehat{\partial} f(\overline{x}) = \Big\{ v \in \mathbb{R}^n : \liminf_{\|h\| \to 0} \frac{1}{\|h\|} \Big[f(\overline{x} + h) - f(\overline{x}) - \langle v, h \rangle \Big] \geqslant 0, \ h \neq 0 \Big\}.$$

(b) *O* ***subdiferencial-limite*** *(ou simplemente subdiferencial) de* f *em* $\overline{x} \in \operatorname{dom} f$*, denotado por* $\partial f(\overline{x})$*, é definido como segue*

$$\partial f(\overline{x}) := \{ v \in \mathbb{R}^n : \exists x^k \to \overline{x}, f(x^k) \to f(\overline{x}), v^k \in \widehat{\partial} f(x^k) \ \text{ e } \ v^k \to v \}.$$

Observação 1. *Da definição acima segue que*

$$\widehat{\partial} f(x) \subset \partial f(x) \ \textit{para todo} \ x \in \mathbb{R}^n.$$

Os exemplos 2 e 4 mostram que a inclusão acima pode ser estrita.

Quando f é uma função convexa os subdiferenciais Fréchet e limite tem a seguinte caracterização.

Proposição 1. *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *uma função convexa tal que* $\overline{x} \in \operatorname{dom} f$*. Então,*

$$\widehat{\partial} f(\overline{x}) = \{ v \in \mathbb{R}^n : f(y) \geqslant f(\overline{x}) + \langle v, y - \overline{x} \rangle, \ y \in \mathbb{R}^n \} = \partial f(\overline{x}). \tag{1.1}$$

Demonstração. Mostraremos inicialmente que

$$\widehat{\partial} f(\overline{x}) = \{ v \in \mathbb{R}^n : f(y) \geqslant f(\overline{x}) + \langle v, y - \overline{x} \rangle, \ y \in \mathbb{R}^n \}. \tag{1.2}$$

É claro que $\{ v \in \mathbb{R}^n : f(y) \geqslant f(\overline{x}) + \langle v, y - \overline{x} \rangle, \ y \in \mathbb{R}^n \} \subset \widehat{\partial} f(\overline{x})$. Por outro lado, se $\overline{v} \in \widehat{\partial} f(\overline{x})$, então

$$f(y) \geqslant f(\overline{x}) + \langle \overline{v}, y - \overline{x} \rangle + o(\|y - \overline{x}\|),$$

onde $\lim_{y \to \overline{x}} o(\|y - \overline{x}\|)/\|y - \overline{x}\| = 0$. Segue que para todo $z \in \mathbb{R}^n$ e para todo $t \in (0, 1]$,

$$f((1 - t)\overline{x} + tz) \geqslant f(\overline{x}) + t\langle \overline{v}, z - \overline{x} \rangle + o(\|t(z - \overline{x})\|).$$

Como f é convexa,

$$f(z) \geqslant f(\bar{x}) + \langle \bar{v}, z - \bar{x} \rangle + o(\|t(z - \bar{x})\|)/t.$$

Fazendo $t \to 0^+$ obtemos,

$$\bar{v} \in \{v \in \mathbb{R}^n : f(y) \geqslant f(\bar{x}) + \langle v, y - \bar{x} \rangle,\ y \in \mathbb{R}^n\}.$$

Por fim, afirmamos que

$$\partial f(\bar{x}) \subset \{v \in \mathbb{R}^n : f(y) \geqslant f(\bar{x}) + \langle v, y - \bar{x} \rangle,\ y \in \mathbb{R}^n\}. \tag{1.3}$$

De fato, se $\bar{v} \in \partial f(\bar{x})$, existem $x^k \to \bar{x}$, tal que $f(x^k) \to f(\bar{x})$ e $v^k \in \widehat{\partial} f(x^k)$ com $v^k \to \bar{v}$. Como $v^k \in \widehat{\partial} f(x^k)$, tem-se que para todo $y \in \mathbb{R}^n$,

$$f(y) \geqslant f(x^k) + \langle v^k, y - x^k \rangle.$$

Fazendo $k \to +\infty$, obtemos

$$f(y) \geqslant f(\bar{x}) + \langle \bar{v}, y - \bar{x} \rangle.$$

Isso prova (1.3).

De (1.2) e (1.3) conclui-se que,

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

Portanto,

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

□

Exemplo 1. *Seja* $f = i_{[0,1]} : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$*, definida por*

$$i_{[0,1]}(x) = \begin{cases} 0, & se \quad x \in [0,1] \\ +\infty, & se \quad x \notin [0,1]. \end{cases}$$

Como o intervalo $[0,1]$ *é convexo, segue que a função indicadora* $i_{[0,1]}$ *é convexa. Logo* $\widehat{\partial} f(x) = \partial f(x)$ *para todo* $x \in \operatorname{dom} f$*. Calculemos então* $\partial f(x)$*.*

- *Se* $x \notin [0,1]$*, então* $i_{[0,1]}(x) = +\infty$*, o que implica que* $\partial f(x) = \emptyset$*;*

- *Se* $x \in (0,1)$*, então*

$$\partial f(x) = \{v \in \mathbb{R} : f(y) \geqslant v(y - x),\ y \in \mathbb{R}\} = \{v \in \mathbb{R} : vx \geqslant vy,\ y \in \mathbb{R}\}.$$

Analisaremos os seguites subcasos.

- *Se* $y \neq [0,1]$*, então* $\partial f(x) = \mathbb{R}$*;*

- *Se* $y \in (0,1)$*, então* $\partial f(x) = \{v \in \mathbb{R} : vx \geqslant vy\}$*. Supondo* $v \neq 0$*, é possível escolher* y *de forma a chegar a uma contradição. Isso nos diz que* $\partial f(x) = \{0\}$*;*

- *Se* $y = 0$, *então* $\partial f(x) = \{v \in \mathbb{R} : vx \geqslant 0\}$, *logo,* $\partial f(x) = [0, +\infty)$;

- *Se* $y = 1$, *tem-se* $\partial f(x) = \{v \in \mathbb{R} : v(x-1) \geqslant 0\} = (-\infty, 0]$.

Concluímos que se $x \in (0,1)$, *então* $\partial f(x) = \mathbb{R} \cap \{0\} \cap [0, +\infty) \cap (-\infty, 0] = \{0\}$.

- *Se* $x = 0$, *segue que* $\partial f(0) = \{v \in \mathbb{R} : f(y) \geqslant vy,\ y \in \mathbb{R}\}$. *Assim, para* $y \notin [0,1]$ *temos que* $\partial f(0) = \mathbb{R}$, *e para* $y \in [0,1]$ *temos que* $\partial f(0) = (-\infty, 0]$. *Considerando a interseção dos dois casos, temos que* $\partial f(0) = (-\infty, 0]$;

- *Se* $x = 1$, *temos que* $\partial f(1) = \{v \in \mathbb{R} : f(y) \geqslant v(y-1),\ y \in \mathbb{R}\}$. *Daí, se* $y \notin [0,1]$ *ou se* $y = 1$, *temos que* $\partial f(1) = \mathbb{R}$. *Por outro lado, se* $y \in [0,1)$ *obtemos que* $\partial f(1) = [0, +\infty)$. *Logo, tomando a interseção dos três casos anteriores, temos que* $\partial f(1) = [0, +\infty)$.

Portanto,

$$\partial f(x) = \begin{cases} \emptyset, & se \quad x < 0, \\ (-\infty, 0\big], & se \quad x = 0, \\ \{0\}, & se \quad x \in (0,1), \\ \big[0, +\infty), & se \quad x = 1, \\ \emptyset, & se \quad x > 1. \end{cases}$$

A concavidade de uma função não garante a igualdade entre os subdiferenciais Fréchet e limite, como mostra o próximo exemplo.

Exemplo 2. *Se* $f : \mathbb{R} \to \mathbb{R}$ *é tal que* $f(x) = -|x|$, *então*

$$\widehat{\partial} f(x) = \begin{cases} \{-1\}, & se \quad x > 0, \\ \{1\}, & se \quad x < 0, \\ \emptyset, & se \quad x = 0, \end{cases} \tag{1.4}$$

$$\partial f(x) = \begin{cases} \{-1\}, & se \quad x > 0, \\ \{1\}, & se \quad x < 0, \\ \{-1, 1\}, & se \quad x = 0. \end{cases} \tag{1.5}$$

Com efeito, se $x = 0$ *e* h *é um número não-nulo suficientemente pequeno, obtemos*

$$\begin{aligned}
\frac{1}{|h|}\Big[f(x+h) - f(x) - \langle v, h\rangle\Big] &= \frac{1}{|h|}\Big[-|h| - vh\Big] \\
&= -1 - v\frac{h}{|h|} \\
&= \begin{cases} -1 - v, & \textit{se} \quad h > 0, \\ v - 1, & \textit{se} \quad h < 0. \end{cases}
\end{aligned}$$

Logo, se $v \in \widehat{\partial} f(0)$ *e* $h > 0$ *temos que* $v \leqslant -1$*. Se* $v \in \widehat{\partial} f(0)$ *e* $h < 0$ *temos que* $v \geqslant 1$*. Tomando a interseção dos dois casos, vemos que* $\widehat{\partial} f(0) = \emptyset$*.*

Suponha agora que $x > 0$*. Daí,*

$$\begin{aligned}
\frac{1}{|h|}\Big[f(x+h) - f(x) - \langle v, h\rangle\Big] &= \frac{1}{|h|}\Big[-|x+h| + |x| - vh\Big] \\
&= \frac{1}{|h|}\Big[-|x+h| + x - vh\Big] \\
&= \begin{cases} -\dfrac{h}{|h|}(1+v), & \textit{se} \quad x \geqslant -h, \\ \dfrac{1}{|h|}(2x + h - vh), & \textit{se} \quad x < -h. \end{cases}
\end{aligned}$$

Dado $x > 0$*, para* h *suficientemente pequeno podemos supor que* $0 < |h| < x$*. Assim,*

$$\frac{1}{|h|}\Big[f(x+h) - f(x) - \langle v, h\rangle\Big] = \begin{cases} -1 - v, & \textit{se} \quad h > 0, \\ 1 + v, & \textit{se} \quad h < 0. \end{cases}$$

Logo, se $v \in \widehat{\partial} f(x)$ *e* $h > 0$ *temos que* $v \leqslant -1$*. Se* $v \in \widehat{\partial} f(x)$ *e* $h < 0$ *temos que* $v \geqslant -1$*. Tomando novamente a interseção dos dois casos, vemos que se* $x > 0$*, então* $\widehat{\partial} f(x) = \{-1\}$*.*

Um argumento análogo mostra que se $x < 0$*, então* $\widehat{\partial} f(x) = \{1\}$*. Isso conclui a prova de* (1.4)*. A prova de* (1.5) *segue diretamente de* (1.4)*.*

Proposição 2. *Se* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *é diferenciável em* $\overline{x}$*, então* $\widehat{\partial} f(\overline{x}) = \{\nabla f(\overline{x})\}$*. Além disso, se* f *é de classe* C^1 *em uma vizinhança de* $\overline{x}$*, então* $\partial f(\overline{x}) = \{\nabla f(\overline{x})\}$*.*

Demonstração. Como f é diferenciável em $\overline{x}$, temos

$$f(y) = f(\overline{x}) + \langle \nabla f(\overline{x}), y - \overline{x}\rangle + o(\|y - \overline{x}\|) \quad \forall y \in \mathbb{R}^n,$$

isto é, $\nabla f(\overline{x}) \in \widehat{\partial} f(\overline{x})$.

Suponha agora que exista $\overline{v} \in \widehat{\partial} f(\overline{x})$, $\overline{v} \neq \nabla f(\overline{x})$. Daí, para todo $h \in \mathbb{R}^n$,

$$o(\|h\|) + \langle \overline{v}, h\rangle \leqslant f(\overline{x} + h) - f(\overline{x}) = \langle \nabla f(\overline{x}), h\rangle + o^*(\|h\|),$$

o que implica em,

$$\langle \bar{v} - \nabla f(\bar{x}), h\rangle \leqslant o^*(\|h\|) - o(\|h\|).$$

Tomando

$$h^k = \frac{\bar{v} - \nabla f(\bar{x})}{k\|\bar{v} - \nabla f(\bar{x})\|},$$

temos que $h^k \to 0$ $(k \to \infty)$ e

$$\|\bar{v} - \nabla f(\bar{x})\|/k \leqslant o^*(1/k) - o(1/k),$$

o que é possível somente quando $\bar{v} - \nabla f(\bar{x}) = 0$. Essa contradição mostra que $\widehat{\partial} f(\bar{x}) = \{\nabla f(\bar{x})\}$.

Para mostrar que $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$, tomemos um ponto $\bar{v} \in \partial f(\bar{x})$, logo que existem $x^k \to \bar{x}$ tal que $f(x^k) \to f(\bar{x})$ e $v^k \in \widehat{\partial} f(x^k)$ com $v^k \to \bar{v}$. Seja U uma vizinhança de $\bar{x}$ onde f é de classe C^1. Como $x^k \to \bar{x}$, temos que existe $k_0 \in \mathbb{N}$ tal que $\forall k \geqslant k_0$, $x^k \in U$. Assim,

$$\{v^k\} = \{\nabla f(x^k)\} = \widehat{\partial} f(x^k).$$

Logo,

$$\bar{v} = \lim_{k\to\infty} v^k = \lim_{k\to\infty} \nabla f(x^k) = \nabla f(\bar{x}),$$

ou seja, $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$. $\square$

A seguir, apresentaremos um exemplo de uma função que não é convexa mas os subdiferenciais Fréchet e limite coincidem.

Exemplo 3. *Seja* $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \sqrt{|x|}$. *Então*

$$\widehat{\partial} f(x) = \partial f(x) = \begin{cases} \left\{\dfrac{1}{2\sqrt{|x|}}\right\}, & \textit{se} \quad x \neq 0, \\ (-\infty, +\infty), & \textit{se} \quad x = 0. \end{cases} \tag{1.6}$$

De fato, se $x = 0$, *então para todo* $v \in \mathbb{R}$,

$$\frac{\sqrt{|h|} - vh}{|h|} = \frac{1}{\sqrt{|h|}} - v\frac{h}{|h|} \geqslant \frac{1}{\sqrt{|h|}} - |v|,$$

o que implica em

$$\liminf_{|h|\to 0} \frac{\sqrt{|h|} - vh}{|h|} \geqslant \liminf_{|h|\to 0} \Big(\frac{1}{\sqrt{|h|}} - |v|\Big) = +\infty.$$

Portanto, $\widehat{\partial} f(0) = (-\infty, +\infty)$. *Além disso,* $(-\infty, +\infty) = \widehat{\partial} f(0) \subset \partial f(0) \subset (-\infty, +\infty)$, *o que implica que* $\partial f(0) = (-\infty, +\infty)$.

Por outro, se $x \neq 0$, f *é de classe* C^1 *e*

$$\widehat{\partial} f(x) = \partial f(x) = \{f'(x)\} = \{1/2\sqrt{|x|}\}.$$

Isso prova (1.6).

Quando uma função $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ é convexa e diferenciável em um ponto $\bar{x} \in \operatorname{dom} f$, vê-se que $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$. Entretanto, se não tivermos a hipótese de convexidade, essa igualdade nem sempre é verdadeira, como mostra o próximo exemplo.

Exemplo 4. *Seja* $f : \mathbb{R} \to \mathbb{R}$ *definida por*

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

Então $\widehat{\partial} f(0) = \{0\}$ *e* $\partial f(0) = [-1, 1]$. *De fato, como* f *é diferenciável em* 0 *temos que* $\widehat{\partial} f(0) = \{f'(0)\} = \{0\}$. *Além disso, se* $x_0 \in [-1, 1] \setminus \{0\}$, *então existe* $v \in \mathbb{R}$ *tal que* $\cos v = x_0$. *Se* $x^k = 1/(x_0 + k\pi)$ *e* $v^k = 2x^k \sin(1/x^k) - \cos(1/x^k)$, *vemos que* $v \in \partial f(0)$. *Por outro lado, se* $v \in \partial f(0) \setminus \{0\}$, *existem* $x^k \to 0$ *tal que* $f(x^k) \to 0$ *e* $v^k \in \widehat{\partial} f(x^k)$ *com* $v^k \to v$. *Segue que* $v^k = 2x^k \sin(1/x^k) - \cos(1/x^k)$ *e, consequentemente,*

$$|v| \leqslant \lim_{k\to\infty} (|v - v^k| + |v^k|) \leqslant \lim_{k\to\infty} (|v - v^k| + 2|x^k| + 1) = 1.$$

Portanto, $\partial f(0) = [-1, 1]$.

Proposição 3. *Sejam* $f_1 : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *e* $f_2 : \mathbb{R}^n \to \mathbb{R}$ *funções, onde* $\bar{x} \in \operatorname{dom} f_1$, *e* f_2 *é de classe* C^1 *em um vizinhança de* $\bar{x}$. *Se* $f = f_1 + f_2$, *então*

(a) $\widehat{\partial} f(\bar{x}) = \widehat{\partial} f_1(\bar{x}) + \nabla f_2(\bar{x})$;

(b) $\partial f(\bar{x}) = \partial f_1(\bar{x}) + \nabla f_2(\bar{x})$.

Demonstração. (a) Sejam U uma vizinhança de $\bar{x}$ onde f_2 é de classe C^1, e h suficientemente pequeno tal que $\bar{x} + h \in U$. Assim,

$$f_2(\bar{x} + h) = f_2(\bar{x}) + \langle \nabla f_2(\bar{x}), h \rangle + o^{**}(||h||).$$

Tome $\bar{v} \in \widehat{\partial} f_1(\bar{x}) + \nabla f_2(\bar{x})$; então $\bar{v} = \bar{w} + \nabla f_2(\bar{x})$, $\bar{w} \in \widehat{\partial} f_1(\bar{x})$. Segue que,

$$f_1(\bar{x} + h) \geqslant f_1(\bar{x}) + \langle \bar{w}, h \rangle + o^{*}(||h||).$$

Mortraremos que se h é suficientemente próximo de zero, então

$$f(\bar{x} + h) - f(\bar{x}) - \langle \bar{w}, h \rangle \geqslant o(||h||).$$

De fato,

$$\begin{aligned} f(\bar{x} + h) - f(\bar{x}) - \langle \bar{v}, h \rangle &= f_1(\bar{x} + h) + f_2(\bar{x} + h) - f_1(\bar{x}) - f_2(\bar{x}) - \langle \bar{w} + \nabla f_2(\bar{x}), h \rangle \\ &= f_1(\bar{x} + h) - f_1(\bar{x}) - \langle \bar{w}, h \rangle + f_2(\bar{x} + h) - f_2(\bar{x}) - \langle \nabla f_2(\bar{x}), h \rangle \\ &\geqslant o^{*}(||h||) + o^{**}(||h||) = o(||h||). \end{aligned}$$

Logo, $\bar{v} \in \widehat{\partial} f(\bar{x})$.

Por outro lado, se $\bar{v} \in \widehat{\partial} f(\bar{x})$, mostraremos que $\bar{v} - \nabla f_2(\bar{x}) \in \widehat{\partial} f_1(\bar{x})$. Com efeito,

$$\begin{aligned} f_1(\bar{x}+h) - f_1(\bar{x}) - \langle \bar{v} - \nabla f_2(\bar{x}), h\rangle &= f(\bar{x}+h) - f(\bar{x}) - \langle \bar{v}, h\rangle - \\ &\quad \Big[f_2(\bar{x}+h) - f_2(\bar{x}) - \langle \nabla f_2(\bar{x}), h\rangle\Big] \\ &\geqslant o(\|h\|) + o^{**}(\|h\|) = o^{*}(\|h\|). \end{aligned}$$

Portanto, $\bar{v} - \nabla f_2(\bar{x}) \in \widehat{\partial} f_1(\bar{x})$. Daí, $\bar{v} = \bar{v} - \nabla f_2(\bar{x}) + \nabla f_2(\bar{x}) \in \widehat{\partial} f_1(\bar{x}) + \nabla f_2(\bar{x})$.

(b) Se $\bar{v} \in \partial f_1(\bar{x}) + \nabla f_2(\bar{x})$, então $\bar{v} = \bar{w} + \nabla f_2(\bar{x})$, $\bar{w} \in \partial f(\bar{x})$. Segue que existem $y^k \to \bar{x}$ tal que $f_1(y^k) \to f_1(\bar{x})$ e $w^k \in \widehat{\partial} f_1(y^k)$ com $w^k \to \bar{w}$.

Defina,

$$x^k := y^k \;\text{ e }\; v^k := w^k + \nabla f_2(y^k) \in \widehat{\partial} f_1(x^k) + \nabla f_2(x^k) = \widehat{\partial} f(x^k).$$

Assim, $x^k \to \bar{x}$, $f(x^k) = f_1(x^k) + f_2(x^k) \to f_1(\bar{x}) + f_2(\bar{x}) = f(\bar{x})$ e $v^k = w^k + \nabla f_2(y^k) \to \bar{w} + \nabla f_2(\bar{x}) = \bar{v}$. Isso implica que $\bar{v} \in \partial f(x^k)$.

Se $\bar{v} \in \partial f(x^k)$, devemos mostrar que $\bar{v} - \nabla f_2(\bar{x}) \in \partial f_1(x^k)$. Como $\bar{v} \in \partial f(x^k)$, existem $x^k \to \bar{x}$ tal que $f(x^k) \to f(\bar{x})$ e $v^k \in \widehat{\partial} f(x^k)$ com $v^k \to \bar{v}$. Assim, $v^k = w^k + \nabla f_2(y^k)$, $w^k \in \partial f_1(x^k)$. E consequentemente, $w^k \to \bar{v} - \nabla f_2(\bar{x})$. Portanto, $\bar{v} - \nabla f_2(\bar{x}) \in \partial f_1(x^k)$. □

Proposição 4. *Seja* $(x^k, v^k)_{k\in\mathbb{N}}$ *uma sequência em* $\mathbb{R}^n \times \mathbb{R}^n$ *tal que* $(x^k, v^k) \in \mathrm{Graf}\,\partial f$ *para todo* $k \in \mathbb{N}$. *Se* (x^k, v^k) *converge para* (x, v), *e* $f(x^k)$ *converge para* $f(x)$, *então* $(x, v) \in \mathrm{Graf}\,\partial f$.

Demonstração. Use diagonal de Cantor ou veja página 441 de [3]. □

Exemplo 5. *O resultado acima não vale para o subdiferencial de Fréchet, basta tomar* $f : \mathbb{R} \to \mathbb{R}$, $f(x) = -|x|$ *e* $(x^k, v^k) = (-1/k, 1)$.

Definição 3. *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. *Um ponto* $\bar{x} \in \mathbb{R}^n$ *é dito ponto crítico de* f *se* $0 \in \partial f(\bar{x})$. *O conjunto dos pontos críticos de* f *é denotado por*

$$\text{crit } f = \{x \in \mathbb{R}^n : 0 \in \partial f(x)\}.$$

Observação 2. *De acordo com Proposição 4, se mostrarmos que existe uma sequência* $(x^k, w^k) \in \mathrm{Graf}\,\partial f$ *tal que* $(x^k, w^k) \to (\bar{x}, 0)$ *e* $f(x^k) \to f(\bar{x})$, *então* $\bar{x}$ *é um ponto crítico.*

Definição 4. *Seja* $C \subset \mathbb{R}^n$ *e* $\bar{x} \in C$.

(a) *O cone normal regular de* C *no ponto* $\bar{x}$, *é definido por*

$$\widehat{N}_C(\bar{x}) = \{v \in \mathbb{R}^n : \langle v, x - \bar{x}\rangle \leqslant o(\|x - \bar{x}\|), \text{ para } x \in C\}.$$

(b) *O cone normal de* C *no ponto* $\overline{x}$ *é dado por*

$$N_C(\overline{x}) = \{v \in \mathbb{R}^n : \exists x^k \to \overline{x},\ x^k \in C \text{ e } v^k \to v \text{ com } v^k \in \widehat{N}_C(x^k)\}.$$

Proposição 5. *Seja* C *um subconjunto não-vazio e fechado de* $\mathbb{R}^n$*. Se* $\overline{x} \in C$*, então*

(i) $\widehat{\partial} i_C(\overline{x}) = \widehat{N}_C(\overline{x})$,

(ii) $\partial i_C(\overline{x}) = N_C(\overline{x})$.

Demonstração. (i) Se $v \in \widehat{\partial} i_C(\overline{x})$, ent ao $\langle v, y - \overline{x}\rangle \leqslant -o(\|y - \overline{x}\|)$, $y \in C$. Portanto, $v \in \widehat{N}_C(\overline{x})$.

Se, por outro lado, $v \in \widehat{N}_C(\overline{x})$, então

$$0 \leqslant o(\|y-\overline{x}\|) - \langle v, y-\overline{x}\rangle \Rightarrow 0 \leqslant \liminf_{y\to\overline{x}} \frac{1}{\|y-\overline{x}\|}[i_C(y) - i_C(\overline{x}) - \langle v, y-\overline{x}\rangle] \Rightarrow v \in \widehat{\partial} f(\overline{x}).$$

Portanto, $\widehat{N}_C(\overline{x}) = \widehat{\partial} f(\overline{x})$.

(ii) Como $\overline{x} \in C$, a condiç ao $i_C(x^k) \to i_C(\overline{x})$ implica que $i_C(x^k) = i_C(\overline{x}) = 0$. Daí,

$$\begin{aligned}
\partial i_C(\overline{x}) &= \{v \in \mathbb{R}^n : \exists x^k \to \overline{x}, i_C(x^k) \to i_C(\overline{x}), v^k \in \widehat{\partial} i_C(x^k) \text{ e } v^k \to v\} \\
&= \{v \in \mathbb{R}^n : \exists x^k \to \overline{x}, x^k \in C, v^k \in \widehat{\partial} i_C(x^k) \text{ e } v^k \to v\} \\
&= \{v \in \mathbb{R}^n : \exists x^k \to \overline{x}, x^k \in C, v^k \in \widehat{N}_C(\overline{x}) \text{ e } v^k \to v\} \\
&= N_C(\overline{x}).
\end{aligned}$$

Onde a terceira igualdade vale por (i). □

Corolário 1. *Seja* C *um subconjunto não-vazio, fechado e convexo de* $\mathbb{R}^n$*. Se* $\overline{x} \in C$*, então*

$$\widehat{N}_C(\overline{x}) = \{v \in \mathbb{R}^n : \langle v, x - \overline{x}\rangle \leqslant 0,\ \textit{para } x \in C\} = N_C(\overline{x}).$$

Demonstração. Como C é um conjunto convexo, segue da Proposição 5 que $\widehat{N}_C(\overline{x}) = N_C(\overline{x})$. Por outro lado, é claro que $\{v \in \mathbb{R}^n : \langle v, x - \overline{x}\rangle \leqslant 0,\ \textit{para } x \in C\} \subset \widehat{N}_C(\overline{x})$. Além disso, se $v \in \widehat{N}_C(\overline{x})$, tem-se que,

$$\langle v, x - \overline{x}\rangle \leqslant o(\|x - \overline{x}\|),$$

para $x \in C$. Logo, para todo $x \in C$ e para todo $t \in (0, 1]$,

$$\langle v, t(x - \overline{x})\rangle \leqslant o(\|t(x - \overline{x})\|),$$

e portanto,

$$\langle v, x - \overline{x}\rangle \leqslant o(\|t(x - \overline{x})\|)/t.$$

Fazendo $t \to 0^+$, obtemos $\langle v, x - \overline{x}\rangle \leqslant 0$. □

Definição 5. *Seja* C *um subconjunto fechado de* $\mathbb{R}^n$*. A projeção em* C*, denotada por* P_C*, é a aplicação ponto-conjunto* $P_C : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *definida por* $P_C(x) := \operatorname{argmin}\{\|x - z\| : z \in C\}$.

Proposição 6. *Seja* $C \subset \mathbb{R}^n$ *um conjunto fechado não-vazio. Então a projeção de* y *sobre* C *existe para todo* $y \in \mathbb{R}^n$.

Demonstração. Veja página 11 de [9]. □

O Teorema seguinte mostra que se C é convexo e fechado, então a projeção é única.

Teorema 1. (Teorema da Projeção)

Seja $C \subset \mathbb{R}^n$ *um conjunto convexo e fechado.*

Então para todo $x \in \mathbb{R}^n$*, a projeção de* x *sobre* C*,* $P_C(x)$*, existe e é única.*

Além disso, $\bar{x} = P_C(x)$ *se, e somente se,*

$$\bar{x} \in C, \quad \langle x - \bar{x}, y - \bar{x} \rangle \leqslant 0 \quad \forall y \in C,$$

ou, equivalentemente,

$$\bar{x} \in C, \quad x - \bar{x} \in N_C(x).$$

Demonstração. Veja página 102 de [9]. □

Corolário 2. (Operador de projeção é não-expansivo)

Seja $C \subset \mathbb{R}^n$ *um conjunto convexo e fechado.*

Então para $x \in \mathbb{R}^n$ *e* $y \in \mathbb{R}^n$ *quaisquer,*

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

Demonstração. Veja página 106 de [9]. □

1.2 Prox-regularidade

Definição 6. *Um subconjunto fechado* $F \subset \mathbb{R}^n$ *é dito prox-regular se o operador projeção* P_F *é um único valor em torno de cada ponto* x *em* F.

Exemplo 6. *Segue do Teorema 1 que os conjuntos convexos são prox-regulares.*

Exemplo 7. *O conjunto* $D = \{x \in \mathbb{R}^2_+ : x_1 + x_2 = 1\}$ *com a norma* $\|x\|_1 = |x_1| + |x_2|$ *não é prox-regular. De fato,* $d(0, x) = 1$, $\forall x \in D$*. Logo uma projeção de* 0 *sobre* D *em relação a esta norma é qualquer ponto de* D.

A seguir apresentaremos um importante teorema de Análise Variacional.

Teorema 2. [12] *Seja* F *um conjunto fechado prox-regular de* $\mathbb{R}^n$*. Seja* $g : \mathbb{R}^n \to \mathbb{R}$ *dada por* $g(x) = \frac{1}{2}d(x, F)^2$*. Então para cada* $\overline{x} \in F$ *existe* $r > 0$ *tal que:*

(a) *A projeção* P_F *é um único valor na* $B(\overline{x}, r)$*;*

(b) *a função* g *é* C^1 *na* $B(\overline{x}, r)$ *e* $\nabla g(x) = x - P_F(x)$*;*

(c) *a aplicação gradiente* ∇g *é* 1*-Lipschitz contínua na* $B(\overline{x}, r)$.

O ítem (c) não está explícito em [12], entretanto, ele está provado em [14].

1.3 Conjuntos algébricos e semi-algébricos

Definição 7. *Um conjunto* $A \subset \mathbb{R}^n$ *é dito algébrico, quando existe uma função polinomial* $f : \mathbb{R}^n \to \mathbb{R}$*, tal que,* $A = \{x \in \mathbb{R}^n : f(x) = 0\}$*, ou seja,* $A = f^{-1}(0)$.

Exemplo 8. *Seja* $\mathbb{S}^n = \left\{x \in \mathbb{R}^{n+1} : \sum_{i=1}^{n+1} x_i^2 = 1\right\}$*. Então, se tomarmos* $f(x) = \sum_{i=1}^{n+1} x_i^2 - 1$*, temos que* $\mathbb{S}^n = f^{-1}(0)$.

Proposição 7. *Se* A, B *subconjuntos algébricos de* $\mathbb{R}^n$*, então* $A \cup B$ *e* $A \cap B$ *são subconjuntos algébricos de* $\mathbb{R}^n$*.*

Demonstração. Como A, B são conjuntos algébricos, existem funções polinomialis f_1 e f_2 tais que, $A = f_1^{-1}(0)$ e $B = f_2^{-1}(0)$. O resultado segue observando que $A \cup B = (f_1 \cdot f_2)^{-1}(0)$, $A \cap B = (f_1^2 + f_2^2)^{-1}(0)$, e que as funções $(f_1 \cdot f_2)$ e $(f_1^2 + f_2^2)$ são polinomiais. $\square$

Proposição 8. *Se* B *é um subconjunto algébrico de* $\mathbb{R}^n$ *e* $F : \mathbb{R}^m \to \mathbb{R}^n$ *é uma aplicação polinomial, então* $F^{-1}(B)$ *é um subconjunto algébrico de* $\mathbb{R}^m$*.*

Demonstração. Como B é um subconjunto algébrico de $\mathbb{R}^n$, existe uma função polinomial $\varphi : \mathbb{R}^n \to \mathbb{R}$ tal que $B = \varphi^{-1}(0)$. Sabemos que $F^{-1}(B) = \{x \in \mathbb{R}^n : \exists y \in B, F(x) = y\}$. Logo, para todo $x \in F^{-1}(B)$, existe $y \in B$ tal que $\varphi(F(x)) = \varphi(y) = 0$, ou seja, $F^{-1}(B) = (\varphi \circ F)^{-1}(0)$. Como $\varphi \circ F$ é uma função polinomial, temos que $F^{-1}(B)$ é um subconjunto algébrico de $\mathbb{R}^m$. $\square$

Entretanto, a imagem de um conjunto algébrico por uma aplicação polinomial nem sempre é um conjunto algébrico. Veja o exemplo a seguir.

Exemplo 9. *Sejam* $\pi : \mathbb{R}^2 \to \mathbb{R}$ *a projeção* $\pi(x, y) = x$ *e* $\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$. *Vimos no Exemplo 8 que* $\mathbb{S}^1$ *é um conjunto algébrico, mas* $\pi(\mathbb{S}^1) = [-1, 1]$ *não é um conjunto algébrico. De fato, para todo polinômio* $f : \mathbb{R} \to \mathbb{R}$, *temos duas possibilidades para* $f^{-1}(0)$. *Se* f *não for o polinômio identicamente nulo, temos que* $f^{-1}(0)$ *é o conjunto finito de raízes do polinômio. Caso contrário,* $f^{-1}(0) = \mathbb{R}$. *Logo não exite um polinômio* $f : \mathbb{R} \to \mathbb{R}$, *tal que,* $f^{-1}(0) = [-1, 1]$. *Portanto, o intervalo* $[-1, 1]$ *não é um conjunto algébrico.*

A seguir, definiremos uma classe mais geral de conjuntos algébricos, a saber, a classe dos conjuntos semi-algébricos.

Definição 8. *Um subconjunto* S *de* $\mathbb{R}^n$ *é dito semi-algébrico se existe um número finito de funções polinômiais reais* $P_{ij}, Q_{ij} : \mathbb{R}^n \to \mathbb{R}$ *tais que*

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

Definição 9. *Uma função* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *é dita semi-algébrica se o seu gráfico, isto é,* Graf $f = \{(x, \lambda) \in \mathbb{R}^{n+1} : f(x) = \lambda\}$, *é um subconjunto semi-algébrico de* $\mathbb{R}^{n+1}$. *Analogamente, uma aplicação ponto-conjunto* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *é uma aplicação semi-algébrica se* Graf $F = \{(x, y) \in \mathbb{R}^{n+m} : y \in F(x)\}$ *é um subconjunto semi-algébrico de* $\mathbb{R}^{n+m}$.

Definição 10. *Sejam* $S_1 \subset \mathbb{R}^m$ *e* $S_2 \subset \mathbb{R}^n$ *conjuntos semi-algébricos. Uma aplicação* $f : S_1 \to S_2$ *dada por* $f = (f_1, ..., f_n)$ *é semi-algébrica se* f_i *é uma função semi-algébrica* $\forall i \in \{1, ..., n\}$.

Observe que todo conjunto algébrico é um conjunto semi-algébrico. De fato, se A é algébrico, existe uma função polinomial $f : \mathbb{R}^n \to \mathbb{R}$ tal que $A = \{x \in \mathbb{R}^n : f(x) = 0\}$. Daí, basta notar que se $g_1(x) = f(x)$ e $g_2(x) = -f(x)$, então

$$\begin{aligned} A &= \{x \in \mathbb{R}^n : f(x) = 0\} \\ &= \{x \in \mathbb{R}^n : f(x) = 0\} \cup \Big(\bigcap_{i=1}^{2} \{x \in \mathbb{R}^n : g_i(x) < 0\} \Big). \end{aligned}$$

Exemplo 10. *Toda aplicação polinomial* $p : \mathbb{R}^m \to \mathbb{R}^n$ *é uma aplicação semi-algébrica. De fato,*

$$\text{Graf } p = \{(x, y) \in \mathbb{R}^m \times \mathbb{R}^n : y = p(x)\} = q^{-1}(0)$$

onde $q(x, y) = y - p(x)$ *é uma função polinomial.*

Proposição 9. *Na reta, um conjunto semi-algébrico é a reunião finita de intervalos abertos, fechados, semi-abertos e pontos.*

Demonstração. Veja página 25 de [5]. □

Teorema 3. (Tarski-Seidenberg). *Considere a aplicação projeção* $\pi : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m$ *definida por* $\pi(x, y) = x$. *Então, para todo subconjunto semi-algébrico* A *de* $\mathbb{R}^{m+n}$, $\pi(A)$ *é um subconjunto semi-algébrico de* $\mathbb{R}^m$.

Demonstração. Veja página 26 de [5]. □

Vejamos algumas propriedades de conjuntos semi-algébricos:

(i) Seja $A \subset \mathbb{R}^n$ um conjunto semi-algébrico. Então os conjuntos $\overline{A}$, int$A$ e ∂A são conjuntos semi-algébricos;

(ii) Sejam $A, B \subset \mathbb{R}^n$ conjuntos semi-algébricos. Então, $A \cup B$, $A \cap B$, $A - B$, $B - A$, A^c e B^c são conjuntos semi-algébricos;

(iii) Sejam $A \subset \mathbb{R}^n$ e $B \subset \mathbb{R}^m$ conjuntos semi-algébricos. Então $A \times B \subset \mathbb{R}^n \times \mathbb{R}^m$ é um conjunto semi-algébrico.

Corolário 3. *Sejam* $A \subset \mathbb{R}^m$ *e* $B \subset \mathbb{R}^n$ *conjuntos semi-algébricos e* $f : A \to B$ *uma aplicação semi-algébrica. Se* $S_1 \subset A$ *é um conjunto semi-algébrico, então* $f(S_1)$ *é um conjunto semi-algébrico. Além disso, se* $S_2 \subset B$ *é um conjunto semi-algébrico, então* $f^{-1}(S_2)$ *é um conjunto semi-algébrico.*

Demonstração. Note que Graf $f|_{S_1} = (S_1 \times B) \cap (\text{Graf } f)$ é um conjunto semi-algébrico. Definindo $\pi_2 : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$, $\pi_2(x, y) = y$, obtemos $\pi_2(\text{Graf } f|_{S_1}) = f(S_1)$. Logo, pelo Teorema 3, $f(S_1)$ é um conjunto semi-algébrico. Para $f^{-1}(S_2)$, basta notar que $f^{-1}(S_2) = \pi_1((A \times S_2) \cap (\text{Graf } f))$, onde π_1 é a projeção na primeira coordenada. □

Corolário 4. *A imagem de um conjunto semi-algébrico por uma aplicação polinomial é um conjunto semi-algébrico.*

Demonstração. Segue do Corolário 3 e do Exemplo 10. □

Proposição 10. *Sejam* $A \subset \mathbb{R}^p$ *e* $B \subset \mathbb{R}^n$ *conjuntos semi-algébricos e* $f : A \to \mathbb{R}^n$ *e* $g : B \to \mathbb{R}^m$ *aplicações semi-algébricas tal que* $f(A) \subset B$. *Então* $g \circ f : A \subset \mathbb{R}^p \to \mathbb{R}^m$ *é uma aplicação semi-algébrica.*

Demonstração. Por hipótese, Graf f é um conjunto semi-algébrico. Por outro lado, o conjunto $f(A) \times g(f(A)) = \{(f(x), g(f(x))) \in \mathbb{R}^n \times \mathbb{R}^m : x \in A\}$ também é semi-algébrico. Daí o conjunto

$$\begin{aligned} \Omega &= (\text{Graf } f) \times f(A) \times g(f(A)) \\ &= \{(x, f(x), f(x), g(f(x))) \in \mathbb{R}^p \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m : x \in A\} \end{aligned}$$

é semi-algébrico. Tomando a projeção $\pi : \mathbb{R}^p \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p \times \mathbb{R}^m$ pondo $\pi(x, y, z, w) = (x, w)$, segue que $\pi(\Omega) = \{(x, g(f(x))) \in \mathbb{R}^p \times \mathbb{R}^m : x \in A\} = \text{Graf } g \circ f$ é um conjunto semi-algébrico. Portanto $g \circ f$ é uma aplicação semi-algébrica. □

Proposição 11. *Sejam* S *um subconjunto semi-algébrico não-vazio de* $\mathbb{R}^m$, $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *uma função polinomial real e* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *definida por*

$$f(x) = \sup\{g(x, y) : y \in S\}.$$

Então f *é uma função semi-algébrica.*

Demonstração. Para isso, basta mostrarmos que

$$\text{Graf } f = \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R} : \forall y \in S, g(x, y) \leqslant \lambda\} \cap \{(x, \mu) \in \mathbb{R}^n \times \mathbb{R} : f(x) \geqslant \mu\}$$

é um conjunto semi-algébrico. De fato, usando as propriedades dos conjuntos semi-algébricos, vemos que

$$\begin{aligned} \Omega &= \{(x, \lambda, y) \in \mathbb{R}^n \times \mathbb{R} \times S : g(x, y) > \lambda\} \\ &= \{(x, \lambda, y) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m : g(x, y) > \lambda\} \cap (\mathbb{R}^n \times \mathbb{R} \times S) \end{aligned}$$

é um conjunto semi-algébrico. Defina

$$\pi : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R} \text{ pondo}, \pi(x, \lambda, y) = (x, \lambda).$$

Pelo Teorema 3, o conjunto $\pi(\Omega) = \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R} : \text{existe } y \in S, g(x, y) > \lambda\}$ é semi-algébrico, logo o seu complementar $\pi(\Omega)^C = \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R} : \forall\, y \in S, g(x, y) \leqslant \lambda\}$, goza da mesma propriedade. Analogamente, mostra-se que o conjunto

$$\{(x, \mu) \in \mathbb{R}^n \times \mathbb{R} : \forall\, y \in S, g(x, y) \geqslant \mu\} = \{(x, \mu) \in \mathbb{R}^n \times \mathbb{R} : f(x) \geqslant \mu\}$$

é semi-algébrico. Portanto, Graf f é um conjunto semi-algébrico. □

Corolário 5. *Sejam* S *um subconjunto semi-algébrico não-vazio de* $\mathbb{R}^m$, $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *uma função polinomial real e* $f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty\}$ *definida por*

$$f(x) = \inf\{h(x, y) : y \in S\}.$$

Então f *é uma função semi-algébrica.*

Demonstração. Ponha $g(x, y) = -h(x, y)$. Daí,

$$f(x) = \inf\{h(x, y) : y \in S\} = \inf\{-g(x, y) : y \in S\} = -\sup\{g(x, y) : y \in S\}.$$

Logo f é uma função semi-algébrica. □

Exemplo 11. *Se S é um subconjunto semi-algébrico não-vazio de $\mathbb{R}^m$, então a função*

$$\psi : \mathbb{R}^m \to \mathbb{R}, \text{ onde } \psi(x) = \text{dist}(x, S)^2$$

é semi-algébrica. De fato, basta aplicarmos a Corolário 5 é função polinomial $g(x, y) = \|x - y\|^2$. Daí, $\psi(x) = \inf\{\|x - y\|^2 : y \in S\}$ é uma função semi-algébrica.

Exemplo 12. *Sejam $F_1, ..., F_p$ conjuntos semi-algébricos não-vazios de $\mathbb{R}^n$. Então a função $f : \mathbb{R}^n \to [0, +\infty)$, dada por*

$$f(x) = \tfrac{1}{2} \textstyle\sum_{i=1}^{p} \text{dist}(x, F_i)^2$$

é semi-algébrica. Com efeito, pelo Exemplo 11, a aplicação $\xi(x) = (\text{dist}(x, F_1)^2, ..., \text{dist}(x, F_i)^2)$ é semi-algébrica. Logo $f(x) = (h \circ \xi)(x)$, onde $h(x) = \dfrac{x_1 + \cdots + x_n}{2}$. Aplicando a Proposição 10, obtemos o resultado.

1.4 Conjuntos analíticos, semi-analíticos e subanalíticos

Definição 11. *Um conjunto $A \subset \mathbb{R}^n$ é dito analítico se para todo $x \in A$ existe uma vizinhança $U_x \subset \mathbb{R}^n$ de x e uma função analítica $f : U_x \to \mathbb{R}$, tal que, $A \cap U_x = f^{-1}(0)$.*

Exemplo 13. *Seja $f : \mathbb{R}^2 \to \mathbb{R}$ definida por $f(x, y) = e^x + \sin x^2 + y^3$. Então o conjunto $A = \{(x, y) \in \mathbb{R}^2 : f(x, y) = 0\}$ é um conjunto analítico.*

Claramente, todo polinômio é uma aplicação analítica, logo todo conjunto algébrico é um conjunto analítico.

Definição 12. *Um conjunto $X \subset \mathbb{R}^n$ é dito semi-analítico básico, se para todo $x \in X$ existe uma vizinhança $U_x \subset \mathbb{R}^n$ de x e funções $f, g_1, ..., g_k$, analíticas em U_x tal que*

$$X \cap U_x = \{x \in \mathbb{R}^n : f(x) = 0\} \cap \left(\bigcap_{i=1}^{k} \{x \in \mathbb{R}^n : g_i(x) > 0\} \right)$$

Definição 13. *Um conjunto semi-analítico é a reunião finita de conjuntos semi-analíticos básicos.*

Observação 3. *Todo polinômio é uma aplicação analítica, logo todo conjunto algébrico é um conjunto analítico. Da mesma forma que todo conjunto algébrico é semi-algébrico,* *fica claro que todo conjunto analítico é semi-analítico. E como todo polinômio é analítico, temos que, todo conjunto semi-algébrico é um conjunto semi-analítico.*

Definição 14. *Seja* $X \subset \mathbb{R}^n$. *Dizemos que a função* $F : X \to \mathbb{R}$ *é semi-analítica se seu gráfico* Graf $F = \{(x, F(x)) : x \in X\} \subset \mathbb{R}^{n+1}$ *é um conjunto semi-analítico.*

Definição 15. *Sejam* $X \subset \mathbb{R}^n$ *e* $Y \subset \mathbb{R}^m$. *Uma aplicação* $F : X \to Y$ *é semi-analítica se suas funções coordenadas são semi-analíticas. Equivalentemente, se seu gráfico* Graf $F \subset \mathbb{R}^{n+m}$ *é um conjunto semi-analítico.*

Observação 4. *O Teorema de Tarski-Seidenberg não é válido para conjuntos semi-analíticos, veja Exemplo 2.14 de* [4].

Definição 16. *Um conjunto* $A \subset \mathbb{R}^n$ *é dito subanalítico se existe um conjunto* $\widetilde{A} \subset \mathbb{R}^m$, *semi-analítico,* $m \geqslant n$, *tal que a projeção* $\pi : \mathbb{R}^m \to \mathbb{R}^n$ *restrita a* $\widetilde{A}$ *é uma aplicação própria e* $A = \pi(\widetilde{A})$.

Com esta definição torna-se naturalmente válido o Teorema de Tarski-Seidenberg. De fato, Sejam $A \subset \mathbb{R}^n$ um conjunto subanalítico e $\pi : \mathbb{R}^n \to \mathbb{R}^k$, $n \geqslant k$, uma projeção. Sabemos que existe um conjunto semi-analítico $\widetilde{A} \subset \mathbb{R}^m$, $m \geqslant n$, tal que a projeção $\widetilde{\pi} : \mathbb{R}^m \to \mathbb{R}^n$ restrita a $\widetilde{A}$ é uma aplicação própria e $A = \widetilde{\pi}(\widetilde{A})$. Logo, $\pi \circ \widetilde{\pi}$ define $B = \pi(A)$ como conjunto subanalítico.

Assim definido, segue que todo conjunto semi-analítico é subanalítico.

Em suma, temos as seguintes inclusões:

Como antes temos as seguintes definições:

Definição 17. *Seja* $A \subset \mathbb{R}^n$. *Dizemos que uma função* $F : A \to \mathbb{R}$ *é subanalítica se seu gráfico* Graf $F = \{(x, F(x)) : x \in A\} \subset \mathbb{R}^{n+1}$ *é um conjunto subanalítico.*

Definição 18. *Sejam* $A \subset \mathbb{R}^n$ *e* $B \subset \mathbb{R}^m$. *Uma aplicação* $F : A \to B$ *é subanalítica se suas funções componentes são subanalíticas ou equivalentemente, se seu gráfico* Graf $F \subset \mathbb{R}^{n+m}$ *é um conjunto subanalítico.*

Vejamos algumas propriedades de conjuntos subanalíticos:

(i) Se $A, B \subset \mathbb{R}^n$ são conjuntos subanalíticos, então $A \cup B$ é um conjunto subanalítico;

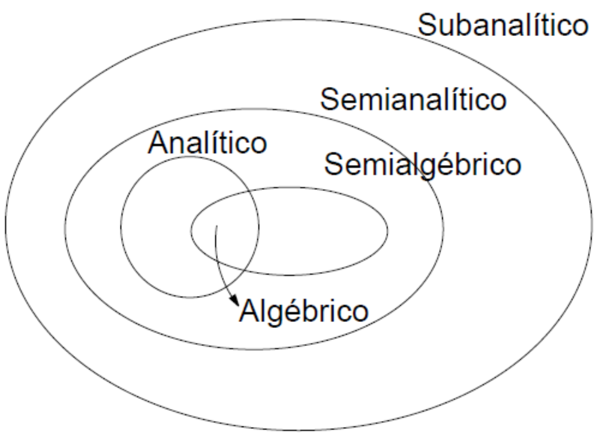

Figura 1.1: Diagrama 1

(ii) (Teorema de Gabrielov) Se $A \subset \mathbb{R}^n$ é um conjunto subanalítico, então A^c é um conjunto subanalítico; (veja Teorema 3.10 de [4])

Como consequência de (i) e (ii), temos que a interseção e a diferença de conjuntos subanalíticos é um conjunto subanalítico.

(iii) Se $F : A \subset \mathbb{R}^m \to \mathbb{R}^n$ é uma aplicação subanalítica, então A e $F(A)$ são conjuntos subanalíticos;

(iv) Se $A \subset \mathbb{R}$ é um conjunto subanalítico, então A é reunião finita de intervalos abertos, fechados, semiabertos e pontos;

(v) Composta se funções subanalíticas ainda é uma função subanalítica;

(vi) Se $A \subset \mathbb{R}$ é um conjunto subanalítico, então $\overline{A}$, int$A$ e ∂A são conjuntos subanalíticos.

Para uma abordagem mais completa sobre conjuntos semi-anlíticos e subanalíticos veja [4].

1.5 Estrutura o-minimal

Definição 19. [11] *Seja* $\mathcal{M} = \bigcup\limits_{n \in \mathbb{N}} \mathcal{M}_n$, *onde cada* $\mathcal{M}_n$ *é uma família de subconjuntos de* $\mathbb{R}^n$. *Dizemos que a coleção* $\mathcal{M}$ *é uma estrutura o-minimal em* $(\mathbb{R}, +, \cdot)$ *se:*

(1) *cada* $\mathcal{M}_n$ *é uma álgebra booleana, ou seja,* $\emptyset \in \mathcal{M}_n$ *e para cada* $A, B \in \mathcal{M}_n$*,* $A \cup B$*,* $A \cap B$*, e* $\mathbb{R}^n \backslash A$ *pertencem a* $\mathcal{M}_n$*;*

(2) *se* $A \in \mathcal{M}_n$ *e* $B \in \mathcal{M}_m$*, então* $A \times B \in \mathcal{M}_{n+m}$*;*

(3) *se* $A \in \mathcal{M}_{n+m}$ *e* $\pi : \mathbb{R}^{n+m} \to \mathbb{R}^n$ *é a projeção das* n *primeiras coordenadas, então* $\pi(A) \in \mathcal{M}_n$*;*

(4) *se* $f, g_1, g_2, ..., g_k \in \mathbb{Q}[X_1, ..., X_n]$*, então* $\{x \in \mathbb{R}^n : f(x) = 0, g_1(x) > 0, ..., g_k(x) > 0\} \in \mathcal{M}_n$;

(5) $\mathcal{M}_1$ *consiste de todas as uniões de intervalos abertos e pontos.*

Definição 20. *Para uma estrutura o-minimal* $\mathcal{M}$ *fixada em* $(\mathbb{R}, +, \cdot)$ *dizemos que* A *é um* $\mathcal{M}$*-conjunto se* $A \in \mathcal{M}_n$ *para algum* $n \in \mathbb{N}$.

Definição 21. *Dizemos que* $f : A \to \mathbb{R}$*, onde* $A \subset \mathbb{R}^n$*, é uma* $\mathcal{M}$*-função se o gráfico de* f *é um* $\mathcal{M}$*-conjunto.*

Definição 22. *Uma função* $f : \mathbb{R}^n \to \mathbb{R}$ *é dita ser* ***definível*** *ou* ***definable*** *em* $\mathcal{M}$ *se seu gráfico pertence a* $\mathcal{M}_{n+1}$*. Além disso,* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *é dita definível em* $\mathcal{M}$ *se a imagem inversa* $f^{-1}(+\infty)$ *é um subconjunto definível de* $\mathbb{R}^n$ *e* f *restrita à seu domínio efetivo é uma função definível.*

Definição 23. *Dizemos que* $f : A \to \mathbb{R}^m$*, onde* $A \subset \mathbb{R}^n$*, é uma* $\mathcal{M}$*-aplicação se o gráfico de* f *é um* $\mathcal{M}$*-conjunto, ou equivalentemente, se* f_i *são* $\mathcal{M}$*-funções para cada* $i = 1, ..., m$*.*

Claramente os conjuntos semi-algébricos e subanalíticos são $\mathcal{M}$-conjuntos, e as funções semi-algébricas e subanalíticas são $\mathcal{M}$-funções.

Proposição 12. *Sejam* $A \subset \mathbb{R}^m$ *e* $B \subset \mathbb{R}^n$ $\mathcal{M}$*-conjuntos e* $f : A \to B$ *uma* $\mathcal{M}$*-aplicação. Se* $S_1 \subset A$ *é um* $\mathcal{M}$*-conjunto, então* $f(S_1)$ *é um* $\mathcal{M}$*-conjunto. Além disso, se* $S_2 \subset B$ *é um* $\mathcal{M}$*-conjunto, então* $f^{-1}(S_2)$ *é um* $\mathcal{M}$*-conjunto.*

Demonstração. Note que Graf $f|_{S_1} = (S_1 \times B) \cap (\text{Graf } f)$ é um $\mathcal{M}$-conjunto. Definindo $\pi_2 : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$, $\pi_2(x, y) = y$, obtemos $\pi_2(\text{Graf } f|_{S_1}) = f(S_1)$. Logo, $f(S_1)$ é um $\mathcal{M}$-conjunto. Para $f^{-1}(S_2)$, basta notar que $f^{-1}(S_2) = \pi_1((A \times S_2) \cap (\text{Graf } f))$, onde π_1 é a projeção na primeira coordenada. $\square$

Proposição 13. *Sejam* $A \subset \mathbb{R}^p$ *e* $B \subset \mathbb{R}^n$ $\mathcal{M}$-*conjuntos e* $f : A \to \mathbb{R}^n$ *e* $g : B \to \mathbb{R}^m$ $\mathcal{M}$-*aplicação tal que* $f(A) \subset B$. *Então* $g \circ f : A \subset \mathbb{R}^p \to \mathbb{R}^m$ *é uma* $\mathcal{M}$-*aplicação.*

Demonstração. Por hipótese, Graf f é um $\mathcal{M}$-conjunto. Por outro lado, o conjunto $f(A) \times g(f(A)) = \{(f(x), g(f(x))) \in \mathbb{R}^n \times \mathbb{R}^m : x \in A\}$ também é um $\mathcal{M}$-conjunto. Daí o conjunto

$$\begin{aligned}\Omega &= (\text{Graf } f) \times f(A) \times g(f(A)) \\ &= \{(x, f(x), f(x), g(f(x))) \in \mathbb{R}^p \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m : x \in A\}\end{aligned}$$

é um $\mathcal{M}$-conjunto. Tomando a projeção $\pi : \mathbb{R}^p \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p \times \mathbb{R}^m$ pondo $\pi(x, y, z, w) = (x, w)$, segue que $\pi(\Omega) = \{(x, g(f(x))) \in \mathbb{R}^p \times \mathbb{R}^m : x \in A\} =$ Graf $g \circ f$ é um $\mathcal{M}$-conjunto. Portanto $g \circ f$ é uma $\mathcal{M}$-aplicação. $\square$

A seguir enunciaremos alguns resultados básicos da estrutura o-minimal, cujas respectivas demonstrações podem ser encontradas em [11].

Lema 1. *Sejam* $G : A \to \mathbb{R}^m$ *uma* $\mathcal{M}$-*aplicação e* $f : A \to \mathbb{R}$ *uma* $\mathcal{M}$-*funão tal que* $f(x) \geqslant 0$ *para todo* $x \in A$. *Defina* $\varphi : G(A) \to \mathbb{R}$ *por*

$$\varphi(y) = \inf_{x \in G^{-1}(y)} f(x).$$

Então φ *é uma* $\mathcal{M}$-*função.*

Corolário 6. *Seja* A *um* $\mathcal{M}$-*conjunto em* $\mathbb{R}^n$. *Então* $\operatorname{dist} : \mathbb{R}^n \to \mathbb{R}$ *é uma* $\mathcal{M}$-*função, onde* $\operatorname{dist}(x) = \inf\limits_{y \in A} \|x - y\|$.

Corolário 7. *Seja* A *uma* $\mathcal{M}$-*conjunto em* $\mathbb{R}^n$. *Então o fecho e o interior de* A *são* $\mathcal{M}$-*conjuntos.*

Lema 2. (Lema da Monotonicidade)
Seja $f : (a, b) \to \mathbb{R}$ *uma* $\mathcal{M}$-*função. Então existem números reais* $a = a_0 < a_1 < ... < a_k = b$ *tal que* f *é continuamente diferenciável em cada intervalo* (a_i, a_{i+1}). *Além disso,* f' *é uma* $\mathcal{M}$-*função e a função* f *é estritamente monótona ou constante em cada intervalo* (a_i, a_{i+1}).

Lema 3. *Seja* $f : U \to \mathbb{R}^k$ *uma* $\mathcal{M}$-*função diferenciável onde* U *é uma aberto em* $\mathbb{R}^n$. *Então* $\partial f / \partial x_j$, $j = 1, ..., n$ *são* $\mathcal{M}$-*funções, e daí* ∇f *é uma* $\mathcal{M}$-*aplicação.*

Lema 4. (Lema de Seleção de Curva)

Seja A *um* $\mathcal{M}$*-conjunto em* $\mathbb{R}^n$ *e suponha que* $a \in \overline{A\backslash\{a\}}$*. Então existe uma* $\mathcal{M}$*-curva* $\gamma : [0, \varepsilon) \to \mathbb{R}^n$ *de classe* C^1 *em* $[0, \varepsilon)$ *e tal que* $\gamma(0) = a$ *e* $\gamma((0, \varepsilon)) \subset A\backslash\{a\}$*.*

Exemplo 14. *Se* $f : A \to \mathbb{R}$ *é uma* $\mathcal{M}$*-função diferenciável, então*

$$g(x) = \|\nabla f(x)\| - (f(x))^2$$

é uma $\mathcal{M}$*-função.*

Exemplo 15. *Seja* $f : \mathbb{R} \to \mathbb{R}$ *definida por*

$$f(x) = \begin{cases} x^2 \sin \dfrac{1}{x}, & \textit{se} \quad x \neq 0, \\ 0, & \textit{se} \quad x = 0. \end{cases}$$

Se f *fosse uma* $\mathcal{M}$*-função teríamos pelo Lema 2 que* f' *também seria uma* $\mathcal{M}$*-função, logo* $A = \{x \in \mathbb{R} : f'(x) = 0\} = f'^{-1}(0)$ *seria um* $\mathcal{M}$*-conjunto, o que contradiz a Definição 19 ítem (5).*

Apresentaremos agora o Teorema de Pouiseux, que será muito útil no próximo capítulo.

Teorema 4. *Seja* $f : [0, \delta) \to \mathbb{R}$ *uma função subanalítica e contínua. Então existem* $\varepsilon > 0$*,* $k \in \mathbb{N}$ *e* $h(t) = \sum_{i=0}^{\infty} \alpha_i t^i$ *uma função analítica em uma vizinhança de* $0 \in \mathbb{R}$ *tal que* $f(t) = h(t^{\frac{1}{k}})$*, para* $t \in [0, \varepsilon)$*. Portanto, há um desenvolvimento de* f *em série de Pouiseux* $f(t) = \sum_{i=0}^{\infty} \alpha_i t^{\frac{i}{k}}$ *uniformemente convergente.*

Demonstração. Veja página 143 de [10]. $\square$

Capítulo 2

Funções KL

Seja U um subconjunto aberto de $\mathbb{R}^n$. Denotaremos por $C^1(U;\mathbb{R})$ o conjunto das funções $f : U \to \mathbb{R}$ de classe C^1.

Definição 24. [7] *Dizemos que uma função* $f \in C^1(U;\mathbb{R})$ *satisfaz a desigualdade de Łojasiewicz em torno de* $a \in U$ *se existem constantes* $\theta \in (0,1]$, $\delta, c > 0$ *tal que para todo* $x \in U$, $\|x - a\| < \delta$,

$$|f(x) - f(a)|^{1-\theta} \leqslant c\|\nabla f(x)\|. \tag{2.1}$$

O número θ *será chamado de expoente de Łojasiewicz.*

Teorema 5. (Desigualdade de Łojasiewicz. Caso n=1)
Seja f *uma função analítica real em uma vizinhança de um ponto* $a \in \mathbb{R}$ *tal que* $f'(a) = 0$. *Então* f *satisfaz a desigualdade de Łojasiewicz em alguma vizinhança de* a.

Demonstração. Sem perda de generalidade, podemos supor que $a = 0$ e que $f(a) = 0$. Como f é analítica em uma vizinhança da origem, podemos escrever

$$f(x) = \sum_{k=0}^{\infty} a_k x^k, \tag{2.2}$$

e consequentemente,

$$f'(x) = \sum_{k=1}^{\infty} k a_k x^{k-1}. \tag{2.3}$$

Afirmamos que existem constantes $k_0 \in \mathbb{N}, \delta > 0, c_1, c_2$ tal que, se $|x| < \delta$, então

$$|f(x)| \leqslant c_1 |x|^{k_0} \tag{2.4}$$

e

$$|f'(x)| \geqslant c_2 |x|^{k_0 - 1}. \tag{2.5}$$

Mostraremos inicialmente (2.4). Defina $A = \{k \in \mathbb{N} : a_k \neq 0\}$. Seja k_0 o menor elemento de A. Segue de (2.2) que,

$$\frac{f(x)}{x^{k_0}} = a_{k_0} + s(x),$$

onde, $s(x) = a_{k_0+1}x + a_{k_0+2}x^2 + \cdots$. Como $\lim_{x\to 0} s(x) = 0$, existe $\delta_1 > 0$ tal que para $|x| < \delta_1$, tem-se $|s(x)| \leqslant 1$. Logo, para $|x| < \delta_1$,

$$\frac{|f(x)|}{|x|^{k_0}} \leqslant |a_{k_0}| + 1 = c_1,$$

ou seja,

$$|f(x)| \leqslant c_1 |x|^{k_0}.$$

Vamos à prova de (2.5).

Usando (2.3) obtemos,

$$f'(x) = g(x) + r(x),$$

onde,

$$g(x) = k_0 a_{k_0} x^{k_0 - 1} \quad \text{e} \quad r(x) = (k_0 + 1) a_{k_0+1} x^{k_0} + \cdots .$$

Como $\lim_{x\to 0} r(x)/x^{k_0-1} = 0$, existe $\delta_2 > 0$ tal que para $|x| < \delta_2$, $|r(x)| \leqslant \dfrac{k_0|a_{k_0}|}{2}|x|^{k_0-1}$. Logo, para $|x| < \delta_2$,

$$|f'(x)| \geqslant |g(x)| - |r(x)| \geqslant k_0|a_{k_0}||x|^{k_0-1} - \frac{k_0|a_{k_0}|}{2}|x|^{k_0-1} = c_2|x|^{k_0-1}, \ c_2 = \frac{k_0|a_{k_0}|}{2}.$$

Tomando $\delta = \min\{\delta_1, \delta_2\}$, tem-se que para $|x| < \delta$, vale (2.4) e (2.5) simultaneamente. Daí,

$$\left(\frac{1}{c_1}\right)^{\frac{1}{k_0}} |f(x)|^{\frac{1}{k_0}} \leqslant |x| \leqslant \left(\frac{1}{c_2}\right)^{\frac{1}{k_0-1}} |f'(x)|^{\frac{1}{k_0-1}} \Rightarrow c_2 \left(\frac{1}{c_1}\right)^{\frac{k_0-1}{k_0}} |f(x)|^{\frac{k_0-1}{k_0}} \leqslant |f'(x)|.$$

O resultado segue tomando $c = \dfrac{c_1^{1-\frac{1}{k_0}}}{c_2}$ e $\theta = 1/k_0$. □

Funções C^∞, em geral, não verificam a desigualdade de Łojasiewicz, como mostra o

Exemplo 16. *Seja* $f : \mathbb{R} \to \mathbb{R}$ *definida por*

$$f(x) := \begin{cases} e^{-1/x^2} & , \ se \quad x \neq 0 \\ 0 & , \ se \quad x = 0 \end{cases} .$$

Afirmamos que f *é uma função* C^∞ *que não satisfaz a desigualdade de Łojasiewicz em torno da origem. De fato, se* $x \neq 0$ *existe* $f^{(n)}(x)$ *para todo* $n \in \mathbb{N}$. *Resta mostrar que existe* $f^{(n)}(0)$ *para todo* n. *Com efeito, se* $x \neq 0$, $f^{(n)}(x) = p\left(\dfrac{1}{x}\right)e^{-1/x^2}$, *onde* p *é um* *polinômio. Ponha* $y = 1/x$*, e suponha por indução que para todo* $f^{(n-1)}(0) = 0$*. Segue que,*

$$f^{(n)}(0) = \lim_{x\to 0} \frac{f^{(n-1)}(x) - f^{(n-1)}(0)}{x - 0} = \lim_{x\to 0} \widetilde{p}\Big(\frac{1}{x}\Big)e^{-1/x^2} = \lim_{y\to\pm\infty} p(y)/e^{y^2} = 0.$$

Logo, $f \in C^\infty$.

Suponha agora que $x > 0$ *e que existam contantes* $c > 0$ *e* $0 < \theta \leqslant 1$ *tal que* $|\nabla f(x)| \geqslant c|f(x)|^{1-\theta}$*. Daí,*

$$c \geqslant \frac{|f(x)|^{1-\theta}}{|\nabla f(x)|} = \frac{e^{\theta/x^2}}{1/2x^3}.$$

Assim,

$$c \geqslant \lim_{x\to 0^+} \frac{e^{\theta/x^2}}{1/2x^3} = +\infty,$$

o que é uma absurdo.

Lema 5. *Sejam* $U \subset \mathbb{R}^n$ *e* $f \in C^1(U;\mathbb{R})$ *satisfazendo a desigualdade de Łojasiewicz em torno de* $a \in U$ *com expoente* $\theta \in (0,1]$*. Então:*

(a) *A função* f *satisfaz a desigualdade de Łojasiewicz em torno de* a *para todo expoente* $\theta' \in (0,\theta]$*;*

(b) *Se* $n = 1$*, então o ponto* a*, ou pertence ao interior de* $f^{-1}(\{f(a)\})$*, ou é um ponto isolado da fronteira de* $f^{-1}(\{f(a)\})$*. Como consequência, ou* $f(x) \geqslant f(a)$ *para todo* $a \leqslant x \leqslant a+\delta$ *(resp.* $a-\delta \leqslant x \leqslant a$*), ou* $f(x) \leqslant f(a)$ *para todo* $a \leqslant x \leqslant a+\delta$ *(resp.* $a-\delta \leqslant x \leqslant a$*);*

(c) *Se* $n = 1$*, então existe* $\delta' \leqslant \delta$ *tal que a função* f *ou é constante, ou é estritamente crescente, ou é estritamente decrescente em* $[a, a+\delta']$ *(resp.* $[a-\delta', a]$*);*

(d) *Se* $n = 1$*, e se* f *é não constante em* $[a, a+\delta']$ *(resp.* $[a-\delta', a]$*), então existem* $c' > 0$*, tal que*

$$|f(x) - f(a)| \geqslant c'|x-a|^{\frac{1}{\theta}}$$

para todo $a \leqslant x \leqslant a+\delta'$ *(resp.* $a-\delta' \leqslant x \leqslant a$*).*

Demonstração. Como f satisfaz a desigualdade de Łojasiewicz em torno de $a \in U$ se existem constantes $\theta \in (0,1]$, $\delta, c > 0$ tal que para todo $x \in U$, $\|x - a\| < \delta$,

$$|f(x) - f(a)|^{1-\theta} \leqslant c\|\nabla f(x)\|. \tag{2.6}$$

(a) Para todo $\theta' \in (0,\theta]$, defina $g : \mathbb{R}^n \to \mathbb{R}$, $g(x) = (f(x) - f(a))^{\theta-\theta'}$. Como , $\lim_{x\to a} g(x) = 0$, existe $\delta_1 > 0$ tal que para $|x-a| < \delta_1$, $|f(x)-f(a)|^{1-\theta'} \leqslant |f(x)-f(a)|^{1-\theta}$.

Logo, para $|x - a| < \min\{\delta, \delta_1\}$, $x \in U$, temos

$$|f(x) - f(a)|^{1-\theta'} \leqslant c\|\nabla f(x)\|.$$

(b) Assuma que f satisfaz a desigualdade de Łojasiewicz em torno de a, que a não pertença ao interior de $f^{-1}(\{f(a)\})$. Suponha e que a não é ponto isolado da fronteira de $f^{-1}(\{f(a)\})$. Uma vez que o complemento de $f^{-1}(\{f(a)\})$ é um aberto em U, é a união enumerável de intervalos disjuntos $(x^k, y^k)_{k\in\mathbb{N}}$. Como a não é ponto isolado da fronteira de $f^{-1}(\{f(a)\})$, segue que existem sequências $(x^{k_j})_{j\in\mathbb{N}}$, $(y^{k_j})_{j\in\mathbb{N}}$ tais que, $\lim x^{k_j} = \lim y^{k_j} = a$ quando $j \to \infty$, $f(x^{k_j}) = f(y^{k_j}) = f(a)$ e $f(z) \neq f(a)$ para todo $z \in (x^{k_j}, y^{k_j})$, $j \in \mathbb{N}$. Pelo Teorema de Rolle existe $z^{k_j} \in (x^{k_j}, y^{k_j})$ tal que $f'(z^{k_j}) = 0$. Por (2.6) temos,

$$|f(z^{k_j}) - f(a)|^{1-\theta} \leqslant c|f'(z^{k_j})| = 0.$$

Isso mostra que $f(z^{k_j}) = f(a)$, o que é impossível.

(c) Segue diretamente de (b).

(d) Pelo ítem (c) existe existe $\delta' \leqslant \delta$ tal que a função f , ou é estritamente crescente, ou é estritamente decrescente em $[a, a+\delta']$. Suponhamos que $f(x) < f(a)$ em $[a, a+\delta']$. Daí,

$$\frac{1}{\theta}\frac{d}{dx}(f(a) - f(x))^{\theta} = -f'(x)(f(a) - f(x))^{\theta-1} = -\frac{f'(x)}{(f(a) - f(x))^{1-\theta}} \geqslant \frac{1}{c}.$$

Logo,

$$\frac{1}{\theta}(f(a) - f(x))^{\theta} \geqslant \int_a^x \frac{1}{c}ds = \frac{1}{c}(x - a),$$

o que implica em,

$$|f(x) - f(a)| \geqslant \frac{\theta}{c}|x - a|^{\frac{1}{\theta}}.$$

Os outros casos são similares. □

Proposição 14. *Sejam* $U \subset \mathbb{R}$ *um conjunto aberto e* $a \in U$.

(a) *Seja* $f \in C^1(U;\mathbb{R})$ *tal que* $f'(x) = g(x) + r(x)$, *onde* $|g(x)| = c|x - a|^{p-1}$, $p \in \mathbb{N}$, $c > 0$, *e* $r(x) = o(|x - a|^{p-1})$. *Então* f *satisfaz a desigualdade de Łojasiewicz em torno de* a *com expoente de Łojasiewicz* $\theta = \frac{1}{p}$;

(b) *Seja* $f \in C^k(U;\mathbb{R})$, *e assuma que* $f^{(j)}(a) = 0$ *para* $1 \leqslant j \leqslant k-1$, *e* $f^{(k)}(a) \neq 0$. *Então* f *satisfaz a desigualdade de Łojasiewicz com expoente* $\theta = \frac{1}{k}$.

Demonstração. Sem perda de generalidade podemos supor que $a = 0$.

(a) Como $\lim_{x\to 0} r(x)/|x|^{p-1} = 0$, podemos escolher $\delta > 0$ tal que $|r(x)| \leqslant \frac{c}{2}|x|^{p-1}$, para todo $x \in U$, $|x| < \delta$. Então para $|x| < \delta$,

$$|f'(x)| \geqslant |g(x)| - |r(x)| \geqslant c|x|^{p-1} - \frac{c}{2}|x|^{p-1} = \frac{c}{2}|x|^{p-1}. \tag{2.7}$$

Por outro lado, se $0 < x < \delta$, então

$$\begin{aligned}
|f(x) - f(0)| &= \left|\int_0^x f'(s)ds\right| \\
&= \left|\int_0^x g(s)ds + \int_0^x r(s)ds\right| \\
&\leqslant \int_0^x cs^{p-1}ds + \int_0^x |r(s)|ds \\
&\leqslant \frac{3c}{2p}x^p.
\end{aligned}$$

Se $-\delta < x < 0$, então

$$\begin{aligned}
|f(x) - f(0)| &= \left|\int_x^0 f'(s)ds\right| \\
&= \left|\int_x^0 g(s)ds + \int_x^0 r(s)ds\right| \\
&\leqslant \int_x^0 c(-s)^{p-1}ds + \int_x^0 |r(s)|ds \\
&\leqslant \int_x^0 c(-s)^{p-1}ds + \int_x^0 \frac{c}{2}(-s)^{p-1}ds \\
&\leqslant (-1)^p\frac{3c}{2p}x^p \leqslant \frac{3c}{2p}|x|^p.
\end{aligned}$$

Portanto, para $|x| < \delta$,

$$|f(x) - f(0)| \leqslant \frac{3c}{2p}|x|^p. \tag{2.8}$$

Combinando (2.7) e (2.8), obtemos

$$|f(x) - f(0)|^{1-\frac{1}{p}} \leqslant 3\left(\frac{2}{3c}\right)^{\frac{1}{p}}|f'(x)|.$$

(b) Usando a fórmula de Taylor em f e a hipótese sobre f, obtemos

$$f(x) = f(0) + \frac{f^{(n)}(0)}{n!}x^n + r(x).$$

Assim,

$$f'(x) = \frac{f^{(n)}(0)}{(n-1)!}x^{n-1} + r'(x),$$

com $\lim_{x\to 0} r'(x)/x^{n-1} = 0$. O resultado segue de (a) com $g(x) = \dfrac{f^{(n)}(0)}{(n-1)!}x^{n-1}$. $\square$

O próximo Lema é um resultado clássico que utilizaremos a seguir.

Lema 6 (Desigualdade de Young)**.** *Se* $1 < p < \infty$ *e* $\dfrac{1}{p} + \dfrac{1}{q} = 1$*, então* $\forall a, b \geqslant 0$*,*

$$ab \leqslant \frac{1}{p}a^p + \frac{1}{q}b^q.$$

Demonstração. Se $ab = 0$, o resultado também é óbvio, então suponha que $ab \neq 0$. Defina $\phi(t) = t^p/p - t$, $t \in (0, \infty)$. Note que $\phi''(t) = (p-1)t^{p-2} > 0$, logo f é convexa e como $t_0 = 1$ é ponto crítico de ϕ, tem-se que $\phi(1) \leqslant \phi(t)$, $\forall t \in (0, \infty)$. Em particular, $\phi(1) \leqslant \phi(a/b^{q/p})$. Portanto,

$$ab \leqslant \frac{1}{p}a^p + \frac{1}{q}b^q.$$

□

Lema 7. *Sejam* $p_1, p_2 \in (0,1)$ *dados. Ponha* $r = \dfrac{p_1 p_2}{p_1 + p_2}$*. Então, para todo* $x, y \geqslant 0$*,*

$$(xy)^{1-r} \leqslant x^{1-p_1}y + xy^{1-p_2}.$$

Demonstração. Basta usar a Desigualdade de Young com $a = 1/x^r$, $b = 1/y^r$, $p = p_1/r$ e $q = p_2/r$. □

Proposição 15. *Sejam* $U \subset \mathbb{R}$*,* $a \in U$ *e* $f, g \in C^1(U; \mathbb{R})$ *tais que* $f(a) = g(a) = 0$*. Assuma que* f *e* g *satisfazem a desigualdade de Łojasiewicz em torno do ponto* a *com constantes* θ*,* $\theta' \in [0,1]$*,* c*,* $c' > 0$ *e* δ*,* $\delta' > 0$*, respectivamente. Então a função* $f \cdot g$ *satisfaz a desigualdade de Łojasiewicz com constantes* $\theta'' := \frac{\theta\theta'}{\theta+\theta'}$*,* $c'' := \max\{c, c'\}$ *e* $\delta'' := \min\{\delta, \delta'\}$*.*

Demonstração. Sejam f e g como na hipótese e seja $x \in U$ tal que $|x - a| < \delta$. Pelo Lema 5 (c) temos que para todo $|x - a| < \delta$,

$$f'(x)f(x)g'(x)g(x) \geqslant 0.$$

Além disso, usando o Lema 7 obtemos,

$$\begin{aligned}
|(fg)'(x)| &= |f'(x)g(x) + f(x)g'(x)| \\
&= |f'(x)||g(x)| + |f(x)||g'(x)| \\
&\geqslant \frac{1}{c}|f(x)|^{1-\theta}|g(x)| + \frac{1}{c'}|f(x)||g(x)|^{1-\theta'} \\
&\geqslant \frac{1}{c''}|f(x)g(x)|^{1-\theta''}.
\end{aligned}$$

□

Proposição 16. *Sejam* $U \subset \mathbb{R}$, $a \in U$ *e* $f \in C^2(U;\mathbb{R})$ *tais que* f' *satisfaz a desigualdade de Łojasiewicz em torno de* a *com expoente de Łojasiewicz* $\theta' \in (0,1]$. *Então* f *satisfaz a desigualdade de Łojasiewicz em torno de* a *com expoente de Łojasiewicz* $\theta'' = 1$ *se* $f'(a) \neq 0$ *e* $\theta'' = \frac{\theta'}{1+\theta'}$ *se* $f'(a) = 0$.

Demonstração. Se $f'(a) \neq 0$, existe $\delta > 0$ tal que para todo $|x| < \delta$, tem-se

$$\frac{2}{|f'(a)|}|f'(x)| \geqslant 1.$$

Logo f satisfaz a desigualdade de Łojasiewicz com $\theta = 1$. Assuma agora que $f'(a) = 0$. Por hipótese, existem $\theta' \in (0,1]$, $c > 0$ e $\delta > 0$ tal que para $|x| < \delta$,

$$c|f''(x)| \geqslant |f'(x)|^{1-\theta'}.$$

Substituindo f por $-f$ se necessário, pelo Lema 5 (b) podemos assumir que $f'(x) \geqslant 0$ para $a \leqslant x \leqslant a + \delta$. Então pelo Lema 5 existe $0 < \delta' \leqslant \delta$ tal que para todo $a \leqslant x \leqslant a + \delta'$, $f''(x) \geqslant 0$. Logo,

$$\frac{c}{\theta'+1}\frac{d}{dx}f'(x)^{\theta'+1} = cf'(x)^{\theta'}f''(x) \geqslant f'(x)^{\theta'}f'(x)^{1-\theta'} = f'(x), \quad a \leqslant x \leqslant a+\delta'.$$

Integrando ambos os lados de a a x obtemos,

$$|f(x) - f(a)| \leqslant \frac{c}{\theta'+1}f'(x)^{\theta'+1}, \quad a \leqslant x \leqslant a+\delta',$$

ou seja,

$$|f(x) - f(a)|^{1-\theta''} \leqslant \frac{c}{\theta'+1}f'(x), \quad a \leqslant x \leqslant a+\delta',$$

onde $\theta'' = \dfrac{\theta'}{1+\theta'}$. Analogamente, mostra-se que

$$|f(x) - f(a)|^{1-\theta''} \leqslant \frac{c}{\theta'+1}f'(x), \quad a-\delta' \leqslant x \leqslant a.$$

□

Proposição 17. *Sejam* $U \subset \mathbb{R}^n$ *um aberto,* $f \in C^1(U;\mathbb{R})$ *e* $a \in U$. *Denote por* $\mathbb{S}^{n-1}$ *a espera unitária em* $\mathbb{R}^n$. *Assuma que*

(1) *para todo* $h \in \mathbb{S}^{n-1}$ *a função* $\lambda \mapsto f(a + \lambda h)$ *satisfaz a desigualdade de Łojasiewicz em torno de* 0, *e*

(2) *o expoente de Łojasiewicz e as constantes* c, $\delta > 0$ *podem ser escolhidas independente de* $h \in \mathbb{S}^{n-1}$.

Então a função f *satisfaz a desigualdade de Łojasiewicz em torno de* a *com expoente de Łojasiewicz* θ.

Demonstração. Seja $h \in \mathbb{S}^{n-1}$. Pela hipótese (1), existem constantes $\theta \in (0,1]$, $c > 0$, $\delta > 0$ tal que, para todo $\lambda \in \mathbb{R}$, $|\lambda| < \delta$,

$$|f(a+\lambda h) - f(a)|^{1-\theta} \leqslant c||\langle \nabla f(a+\lambda h), h\rangle||.$$

Pela desigualdade de Cauchy-Schwarz temos,

$$|f(a+\lambda h) - f(a)|^{1-\theta} \leqslant c||\nabla f(a+\lambda h)||. \tag{2.9}$$

Para todo $x \in U$, $||x-a|| < \delta$, existe $h_1 \in \mathbb{S}^{n-1}$ tal que $x = a + \lambda h_1$, com $|\lambda| < \delta$. Como c e $\delta > 0$ podem ser escolhidos independente de $h \in \mathbb{S}^{n-1}$, tem-se que para h_1 vale (2.9). Assim,

$$|f(x) - f(a)|^{1-\theta} \leqslant c||\nabla f(x)||,$$

para todo $x \in U$, $||x-a|| < \delta$.

□

Lema 8. *Seja* $H : \mathbb{R}^n \to \mathbb{R}^n$ *uma transformação linear invertível. Então existe* $c > 0$ *tal que* $||Hx|| \geqslant c||x||$ *para todo* $x \in \mathbb{R}^n$.

Demonstração. Seja $c = 1/||H^{-1}||$. Para todo $x \in \mathbb{R}^n$, temos

$$||x|| = ||H^{-1}(Hx)|| \leqslant ||H^{-1}||||Hx|| = 1/c||H||,$$

donde $||Hx|| \geqslant c||x||$. □

Proposição 18. *Sejam* $U \subset \mathbb{R}^n$ *um aberto,* $f \in C^1(U;\mathbb{R})$ *e assuma que* f *satisfaz a desigualdade de Łojasiewicz em torno de* $a \in U$ *com constantes* c, $\delta > 0$ *e com expoente de Łojasiewicz* θ. *Seja* $V \subset \mathbb{R}^n$ *um aberto e seja* $g \in C^1(V;\mathbb{R}^n)$ *tal que* $g(b) = a$ *para algum* $b \in V$ *e tal que* g *é um difeomorfismo local em torno de* b. *Então a composição* $f \circ g$ *satisfaz a desigualdade de Łojasiewicz em torno de* b *com expoente de Łojasiewicz* θ.

Demonstração. Uma vez que g é um difeomorfismo local em torno de b, a derivada $\nabla g(x) \in \mathbb{R}^{n\times n}$ é invertível para todo x numa vizinhança U de b, e a inversa é contínua. Como g é contínua em b, existe $\delta' > 0$ tal que, para todo $||x-b|| < \delta'$, tem-se $||g(x)-a|| < \delta$. Pelo Lema 8, para todo $x \in U$, existe $c' > 0$ tal que

$$\|(\nabla g(x))^T\| \geqslant c'\|x\|,$$

onde $(\nabla g(x))^T$ denota a transposta de $\nabla g(x)$. Segue que para $x \in U$ e $\|x - b\| < \delta'$,

$$\begin{aligned}
\|\nabla(f \circ g)(x)\| &= \|(\nabla g(x))^T \nabla f(g(x))\| \\
&\geqslant c'\|\nabla f(g(x))\| \\
&\geqslant \frac{c'}{c}\|f(g(x)) - f(a)\|^{1-\theta} \\
&= \frac{c'}{c}\|f(g(x)) - f(g(b))\|^{1-\theta}.
\end{aligned}$$

□

2.1 Propriedade Kurdyka-Łojasiewicz

Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função semi-contínua inferiormente. Para η_1, η_2 tal que $-\infty < \eta_1 < \eta_2 \leqslant +\infty$, denotamos por

$$[\eta_1 < f < \eta_2] = \{x \in \mathbb{R}^n : \eta_1 < f(x) < \eta_2\} \text{ e } \operatorname{dist}(0, \partial f(x)) = \inf\{\|v\| : v \in \partial f(x)\}.$$

Definição 25. (Propriedade Kurdyka-Łojasiewicz)

(a) *Diz-se que uma função* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *tem a propriedade Kurdyka-Łojasiewicz em* $x^* \in$ *dom* ∂f *se existe* $\eta \in (0, +\infty]$, *uma vizinhança* U *de* x^* *e uma função côncava contínua* $\varphi : [0, \eta) \to \mathbb{R}_+$ *tal que:*

(i) $\varphi(0) = 0$,

(ii) φ *é* C^1 *em* $(0, \eta)$,

(iii) *para todo* $s \in (0, \eta)$, $\varphi'(s) > 0$,

(vi) *para todo* $x \in U \cap [f(x^*) < f < f(x^*) + \eta]$, *tem-se a desigualdade Kurdyka-Łojasiewicz*

$$\varphi'(f(x) - f(x^*))\operatorname{dist}(0, \partial f(x)) \geqslant 1. \tag{2.10}$$

(b) *As funções semicontínuas inferiomente quando satisfazem a desigualdade Kurdyka-Łojasiewicz em cada ponto do dom* ∂f *são chamadas funções KL.*

Lema 9. *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *uma função própria semicontínua inferiormente. Seja* $\overline{x} \in \operatorname{dom} \partial f$ *um ponto não crítico de f, isto é,* $0 \notin \partial f(\overline{x})$. *Se*

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

então

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

Demonstração. Suponha por absurdo que existe uma sequência (δ_k) com $\delta_k > 0$, $\delta_k \to 0$ e uma sequência (x^k) tal que

$$\|x^k - \overline{x}\| + |f(x^k) - f(\overline{x})| < \delta_k, \text{ e } \text{dist}(0, \partial f(x^k)) < \delta_k. \tag{2.11}$$

Como $\partial f(x^k)$ é fechado, existe uma sequência $\nu^k \in \partial f(x^k)$ tal que $\|\nu^k\| = \text{dist}(0, \partial f(x^k))$. Segue que $\|\nu^k\| < \delta_k$, e consequentemente, $\nu^k \to 0$. Por outro lado fazendo $k \to +\infty$ em (2.11), obtemos

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

Logo, segue da Proposição 4 que $0 \in \partial f(\overline{x})$, o que contradiz a hipótese. $\square$

Proposição 19. *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função própria semi-contínua inferiormente e $\overline{x} \in \text{dom}\partial f$ tal que $0 \notin \partial f(\overline{x})$. Então a desigualdade Kurdyka-Lojasiewicz é válida em $\overline{x}$.*

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

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

Sejam $\varphi(t) := t/\delta$, $U := B(\overline{x}, \delta/2)$, $\eta := \delta/2$. Daí, para todo $x \in U \cap [f(\overline{x}) < f < f(\overline{x}) + \eta]$, temos

$$\|x - \overline{x}\| + |f(x) - f(\overline{x})| < \delta.$$

Segue do Lema 9 que para todo $x \in \text{dom}\partial f$,

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

$\square$

Vejamos então alguns exemplos de funções KL.

Exemplo 17. *O Teorema 5 mostra que funções $f : \mathbb{R} \to \mathbb{R}$ analíticas em uma vizinhança de um ponto $a \in \mathbb{R}^n$, satisfazem a desigualdade (2.10) com $\varphi(t) = \dfrac{c}{\theta}t^\theta$, e $\eta \in (0, +\infty]$. De fato, para todo $x \in (-\delta, \delta) \cap [0 < f < \eta]$, temos*

$$\varphi'(f(x) - f(a))\text{dist}(0, \partial f(x)) = \frac{c|f'(x)|}{(f(x) - f(a))^{1-\theta}} \geqslant 1.$$

Exemplo 18. *Voltemos ao Exemplo 16, isto é, seja* $f : \mathbb{R} \to \mathbb{R}$ *definida por*

$$f(x) := \begin{cases} e^{-1/x^2} & , \text{ se } \quad x \neq 0 \\ 0 & , \text{ se } \quad x = 0 \end{cases}.$$

Mostraremos que f *não satisfaz a Propriedade Kurdyka-Łojasiewicz em* $x^* = 0$*. Suponha por absurdo que existem* $\eta \in (0, +\infty]$*, uma vizinhança* U *de* 0 *e uma função côncava contínua* $\varphi : [0, \eta) \to \mathbb{R}_+$ *tal que:* $\varphi(0) = 0$*;* φ *é* C^1 *em* $(0, \eta)$*; para todo* $s \in (0, \eta)$*,* $\varphi'(s) > 0$ *e que para todo* $x \in U \cap [0 < f < \eta]$*, tem-se*

$$\varphi'(f(x))|f'(x)| \geqslant 1.$$

Daí,

$$\varphi'(f(x))\frac{2e^{-1/x^2}}{|x|^3} \geqslant 1 \Rightarrow \varphi'(f(x)) \geqslant \frac{|x|^3}{2e^{-1/x^2}}.$$

Logo,

$$\varphi'(0) = \varphi'(f(0)) = \lim_{x \to 0^+} \varphi'(f(x)) \geqslant \lim_{x \to 0^+} \frac{e^{1/x^2}}{1/2x^3} = +\infty.$$

Definição 26. *Dizemos que uma função* $f : \mathbb{R}^n \to \mathbb{R}$ *de classe* C^2 *é uma* ***função de Morse*** *se cada ponto crítico* $\overline{x}$ *de* f *é não degenerado, isto é, se o hessiano* $\nabla^2 f(\overline{x})$ *de* f *em* $\overline{x}$ *tem todos os seus autovalores diferentes de zero.*

Afirmamos que se f é uma função de Morse e se $\overline{x}$ é um ponto crítico de f, então f goza da Propriedade Kurdyca-Łojasiewicz em $\overline{x}$. De fato, seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função de Morse e $\overline{x}$ um ponto crítico de f. Seja ainda $U = B(\overline{x}, \delta)$ tal que não existe outro ponto crítico em U. Usando a fórmula de Taylor em f e em ∇f, obtemos

$$f(x) - f(\overline{x}) = \langle \nabla^2 f(\overline{x})(x - \overline{x}), x - \overline{x} \rangle + \rho(x)\|x - \overline{x}\|^2, \text{ onde } \lim_{x \to \overline{x}} \rho(x) = 0,$$

e

$$\nabla f(x) = \nabla^2 f(\overline{x})(x - \overline{x}) + \widetilde{\rho}(x)\|x - \overline{x}\|, \text{ onde } \lim_{x \to \overline{x}} \widetilde{\rho}(x) = 0.$$

Pelo Lema 8 existe $c > 0$ tal que $\|\nabla^2 f(\overline{x})(x - \overline{x})\| \geqslant c\|x - \overline{x}\|$. Além disso, existem $\varepsilon_1 > 0$ e $\delta_1 > 0$ tal que para $\|x - \overline{x}\| < \delta_1$, $\|\rho(x)\| \leqslant \varepsilon_1$, e existe $\delta_2 > 0$ tal que para $\|x - \overline{x}\| < \delta_2$, $\|\widetilde{\rho}(x)\| < \frac{c}{2}$. Tome $\widetilde{\delta} = \min\{\delta, \delta_1, \delta_2\}$. Daí,

$$\begin{aligned} |f(x) - f(\overline{x})| &\leqslant \|\nabla^2 f(\overline{x})\|\|x - \overline{x}\|^2 + \|\rho(x)\|\|x - \overline{x}\|^2 \\ &\leqslant (\|\nabla^2 f(\overline{x})\| + \varepsilon_1)\|x - \overline{x}\|^2, \end{aligned}$$

e, portanto,

$$|f(x) - f(\overline{x})| \leqslant c_1\|x - \overline{x}\|^2, \text{ onde } c_1 = \|\nabla^2 f(\overline{x})\| + \varepsilon_1.$$

Tem-se ainda que

$$\begin{aligned}
\|\nabla f(x)\| &\geqslant \|\nabla^2 f(\overline{x})(x-\overline{x})\| - \|\widetilde{\rho}(x)\|\|x-\overline{x}\| \\
&\geqslant c\|x-\overline{x}\| - \frac{c}{2}\|x-\overline{x}\| \\
&= c_2\|x-\overline{x}\|, \text{ onde } c_2 = \frac{c}{2}.
\end{aligned}$$

Na Definição 25, tome $U = B(\overline{x}, \widetilde{\delta})$, $\eta = \widetilde{\delta}$ e $\varphi(s) = 2\frac{\sqrt{c_1 s}}{c_2}$. Logo, para todo $x \in U \cap [0 < f(x) - f(\overline{x}) < \eta]$,

$$\varphi'(f(x)-f(\overline{x}))\mathrm{dist}(0, \partial f(x)) = \frac{\sqrt{c_1}}{c_2\sqrt{f(x)-f(\overline{x})}}\|\nabla f(x)\| \geqslant \frac{\|\nabla f(x)\|}{c_2\|x-\overline{x}\|} \geqslant 1.$$

Funções convexas não verificam necessariamente a propriedade de Kurdyka-Łojasiewicz. Como mostra o seguinte teorema.

Teorema 6. *Existe uma funções convexa $f : \mathbb{R}^2 \to \mathbb{R}$ de classe C^2 com $\min\limits_{x\in\mathbb{R}^2} f(x) = 0$ que não satisfaz a desiguadade KL e cujo conjunto de minimizadores é compacto com interior não vazio.*

Demonstração. Veja página 30 de [6]. □

O principal resultado desta seção é o seguinte teorema.

Teorema 7. [11] *Seja $f : U \to \mathbb{R}$ uma $\mathcal{M}$-função diferenciável, onde U é um subconjunto aberto e limitado de $\mathbb{R}^n$. Suponha que $f(x) > 0$ para todo $x \in U$. Então existe $c > 0$, $\rho > 0$ e uma $\mathcal{M}$-função $\Psi : \mathbb{R}_+ \to \mathbb{R}$ estritamente crescente de classe C^1, tal que*

$$\|\nabla(\Psi \circ f)(x)\| \geqslant c, \tag{2.12}$$

para cada $x \in U$, $f(x) \in (0, \rho)$.

Demonstração. Segue do Lema 3 que

$$U \ni x \mapsto \|\nabla f(x)\|$$

é uma $\mathcal{M}$-função.

Podemos supor que $f^{-1}(t) \neq \emptyset$ para qualquer $t > 0$ pequeno. Daí a função $\varphi : (0, \varepsilon) \to \mathbb{R}$ dada por

$$\varphi(t) = \inf\{\|\nabla f(x)\| : x \in f^{-1}(t)\}$$

está bem definida, e pelo Lema 1, φ é uma $\mathcal{M}$-função.

Afirmamos que existe $\varepsilon' > 0$ tal que $\varphi(t) > 0$ para todo $t \in (0, \varepsilon')$.

Suponha por absurdo que existe uma sequência de números reais positivos $(t_n)_{n\in\mathbb{N}}$ tal

que para cada $n \in \mathbb{N}$, $t_n \in (0, \frac{1}{n})$ e $\varphi(t_n) = 0$. Seja
$$\Sigma = \{x \in U : \|\nabla f(x)\| < (f(x))^2\}.$$
Claramente Σ é um $\mathcal{M}$-conjunto, pois $\Sigma = g^{-1}(0, -\infty)$, onde $g(x) = \|\nabla f(x)\| - (f(x))^2$. Seja $x_n \in \Sigma$ uma sequência tal que $f(x_n) = t_n$, em outras palavras, $(x_n, t_n) \in \text{Graf } f|_\Sigma$. Como U é limitado, existe uma subsequência $(x_{n_j})_{j\in\mathbb{N}}$ de $(x_n)_{n\in\mathbb{N}}$ tal que $x_{n_j} \to b$. Então $(b, 0) = \lim_{j\to+\infty}(x_{n_j}, t_{n_j})$, ou seja, $(b, 0) \in \overline{\text{Graf } f|_\Sigma \backslash \{(b, 0)\}}$. No Lema de Seleção de Curva tome $A = \text{Graf } f|_\Sigma$ e $a = (b, 0)$. Logo, existe uma $\mathcal{M}$-curva $\widetilde{\gamma} : (-\delta, \delta) \to \mathbb{R}^n \times \mathbb{R}$ de classe C^1, tal que $\widetilde{\gamma}(0) = (b, 0)$ e $\widetilde{\gamma}((0, \delta)) \subset \text{Graf } f|_\Sigma \backslash \{(b, 0)\}$. Sejam $\widetilde{\gamma}(s) = (\gamma(s), f \circ \gamma(s))$, onde $\gamma(s) \in \Sigma \subset \mathbb{R}^n$ e $h(s) = f \circ \gamma(s)$, para $s \in (0, \delta)$. Como $\widetilde{\gamma}(0) = (b, 0)$, tem-se que $h(0) = 0$ e $\lim_{s\to 0} h(s) = h(0) = 0$. Por outro lado, como $\gamma'(s)$ é contínua existe $A > 0$ tal que $\|\gamma'(s)\| \leqslant A$, $s \in (0, \delta)$. Daí, como $\gamma(s) \in \Sigma$ temos que

$$\begin{aligned} |h'(s)| &= |\langle \nabla(f \circ \gamma)(s), \gamma'(s)\rangle| \\ &\leqslant \|\nabla(f \circ \gamma)(s)\|\|\gamma'(s)\| \\ &< A(f(\gamma(s)))^2 = A(h(s))^2. \end{aligned}$$

Assim,
$$\lim_{s\to 0} |h'(s)| \leqslant \lim_{s\to 0} A(h(s))^2 = 0 \Rightarrow \lim_{s\to 0} |h'(s)| = 0.$$
Pelo Lema 2, h e h' são $\mathcal{M}$-funções e podemos supor que h e h' são monótonas. Como $\lim_{s\to 0} h(s) = \lim_{s\to 0} h'(s) = 0$, segue que h e h' são monótonas crescentes. Assim temos
$$0 < h'(s) \leqslant A(h(s))^2, \;\; s \in (0, \delta).$$
Definindo $\xi : [0, 1] \to \mathbb{R}$, $\xi(t) = h(ts)$, segue do Teorema Fundamental do Cálculo
$$\xi(1) - \xi(0) = \int_0^1 \xi'(t)dt \Rightarrow h(s) = \int_0^1 h'(ts)sdt.$$
Daí, para $0 < t < 1$ e $s \in (0, \delta)$, temos $ts < s$, e consequentemente, $h'(ts) < h'(s)$. Logo
$$h(s) = \int_0^1 h'(ts)sdt \leqslant \int_0^1 sh'(s)dt = sh'(s).$$
Finalmente,
$$0 < h'(s) \leqslant A(h(s))^2 \leqslant As^2(h'(s))^2 \Rightarrow \frac{1}{As^2} \leqslant h'(s).$$
Portanto,
$$+\infty = \lim_{s\to 0} \frac{1}{As^2} \leqslant \lim_{s\to 0} h'(s) = 0.$$
O que é uma contradição.

Assim provamos que existe $\varepsilon > 0$ tal que $\varphi(t) > 0$ para todo $t \in (0, \varepsilon)$.

Definindo

$$\Delta = \{x \in U \backslash f^{-1}(0) : f(x) < \varepsilon, \|\nabla f(x)\| \leqslant 2\varphi(f(x))\},$$

observemos que Δ também é um $\mathcal{M}$-conjunto, e além disso, $\Delta \cap f^{-1}(t) \neq \emptyset$ para todo $t \in (0, \varepsilon)$. Portanto como antes, existe $d \in \overline{U}$ tal que $(d, 0) \in \overline{\text{Graf } f|_\Delta \backslash \{(d, 0)\}}$. Aplicando novamente o Lema de Seleção de Curva para Graf $f|_\Delta$ no ponto $(d, 0)$ obtemos uma $\mathcal{M}$-curva $\widetilde{\eta} : (-\delta, \delta) \to \mathbb{R}^n$ de classe C^1 tal que $\widetilde{\eta}(0) = (d, 0)$ e $\widetilde{\eta}((0, \delta)) \subset \text{Graf } f|_\Delta \backslash \{(d, 0)\}$. Sejam $\widetilde{\eta}(s) = (\eta(s), f \circ \eta(s))$, onde $\eta(s) \in \Delta \subset \mathbb{R}^n$ e $g(s) = f \circ \eta(s)$ para $s \in (0, \delta)$. Assim, $\lim_{s \to 0} g(s) = 0$ e $g(s) > 0$ para cada $s \in (0, \delta)$. Segue do Lema 2 que para todo $\delta' > 0$ a função $g : (0, \delta') \to \mathbb{R}$ é um difeomorfismo em $(0, \rho)$ para algum $\rho > 0$. Ponhamos $\Psi(t) = g^{-1}(t),\ t \in (0, \rho)$. Seja $B > 0$ tal que $\|\eta'(s)\| \leqslant B,\ s \in (0, \delta')$. Tome $x \in U$ tal que $t = f(x) \in (0, \rho)$, e $s = \Psi(t) = g^{-1}(t)$. Como $\eta(s) \in \Delta \subset \mathbb{R}^n$ temos

$$\|\nabla f(\eta(s))\| \leqslant 2\varphi(f(\eta(s))) \leqslant 2\|\nabla f(x)\| \Rightarrow \|\nabla f(x)\| \geqslant \frac{1}{2}\|\nabla f(\eta(s))\|.$$

Além disso,

$$\|\nabla f(\eta(s))\| B \geqslant \|\nabla f(\eta(s))\| \|\eta'(s)\| \geqslant \langle \nabla f(\eta(s)), \eta'(s) \rangle = (f \circ \eta)'(s),$$

e portanto, $\|\nabla f(\eta(s))\| \geqslant \dfrac{1}{B}(f \circ \eta)'(s)$. Logo,

$$\begin{aligned}
\|\nabla(\Psi \circ f)(x)\| &= |\Psi'(f(x))|\, \|\nabla f(x)\| \\
&= \Psi'(f(x))\, \|\nabla f(x)\| \\
&\geqslant \Psi'(t)\frac{1}{2}\|\nabla f(\eta(s))\| \\
&\geqslant \Psi'(t)\frac{(f \circ \eta)'(s)}{2B} \\
&= \Psi'(g(s))\frac{(f \circ \eta)'(s)}{2B} \\
&= \frac{(\Psi \circ g)'(s)}{g'(s)}\frac{(f \circ \eta)'(s)}{2B} = \frac{1}{2B} = c.
\end{aligned}$$

□

Teorema 8. [11] *Seja* $f : \Omega \to \mathbb{R}$ *uma função subanalítica diferenciável em* $\Omega - f^{-1}(0)$, *onde* Ω *é um subconjunto aberto e limitado de* $\mathbb{R}^n$. *Então existem* $C > 0$, $\rho > 0$ *e* $0 \leqslant \alpha < 1$ *tal que*

$$\|\nabla f(x)\| \geqslant C|f(x)|^\alpha$$

para cada $x \in \Omega$ *tal que* $|f(x)| \in (0, \rho)$. *Além disso, se* $\lim_{x \to a} f(x) = 0$ *para algum* $a \in \overline{\Omega}$, *então a desigualdade acima é verdadeira para cada* $x \in \Omega - f^{-1}(0)$ *perto de* a.

Demonstração. Usando o Teorema 7 para o caso subanalítico, podemos aplicar o Teorema 4 para a função Ψ, isto é, existem $\varepsilon > 0$ e $k \in \mathbb{N}$ tal que $\Psi(t) = \sum_{i=0}^{\infty} a_i t^{\frac{i}{k}}$, para $t \in [0, \varepsilon)$. Daí,

$$\begin{aligned}\Psi(t) &= a_0 + a_1 t^{\frac{1}{k}} + a_2 t^{\frac{2}{k}} + a_3 t^{\frac{3}{k}} + ... + a_{k-1} t^{1-\frac{1}{k}} \\ &+ a_k t + a_{k+1} t^{1+\frac{1}{k}} + a_{k+2} t^{1+\frac{2}{k}} + ... \ .\end{aligned}$$

Derivando Ψ, obtemos

$$\begin{aligned}\Psi'(t) &= \frac{a_1}{k} t^{\frac{1}{k}-1} + \frac{2a_2}{k} t^{\frac{2}{k}-1} + \frac{3a_3}{k} t^{\frac{3}{k}-1} + ... + a_{k-1}\Big(1 - \frac{1}{k}\Big) t^{-\frac{1}{k}} \\ &+ a_k + a_{k+1}\Big(1 + \frac{1}{k}\Big) t^{\frac{1}{k}} + a_{k+2}\Big(1 + \frac{2}{k}\Big) t^{\frac{2}{k}} + ... \\ &= t^{\frac{1}{k}-1}\Big[\frac{a_1}{k} + \frac{2a_2}{k} t^{\frac{1}{k}} + \frac{3a_3}{k} t^{\frac{2}{k}} + ... + a_{k-1}\Big(1 - \frac{1}{k}\Big) t^{-\frac{2}{k}+1} \\ &+ a_k t^{1-\frac{1}{k}} + a_{k+1}\Big(1 + \frac{1}{k}\Big) t + a_{k+2}\Big(1 + \frac{2}{k}\Big) t^{1+\frac{1}{k}} + ...\Big].\end{aligned}$$

Portanto,

$$\lim_{t \to 0} \frac{\Psi'(t)}{t^{\frac{1}{k}-1}} = \frac{a_1}{k}.$$

Isso implica que $\Psi'(t)/t^{\frac{1}{k}-1}$ é limitada para t suficientemente pequeno. Sendo assim, existe $M > 0$ tal que $|\Psi'(t)| \leqslant M t^{\frac{1}{k}-1}$. E finalmente,

$$||\nabla f(x)|| = \frac{||\nabla(\Psi \circ f)(x)||}{|\Psi'(f(x))|} \geqslant \frac{c}{M|f(x)|^{\frac{1}{k}-1}} = \frac{c}{M}|f(x)|^{1-\frac{1}{k}}.$$

Tomando $C = \dfrac{c}{M}$ e $\alpha = 1 - \dfrac{1}{k}$, obtemos o resultado. □

Teorema 9. (Desigualdade de Łojasiewicz)

Seja f *uma função analítica real em uma vizinhança da origem em* $\mathbb{R}^n$*, tal que* $f(0) = 0$. *Então existem constantes* c, θ *tais que* $0 < \theta < 1$ *e*

$$c|f(x)|^{\theta} \leqslant |\nabla f(x)| \tag{2.13}$$

em alguma vizinhança de 0.

Demonstração. Como f é analítica, em particular, f é subanalítica fazendo $k = 1$ no Teorema 4. Logo, o resultado segue do Teorema 8 com $a = 0$. □

Capítulo 3

Resultados de convergência para o método de descida inexata

Neste capítulo serão apresentados os principais resultados de convergência para o método de descida inexata. O ponto de partida para a análise desse método é o Lema 10, que sob as condições $\mathcal{H}1$, $\mathcal{H}2$, $\mathcal{H}3$ (Condição suficiente de decrésimo, Condição relativa de erro e Condição de continuidade, respectivamente) fundamenta toda a teoria abordada nos capítulos 3 e 4. Além disso, esse estudo torna-se empolgante devido ao uso e a aplicabilidade das funções KL (Definição 25). O clímax se dá no estudo dos Teoremas 10 e 11, pois tais resultados serão de fundamental importância para o desenvolvimento do capítulo seguinte.

3.1 Resultados de convergência para funções KL

Sejam a e b constantes positivas fixadas, $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função semi-contínua inferiormente e $(x^k)_{k\in\mathbb{N}}$ uma sequência satisfazendo as seguintes condições:

$\mathcal{H}1$. (Condição suficiente de decréscimo). Para cada $k \in \mathbb{N}$,

$$f(x^{k+1}) + a\|x^{k+1} - x^k\|^2 \leqslant f(x^k);$$

$\mathcal{H}2$. (Condição relativa de erro). Para cada $k \in \mathbb{N}$, existe $w^{k+1} \in \partial f(x^{k+1})$ tal que

$$\|w^{k+1}\| \leqslant b\|x^{k+1} - x^k\|;$$

$\mathcal{H}3$. (Condição de continuidade). Existe uma subsequência $(x^{k_j})_{j\in\mathbb{N}}$ e $\widetilde{x}$ tal que

$$x^{k_j} \to \widetilde{x} \;\text{ e }\; f(x^{k_j}) \to f(\widetilde{x}), \text{quando } \; j \to \infty.$$

Considere ainda a seguinte condição:

$\mathcal{H}4$. Para $\delta > 0$ existem $0 < \rho < \delta$ e $\nu > 0$ tal que

$$\left.\begin{array}{c} x \in B(x^*, \rho), f(x) < f(x^*) + \nu \\ y \notin B(x^*, \delta) \end{array}\right\} \Longrightarrow f(x) < f(y) + a\|y - x\|^2.$$

O seguinte resultado está no centro de nossa análise de convergência.

Lema 10. *Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ *uma função semi-contínua inferiormente satisfazendo a propriedade Kurdyka-Lojasiewicz em algum ponto* $x^* \in \mathbb{R}^n$*. Sejam* U*,* η *e* $\varphi : [0, \eta) \to \mathbb{R}_+$ *como na Definição 25 da propriedade KL em* x^**. Sejam* δ*,* $\rho > 0$ *tais que* $B(x^*, \delta) \subset U$ *com* $\rho \in (0, \delta)$*.*

Considere uma sequência $(x^k)_{k \in \mathbb{N}}$ *satisfazendo as condições* $\mathcal{H}1$ e $\mathcal{H}2$. Além disso, assuma que

$$f(x^*) \leqslant f(x^0) < f(x^*) + \eta, \tag{3.1}$$

$$\|x^* - x^0\| + 2\sqrt{\frac{f(x^0) - f(x^*)}{a}} + \frac{b}{a}\varphi(f(x^0) - f(x^*)) < \rho, \tag{3.2}$$

e

$$\forall k \in \mathbb{N}, x^k \in B(x^*, \rho) \Rightarrow x^{k+1} \in B(x^*, \delta) \;\; \textit{com} \;\; f(x^{k+1}) \geqslant f(x^*). \tag{3.3}$$

Então a sequência $(x^k)_{k \in \mathbb{N}}$ satisfaz

$$\forall k \in \mathbb{N}, x^k \in B(x^*, \rho),$$

$$\sum_{k=0}^{+\infty} \| x^{k+1} - x^k \| < +\infty,$$

$$f(x^k) \to f(x^*), \text{ quando } k \to +\infty,$$

e converge para um ponto $\overline{x} \in B(x^*, \delta)$ tal que $f(\overline{x}) \leqslant f(x^*)$.

Se a sequência $(x^k)_{k \in \mathbb{N}}$ também satisfaz a condição $\mathcal{H}3$, então $\overline{x}$ é um ponto crítico de f, e $f(\overline{x}) = f(x^*)$.

Observação 5.

(i) De $\mathcal{H}1$ observamos que a sequência $f(x^k)_{k \in \mathbb{N}}$ é não crescente, e se $f(x^*) < f(x^0) < f(x^*) + \eta$ e $f(x^{k+1}) \geqslant f(x^*)$ então $\varphi(f(x^{k+1}) - f(x^*))$ está bem definida. De fato, $f(x^*) \leqslant f(x^{k+1}) \leqslant f(x^0) < f(x^*) + \eta \Rightarrow (f(x^{k+1}) - f(x^*)) \in [0, \eta)$.

(ii) De $\mathcal{H}2$ segue que o conjunto $\partial f(x^k)$ é não-vazio, e portanto $x^k \in \text{dom } f$.

Antes de demostrarmos o Lema acima, apresentaremos dois resultados preliminares que serão fundamentais para a demonstração do Lema 10, onde serão admitidos todas as suas hipóteses.

Lema 11. *Nas condições do Lema 10, temos*

$$2\|x^{k+1} - x^k\| \leqslant \|x^k - x^{k-1}\| + \frac{b}{a}\Big[\varphi(f(x^k) - f(x^*)) - \varphi(f(x^{k+1}) - f(x^*))\Big]. \tag{3.4}$$

Demonstração. Se $x^{k+1} = x^k$ a desigualdade vale trivialmente. Assim, assumimos que $x^{k+1} \neq x^k$. Por (3.3) e por $\mathcal{H}1$ temos $f(x^*) \leqslant f(x^{k+1}) < f(x^k)$.

Mostremos agora que $w^k \neq 0$ e $x^{k-1} \neq x^k$. Por hipótese $f(x^k) < f(x^*) + \eta$, o que implica $f(x^k) - f(x^*) < \eta$. Por outro lado, como $f(x^*) \leqslant f(x^{k+1}) < f(x^k)$ temos que

$$0 \leqslant f(x^{k+1}) - f(x^*) < f(x^k) - f(x^*) \Rightarrow (f(x^k) - f(x^*)) \in (0, \eta).$$

Por (2.10) temos que

$$\text{dist}(0, \partial f(x^k)) \geqslant \frac{1}{\varphi'(f(x^k) - f(x^*))} > 0,$$

e por conseguinte, $\|w^k\| \geqslant \inf\{\|x\| : x \in \partial f(x^k)\} = \text{dist}(0, \partial f(x^k)) > 0$. Portanto $w^k \neq 0$.

Por $\mathcal{H}2$, $0 < \|w^k\| \leqslant b\|x^k - x^{k-1}\|$, e como $b > 0$, tem-se que $0 < \|x^k - x^{k-1}\| \Rightarrow x^k \neq x^{k-1}$.

Uma vez que $w^k \in \partial f(x^k)$, e usando novamente a (2.10) e $\mathcal{H}2$ obtemos

$$\varphi'(f(x^k) - f(x^*)) \geqslant \frac{1}{\text{dist}(0, \partial f(x^k))} \geqslant \frac{1}{\|w^k\|} \geqslant \frac{1}{b\|x^k - x^{k-1}\|}.$$

Como φ é côncava tem-se que

$$\varphi(y) \leqslant \varphi(x) + \varphi'(x)(y - x) \quad \forall x, y \in [0, \eta).$$

Em particular, para $x = f(x^k) - f(x^*)$ e $y = f(x^{k+1}) - f(x^*)$ obtemos

$$\varphi(f(x^{k+1}) - f(x^*)) \leqslant \varphi(f(x^k) - f(x^*)) + \varphi'(f(x^k) - f(x^*))(f(x^{k+1}) - f(x^k))$$

Daí,

$$\begin{aligned}\varphi(f(x^k) - f(x^*)) - \varphi(f(x^{k+1}) - f(x^*)) &\geqslant \varphi'(f(x^k) - f(x^*))(f(x^k) - f(x^{k+1})) \\ &\geqslant \frac{1}{b\|x^k - x^{k+1}\|}(f(x^k) - f(x^{k+1})) \\ &\geqslant \frac{a\|x^{k+1} - x^k\|^2}{b\|x^k - x^{k-1}\|},\end{aligned}$$

onde a última desigualdade segue de $\mathcal{H}1$. Portanto,

$$\sqrt{\|x^k - x^{k-1}\|(b/a)\Big[\varphi(f(x^k) - f(x^*)) - \varphi(f(x^{k+1}) - f(x^*))\Big]} \geqslant \|x^{k+1} - x^k\|.$$

Usando o fato de $2\sqrt{\alpha\beta} \leqslant \alpha + \beta$ com

$$\alpha = \|x^k - x^{k+1}\| \text{ e } \beta = \frac{b}{a}\Big[\varphi(f(x^k) - f(x^*)) - \varphi(f(x^{k+1}) - f(x^*))\Big],$$

obtemos

$$\|x^k - x^{k-1}\| + \frac{b}{a}\Big[\varphi(f(x^k) - f(x^*)) - \varphi(f(x^{k+1}) - f(x^*))\Big] \geqslant 2\|x^{k+1} - x^k\|.$$

Isso prova (3.4). □

Lema 12. *Para* $k = 1, 2, ...$

$$x^k \in B(x^*, \rho), \tag{3.5}$$

e

$$\|x^{k+1} - x^k\| + \sum_{i=1}^{k} \|x^{i+1} - x^i\| \leqslant \|x^1 - x^0\| + \frac{b}{a}\Big[\varphi(f(x^1) - f(x^*)) - \varphi(f(x^{k+1}) - f(x^*))\Big]. \tag{3.6}$$

Demonstração. Vamos provar as afirmações (3.5) e (3.6) por indução em k. Em (3.3), com $k = 0$, obtemos que $x^1 \in B(x^*, \delta)$ e $f(x^1) \geqslant f(x^*)$. Usando $\mathcal{H}1$ com $k = 0$, temos

$$\|x^1 - x^0\| \leqslant \sqrt{\frac{f(x^0) - f(x^1)}{a}},$$

e o como $f(x^1) \geqslant f(x^*)$ conclui-se que

$$\|x^1 - x^0\| \leqslant \sqrt{\frac{f(x^0) - f(x^1)}{a}} \leqslant \sqrt{\frac{f(x^0) - f(x^*)}{a}}. \tag{3.7}$$

Combinando (3.7) com a desigualdade triângular obtemos

$$\begin{aligned} \|x^* - x^1\| &\leqslant \|x^* - x^0\| + \|x^0 - x^1\| \\ &\leqslant \|x^* - x^0\| + \sqrt{\frac{f(x^0) - f(x^*)}{a}}. \end{aligned}$$

De (3.2) temos que

$$\sqrt{\frac{f(x^0) - f(x^*)}{a}} < \frac{\rho}{2} - \frac{b}{2a}\varphi(f(x^0) - f(x^*)) - \frac{\|x^* - x^0\|}{2}. \tag{3.8}$$

Daí,

$$\begin{aligned} \|x^* - x^1\| &< \|x^* - x^0\| + \frac{\rho}{2} - \frac{b}{2a}\varphi(f(x^0) - f(x^*)) - \frac{\|x^* - x^0\|}{2} \\ &< \frac{\rho}{2} + \frac{\rho}{2} - \frac{b}{2a}\varphi(f(x^0) - f(x^*)) < \rho. \end{aligned}$$

Portanto, $x^1 \in B(x^*, \rho)$. Fazendo $k = 1$ em (3.4) obtemos

$$2\|x^2 - x^1\| \leqslant \|x^1 - x^0\| + \frac{b}{a}\Big[\varphi(f(x^1) - f(x^*)) - \varphi(f(x^2) - f(x^*))\Big].$$

Logo vale (3.6) para $k = 1$. Suponha que (3.5) e (3.6) valem para algum $k \geqslant 1$. Assim decorre da desigualdade triângular, de (3.2) e de (3.7) que

$$\begin{aligned}
\|x^* - x^{k+1}\| &\leqslant \|x^* - x^0\| + \|x^0 - x^1\| + \|x^1 - x^2\| + ... + \|x^k - x^{k+1}\| \\
&= \|x^* - x^0\| + \|x^0 - x^1\| + \sum_{i=1}^{k} \|x^{i+1} - x^i\| \\
&\leqslant \|x^* - x^0\| + 2\|x^0 - x^1\| - \|x^{k+1} - x^k\| \\
&\quad + \frac{b}{a}\Big[\varphi(f(x^1) - f(x^*)) - \varphi(f(x^{k+1}) - f(x^*))\Big] \\
&\leqslant \|x^* - x^0\| + 2\|x^0 - x^1\| + \frac{b}{a}\Big[\varphi(f(x^1) - f(x^*)) - \varphi(f(x^{k+1}) - f(x^*))\Big] \\
&\leqslant \|x^* - x^0\| + 2\sqrt{\frac{f(x^0) - f(x^*)}{a}} + \frac{b}{a}\varphi(f(x^0) - f(x^*)) \\
&\quad - \frac{b}{a}\varphi(f(x^{k+1}) - f(x^*)) \\
&< \rho - \frac{b}{a}\varphi(f(x^{k+1}) - f(x^*)) < \rho.
\end{aligned}$$

Logo $x^{k+1} \in B(x^*, \rho)$. Por outro lado, segue de (3.4) substituindo k por $k+1$ que,

$$2\|x^{(k+1)+1} - x^{k+1}\| \leqslant \|x^{k+1} - x^k\| + \frac{b}{a}\Big[\varphi(f(x^{k+1}) - f(x^*)) - \varphi(f(x^{(k+1)+1}) - f(x^*))\Big].$$

Adicionando (3.6) à desigualdade acima obtemos

$$\sum_{i=1}^{k+1} \|x^{i+1} - x^i\| + \|x^{(k+1)+1} - x^{k+1}\| \leqslant \|x^1 - x^0\| + \frac{b}{a}\Big[\varphi(f(x^1) - f(x^*)) - \varphi(f(x^{(k+1)+1}) - f(x^*))\Big],$$

o que completa a prova por indução. □

Prova do Lema 10

A primeira afirmativa segue do Lema 12. Por outro lado, segue de (3.6) que

$$\sum_{i=1}^{k+1} \|x^{i+1} - x^i\| \leqslant \|x^1 - x^0\| + \frac{b}{a}\varphi(f(x^1) - f(x^*)).$$

Portanto,

$$\sum_{k=1}^{+\infty} \|x^{k+1} - x^k\| < +\infty.$$

O que implica que $(x^k)_{k\in\mathbb{N}}$ converge para algum $\bar{x}$.

Por $\mathcal{H}2$ $w^k \to 0$, e por (3.3), $f(x^k) \to \beta \geqslant f(x^*)$. Supondo $\beta > f(x^*)$, tem-se que $f(x^k) - f(x^*) > 0$. Por outro lado, como φ é C^1 em $(0, \eta)$ temos que $\varphi'(f(x^k) - f(x^*)) \to \varphi'(\beta - f(x^*))$ quando $k \to \infty$ e usando (2.10) obtemos

$$\varphi'(f(x^k) - f(x^*))\|w^k\| \geqslant \varphi'(f(x^k) - f(x^*))\mathrm{dist}(0, \partial f(x^k)) \geqslant 1.$$

Essas afirmativas nos levam a uma contradição quando $k \to \infty$ na desigualdade acima. Assim $\beta = f(x^*)$. Como f é semi-contínua inferiormente

$$f(\overline{x}) \leqslant \liminf_{k\to\infty} f(x^k) = \beta = f(x^*).$$

Se $(x^k)_{k\in\mathbb{N}}$ satisfaz $\mathcal{H}3$, então $\overline{x} = \widetilde{x}$ pois $x^k \to \overline{x}$ e $f(\overline{x}) = f(x^*)$ pois $f(x^k) \to f(x^*)$. Como $(x^k, w^k) \in \mathrm{Graf}\partial f$, $(x^k, w^k) \to (\overline{x}, 0)$ e $f(x^k) \to f(\overline{x})$, segue da Observação 1 que $\overline{x}$ é um ponto crítico de f.

□

Corolário 8. *Sejam* f*,* x^**,* ρ *e* δ *como no Lema 10. Para* $q \geqslant 1$*, considere a família finita* x^0*,...,*x^q *o qual satisfaz* $\mathcal{H}1$ e $\mathcal{H}2$, as condições (3.1), (3.2) e

$$\forall k \in \{0, ..., q\}, \left(x^k \in B(x^*, \rho)\right) \Rightarrow \left(x^{k+1} \in B(x^*, \delta) \text{ com } f(x^{k+1}) \geqslant f(x^*)\right).$$

Então $x^j \in B(x^*, \rho)$ para todo $j = 0, ..., q$.

Demonstração. A prova é análoga à de (3.5). □

Corolário 9. *Se substituirmos a hipótese* (3.3) *no Lema 10 pelo conjunto de pressupostos,*

$$\eta < a(\delta - \rho)^2, \tag{3.9}$$

$$f(x^k) \geqslant f(x^*), \forall k \in \mathbb{N}, \tag{3.10}$$

a conlusão permanece inalterada.

Demonstração. É suficiente provar que (3.9) e (3.10) implicam na condição (3.3).

Por $\mathcal{H}2$ temos que

$$\|x^{k+1} - x^k\| \leqslant \sqrt{\frac{f(x^k) - f(x^{k+1})}{a}}.$$

Por outro lado,

$$f(x^*) \leqslant f(x^{k+1}) \leqslant f(x^k) \leqslant f(x^0) < f(x^*) + \eta \Rightarrow f(x^k) - f(x^{k+1}) < \eta.$$

Assim, segue de (3.9) que

$$\|x^{k+1} - x^k\| \leqslant \sqrt{\frac{\eta}{a}} < \delta - \rho,$$

e portanto

$$\|x^{k+1} - x^*\| \leqslant \|x^{k+1} - x^k\| + \|x^k - x^*\| < (\delta - \rho) + \rho = \delta.$$

Logo, $x^{k+1} \in B(x^*, \delta)$. □

Teorema 10. (Convergência para um ponto crítico)

Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ *uma função semi-contínua inferiormente. Considere* $(x^k)_{k\in\mathbb{N}}$ *uma sequência que satisfaz* $\mathcal{H}1$, $\mathcal{H}2$ e $\mathcal{H}3$.

Se f *tem a propriedade Kurdyka-Lojasiewicz no ponto de acumulação* $\widetilde{x}$*, especificado em* $\mathcal{H}3$*, então a sequência* $(x^k)_{k\in\mathbb{N}}$ *converge para* $\widetilde{x}$ *quando* $k \to +\infty$*, e* $\widetilde{x}$ *é um ponto crítico de* f*. Além disso, a sequência* $(x^k)_{k\in\mathbb{N}}$ *tem comprimento finito, isto é,*

$$\sum_{k=0}^{+\infty} \| x^{k+1} - x^k \| < +\infty.$$

Demonstração. Seja $\widetilde{x}$ um ponto de acumulação de $(x^k)_{k\in\mathbb{N}}$ dado em $\mathcal{H}3$ (isto é, $x^{k_j} \to \widetilde{x}$ e $f(x^{k_j}) \to f(\widetilde{x})$. Uma vez que $(f(x^k))_{k\in\mathbb{N}}$ é uma sequência não-crescente, deduzimos que $f(x^k) \to f(\widetilde{x})$ e $f(x^k) \geqslant f(\widetilde{x})$ para todo $k \in \mathbb{N}$. Como a função f tem a propriedade KL em torno de $\widetilde{x}$, existem φ, U e η como na Definição 25. Seja $\delta > 0$ tal que $B(\widetilde{x}, \delta) \subset U$ e $\rho \in (0, \delta)$.

Como $f(x^k) \to f(\widetilde{x})$, existe $k_1 \in \mathbb{N}$ tal que

$$\forall k \geqslant k_1 \Rightarrow f(\widetilde{x}) \leqslant f(x^k) < f(\widetilde{x}) + \eta,$$

com $\eta < a(\delta - \rho)^2$.

Definamos

$$\lambda_k := \|\widetilde{x} - x^k\| + 2\sqrt{\frac{f(x^k) - f(\widetilde{x})}{a}} + \frac{b}{a}\varphi(f(x^k) - f(\widetilde{x})).$$

Como φ é contínua e $\varphi(0) = 0$, tem-se que $\lim_{j\to+\infty} \lambda_{k_j} = 0$, logo existe $k_{j_0} \in \mathbb{N}$ tal que $\forall k_j \geqslant k_{j_0} \Rightarrow \lambda_{k_j} < \rho$. Tomando $k_0 = \max\{k_1, k_{j_0}\}$, temos que $\forall k \geqslant k_0$, $f(x^k) \in [f(\widetilde{x}), f(\widetilde{x}) + \eta)$ e $\lambda_{k_j} < \rho$. Portanto $\lambda_{k_0} < \rho$, ou seja,

$$\|\widetilde{x} - x^{k_0}\| + 2\sqrt{\frac{f(x^{k_0}) - f(\widetilde{x})}{a}} + \frac{b}{a}\varphi(f(x^{k_0}) - f(\widetilde{x})) < \rho.$$

Seja $(y^k)_{k\in\mathbb{N}}$ definida por $y^k := x^{k_0+k}$. A sequência (y^k) satisfaz as hipóteses (3.1), (3.2), (3.9) e (3.10), assim a conclusão de Lema 10 vale para a sequência (y^k) e consequentemente vale para (x^k). □

Teorema 11. (Convergência local para mínimos locais)
Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}\cup\{+\infty\}$ *uma função semi-contínua inferiormente que tem a propriedade KL em algum minimizador local* x^**. Assuma* $\mathcal{H}4$ em x^*.

Então, para qualquer $r > 0$*, existe* $u \in (0, r)$ *e* $\mu > 0$ *de tal forma que as desigualdades*

$$\| x^0 - x^* \| < u, \; f(x^*) < f(x^0) < f(x^*) + \mu$$

impliquem que qualquer sequência $(x^k)_{k\in\mathbb{N}}$ *a partir de* x^0*, que satisfaz* $\mathcal{H}1$, $\mathcal{H}2$ *tem o comprimento finito, permanece em* $B(x^*, r)$ *e converge para algum ponto crítico* $\overline{x} \in B(x^*, r)$ *de* f *com* $f(\overline{x}) = f(x^*)$*.*

Demonstração. Tome $r > 0$. Uma vez que f satisfaz a propriedade Kurdyka-Łojasiewicz, existe $\eta_0 \in (0, +\infty]$, $\delta \in (0, r)$ e uma função côncava contínua $\varphi : [0, \eta_0) \to \mathbb{R}^+$ como na

Definição 25. Além disso como x^* é um mínimo local, para todo $x \in B(x^*, \delta)$,

$$f(x) \geqslant f(x^*). \tag{3.11}$$

Podemos inferir de $\mathcal{H}4$ que existe $\rho \in (0, \delta)$ e $\nu > 0$ tal que

$$\left.\begin{array}{c} x \in B(x^*, \rho), f(x) < f(x^*) + \nu \\ y \notin B(x^*, \delta) \end{array}\right\} \Longrightarrow f(x) < f(y) + a\|y - x\|^2$$

Seja $\eta = \min\{\eta_0, \nu\}$ e seja $k \in \mathbb{N}$. Se x^k é tal que $f(x^k) < f(x^*) + \eta$ e $\|x^k - x^*\| < \rho$, afirmamos que $x^{k+1} \in B(x^*, \delta)$. De fato, se $x^{k+1} \notin B(x^*, \delta)$ teríamos $x^k \in B(x^*, \rho)$, $f(x^k) < f(x^*) + \eta \leqslant f(x^*) + \nu$, e assim por $\mathcal{H}4$,

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

Esse fato contradiz $\mathcal{H}1$. Logo $x^{k+1} \in B(x^*, \delta)$, e por conseguinte segue de (3.11) que $f(x^{k+1}) \geqslant f(x^*)$. Isto é precisamente a propriedade (3.3) do Lema 10.

Escolha u, $\mu > 0$ tal que

$$u < \frac{\rho}{3}, \quad \mu < \eta, \quad 2\sqrt{\frac{\mu}{a}} + \frac{b}{a}\varphi(\mu) < \frac{2\rho}{3}.$$

Se x^0 satisfaz o conjunto de desigualdades $\| x^0 - x^* \| < u$, $f(x^*) < f(x^0) < f(x^*) + \mu$, então

(i) $0 < f(x^0) - f(x^*) < \mu < \eta$;

(ii) $u + 2\sqrt{\frac{\mu}{a}} + \frac{b}{a}\varphi(\mu) < \frac{\rho}{3} + \frac{2\rho}{3} = \rho$;

(iii) $\|x^* - x^0\| + 2\sqrt{\frac{\mu}{a}} + \frac{b}{a}\varphi(\mu) < \rho$.

Como $\eta = \min\{\eta_0, \nu\}$ e φ é crescente em $(0, \eta_0)$ tem-se

$$\| x^* - x^0 \| + 2\sqrt{\frac{f(x^0) - f(x^*)}{a}} + \frac{b}{a}\varphi(f(x^0) - f(x^*)) < \rho,$$

isso é precisamente a propriedade (3.2) do Lema 10. Usando o Lema 10 podemos concluir que a sequência $(x^k)_{k\in\mathbb{N}}$ tem a propriedade de comprimento finito, permanece em $B(x^*, \rho)$, converge para algum $\bar{x} \in B(x^*, \delta)$, $f(x^k) \to f(x^*)$ e $f(\bar{x}) \leqslant f(x^*)$. Uma vez que $f(x^*)$ é o valor mínimo de f em $B(x^*, \delta)$ tem-se $f(\bar{x}) = f(x^*)$ e $(x^k)_{k\in\mathbb{N}}$ tem a propriedade $\mathcal{H}3$. Assim, $\bar{x}$ é uma ponto crítico de f. □

Observação 6. Vamos verificar que a condição $\mathcal{H}4$ é satisfeita quando $x^* \in \operatorname{dom} f$ é um mínimo local e a função f satisfaz a condição de crescimento global:

$$f(y) \geqslant f(x^*) - \frac{a}{4}\|y - x^*\|^2, \ \forall y \in \mathbb{R}^n \text{ e } a > 0. \tag{3.12}$$

Sejam δ, ρ e ν números reais positivos tal que $\delta > \rho$ e $\dfrac{\delta^2}{4} > \dfrac{\nu}{a} + \rho^2$. Tome $y \in \mathbb{R}^n$ tal que $\|y - x^*\| \geqslant \delta$ e $x \in \mathbb{R}^n$ tal que $\|x - x^*\| \leqslant \rho$ e $f(x) < f(x^*) + \nu$. De (3.12), e da desigualdade triângular podemos inferir que

$$\begin{aligned}
f(y) &\geqslant f(x^*) - \frac{a}{4}\|y - x^*\|^2 \\
&> f(x) - \nu - \frac{a}{4}\|y - x^*\|^2 \\
&= f(x) - \nu - \frac{a}{2}\|y - x^*\|^2 + \frac{a}{4}\|y - x^*\|^2 \\
&\geqslant f(x) - \nu - \frac{a}{2}\|y - x\|^2 - \frac{a}{2}\|x - x^*\|^2 + \frac{a}{4}\|y - x^*\|^2 \\
&\geqslant f(x) - \nu - a\|y - x\|^2 - a\|x - x^*\|^2 + \frac{a}{4}\|y - x^*\|^2 \\
&\geqslant f(x) - \nu - a\|y - x\|^2 - a\rho^2 + \frac{a}{4}\delta^2.
\end{aligned}$$

Daí,

$$f(y) + a\|y - x^*\|^2 \geqslant f(x) + \left(-\nu - a\rho^2 + \frac{a}{4}\delta^2\right) > f(x), \ \forall y \in \mathbb{R}^n \text{ e } a > 0.$$

Teorema 12. (Convergência local para mínimos globais)

Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ *uma função semi-contínua inferiormente com a propriedade KL em algum ponto* x^*, *onde* x^* *é um ponto de mínimo global de* f. *Para cada* $r > 0$, *existe* $u \in (0, r)$ *e* $\mu > 0$ *de tal forma que as desigualdades*

$$\| x^0 - x^* \| < u, \ \textit{min}\, f < f(x^0) < \textit{min}\, f + \mu$$

impliquem que qualquer sequência $(x^k)_{k \in \mathbb{N}}$ *que satisfaz* $\mathcal{H}1$, $\mathcal{H}2$ *e que começa a partir de* x^0 *satisfaz*

(i) $x^k \in B(x^*, r), \forall k \in \mathbb{N}$,

(ii) x^k converge para algum $\overline{x}$ e $\sum_{k=1}^{+\infty} \| x^{k+1} - x^k \| < +\infty$,

(iii) $f(\overline{x}) = \min f$.

Demonstração. Como $f(x^*) \leqslant f(y)$, $\forall y \in \mathbb{R}^n$ temos que

$$f(x^*) - \frac{a}{4}\|y - x^*\|^2 \leqslant f(x^*) \leqslant f(y), \ \forall y \in \mathbb{R}^n.$$

Isso é exatamente a condição (3.12). Logo o resultado segue do Teorema 11. □

Capítulo 4

Análise de convergência de alguns algoritmos

Como aplicação do Capítulo 3, o Capítulo 4 completa esse trabalho com a análise de convergência de algoritmos nos moldes da teoria do Lema 10. Com um algoritmo em mãos, a busca pelas condições $\mathcal{H}1$, $\mathcal{H}2$ e $\mathcal{H}3$ torna-se uma ação frequente e instigante.

4.1 Método do gradiente inexato

4.1.1 Resultado geral de convergência

Antes de estudarmos a convergência do método do gradiente inexato, vamos analisar exemplo a seguir.

Exemplo 19. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função de classe* C^1. *Considere uma sequência* $(A^k)_{k\in\mathbb{N}}$ *de matrizes simétricas definidas positiva em* $\mathbb{R}^{n^2}$ *tal que para cada* $k \in \mathbb{N}$ *os autovalores de* A^k *satisfazem*

$$0 < \frac{a}{2} \leqslant \lambda_i^k \leqslant b,$$

onde a *e* b *são constantes dadas.*

Fixe $x^0 \in \mathbb{R}^n$. *Considere o seguinte algoritmo:*

$$x^{k+1} = x^k - (A^k)^{-1}\nabla f(x^k). \tag{4.1}$$

Afirmamos que (4.1) *satisfaz as condições:*

$$\langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{a}{2}\|x^{k+1} - x^k\|^2 \leqslant 0 \tag{4.2}$$

e

$$\|\nabla f(x^k)\| \leqslant b\|x^{k+1} - x^k\|. \tag{4.3}$$

Para ver (4.2), *basta notar que se* $A \in \mathbb{R}^{n^2}$ *é uma matriz definida positiva, então*

$$\langle Ax, x\rangle \geqslant \lambda\|x\|^2 \ \forall\ x \in \mathbb{R}^n,$$

onde $\{\lambda_1, ..., \lambda_n\}$ *é o conjunto dos autovalores de* A *e* $\lambda = \min\{\lambda_1, ..., \lambda_n\}$. *Daí,*

$$\begin{aligned}\langle \nabla f(x^k), x^{k+1} - x^k\rangle &= \langle -A^k(x^{k+1} - x^k), x^{k+1} - x^k\rangle \\ &\leqslant -\frac{a}{2}\|x^{k+1} - x^k\|^2,\end{aligned}$$

e portanto,

$$\langle \nabla f(x^k), x^{k+1} - x^k\rangle + \frac{a}{2}\|x^{k+1} - x^k\|^2 \leqslant 0.$$

Por outro lado, para verificar (4.3) *basta notar que* $\|A^k\| = \max\{|\lambda_1^k|, ..., |\lambda_n^k|\} = \max\{\lambda_1^k, ..., \lambda_n^k\}$, *e então*

$$\begin{aligned}\|\nabla f(x^k)\|^2 &= \langle \nabla f(x^k), \nabla f(x^k)\rangle \\ &= \langle \nabla f(x^k), -A^k(x^{k+1} - x^k)\rangle \\ &\leqslant \|\nabla f(x^k)\|\|A^k(x^{k+1} - x^k)\| \\ &\leqslant \|\nabla f(x^k)\|\|A^k\|\|x^{k+1} - x^k\| \\ &\leqslant b\|\nabla f(x^k)\|\|x^{k+1} - x^k\|.\end{aligned}$$

Assim,

$$\|\nabla f(x^k)\| \leqslant b\|x^{k+1} - x^k\|.$$

O exemplo 19 ilustra a aplicabilidade do

Algoritmo 1. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função de classe* C^1 *cujo gradiente é Lipschitz contínuo com constante* L *(ou* L*-Lipschitz contínuo). Tomemos alguns parâmetros* a, b *com* $a > L$.

Fixe $x^0 \in \mathbb{R}^n$. *Para* $k = 0, 1, ...$ *considere:*

$$\langle \nabla f(x^k), x^{k+1} - x^k\rangle + \frac{a}{2}\|x^{k+1} - x^k\|^2 \leqslant 0, \tag{4.4}$$

$$\|\nabla f(x^k)\| \leqslant b\|x^{k+1} - x^k\|. \tag{4.5}$$

Para a análise de convergência do Algoritmo 1, faremos uso de um importante Lema de descida.

Lema 13. (Lema de Descida)

Seja $f : \mathbb{R}^n \to \mathbb{R}$ *uma função e* D *um subconjunto convexo de* $\mathbb{R}^n$ *com interior não-vazio. Assuma que* f *é* C^1 *em uma vizinhança de cada ponto em* D *e que* ∇f *e* L*-Lipschitz contínuo em* D*. Então, para quaisquer dois pontos* x*,* y *em* D*,*

$$f(x) \leqslant f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2}\|x - y\|^2. \tag{4.6}$$

Demonstração. Defina $h : [0,1] \to \mathbb{R}$ pondo $h(t) = f(y + t(x - y))$. Pelo Teorema Fundamental do Cálculo,

$$h(1) - h(0) = \int_0^1 h'(t)dt,$$

e por conseguinte,

$$f(x) - f(y) = \int_0^1 \langle \nabla f(y + t(x - y)), y - x \rangle dt.$$

Daí,

$$\begin{aligned} f(x) - f(y) - \langle \nabla f(y), x - y \rangle &= \int_0^1 \langle \nabla f(y + t(x - y)) - \nabla f(y), x - y \rangle dt \\ &\leqslant \int_0^1 \|\nabla f(y + t(x - y)) - \nabla f(y)\| \|x - y\| dt \\ &\leqslant L\|x - y\|^2 \int_0^1 t dt = \frac{L}{2}\|x - y\|^2. \end{aligned}$$

Isso prova (4.6). □

Temos então o seguinte resultado:

Teorema 13. *Assuma que* $f : \mathbb{R}^n \to \mathbb{R}$ *é uma função* C^1*, limitada inferiormente com gradiente* L*-Lipschitz contínuo. Se* f *é uma função KL, então cada sequência limitada* $(x^k)_{k \in \mathbb{N}}$ *gerada pelo Algoritmo 1 converge para algum ponto crítico* $\overline{x}$ *de* f*. Além disso, a sequência* $(x^k)_{k \in \mathbb{N}}$ *tem comprimento finito, isto é,* $\sum_{k=1}^{+\infty} \|x^{k+1} - x^k\| < +\infty$.

Demonstração. Aplicando o Lema 13 aos pontos $x = x^{k+1}$ e $y = x^k$ e usando a desigualdade (4.4) obtemos

$$f(x^{k+1}) - f(x^k) \leqslant \langle \nabla f(x^{k+1}), x^{k+1} - x^k \rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2 \leqslant -\frac{a}{2}\|x^{k+1} - x^k\|^2 + \frac{L}{2}\|x^{k+1} - x^k\|^2,$$

e portanto

$$f(x^{k+1}) + \frac{a - L}{2}\|x^{k+1} - x^k\|^2 \leqslant f(x^k).$$

Uma vez que $a > L$, a condição $\mathcal{H}1$ está satisfeita. Pela Proposição 2, para cada $k \in \mathbb{N}, \nabla f(x^{k+1}) \in \partial f(x^{k+1})$. Por outro lado, fazendo uso de que ∇f é Lipschitz contínuo e da desigualdade (4.5), obtemos

$$\|\nabla f(x^{k+1})\| \leqslant \|\nabla f(x^{k+1}) - \nabla f(x^k)\| + \|\nabla f(x^k)\| \leqslant (L+b)\|x^{k+1} - x^k\|.$$

Assim se verifica a condição $\mathcal{H}2$.

A sequência $(x^k)_{k\in\mathbb{N}}$ foi tomada limitada, logo admite subsequência $(x^{k_j})_{j\in\mathbb{N}}$ convergente. Digamos que $x^{k_j} \to \widetilde{x}$, quando $j \to \infty$. Por continuidade de f, temos que $f(x^{k_j}) \to f(\widetilde{x})$, logo a condição $\mathcal{H}3$ está satisfeita. O resultado segue aplicando o Teorema 10. □

4.1.2 Projeções médias para problemas de viabilidade

Sejam $F_1, ..., F_p$ subconjuntos fechados, não vazios, semi-algébricos, prox-regulares de $\mathbb{R}^n$ tal que

$$\bigcap_{i=1}^{p} F_i \neq \emptyset.$$

Um problema de viabilidade consiste em encontrar algum ponto $\overline{x} \in \cap_{i=1}^{p} F_i$. Resolver este tipo de problema, equivale a garantir a existência de um minimizador global da função $f : \mathbb{R}^n \to [0, +\infty)$ dada por

$$f(x) := \frac{1}{2}\sum_{i=1}^{p} \operatorname{dist}(x, F_i)^2, \tag{4.7}$$

onde $\operatorname{dist}(x, F_i) = \inf\{\|x - y_i\|;\ y_i \in F_i\}$.

Algoritmo da projeção média inexata.

Fixe $\theta \in (0,1)$, $\alpha < \frac{1}{2}$ e $M > 0$ tal que

$$\frac{1-\alpha}{\theta} > \frac{1}{2}. \tag{4.8}$$

Dado $x^0 \in \mathbb{R}^n$, considere o seguinte algoritmo

$$x^{k+1} \in (1-\theta)x^k + \theta\left(\frac{1}{p}\sum_{i=1}^{p} P_{F_i}(x^k)\right) + \varepsilon^k, \tag{4.9}$$

onde $(\varepsilon^k)_{k\in\mathbb{N}}$ é uma sequência de erros satisfazendo

$$\langle \varepsilon^k, x^{k+1} - x^k \rangle \leqslant \alpha\|x^{k+1} - x^k\|^2 \tag{4.10}$$

e

$$\|\varepsilon^k\| \leqslant M\|x^{k+1} - x^k\| \tag{4.11}$$

para todo $k \in \mathbb{N}$.

Analisaremos a convergência do algoritmo acima com o seguinte resultado.

Teorema 14. (Método da projeção média inexata)

Sejam f *como em* (4.7) *e* $F_1, ..., F_p$ *subconjuntos semi-algébricos, prox-regulares de* $\mathbb{R}^n$ *satisfazendo*

$$\bigcap_{i=1}^{p} F_i \neq \emptyset.$$

Se x^0 *é suficientemente próximo de* $\bigcap_{i=1}^{p} F_i$*, então o algoritmo da projeção média inexata se reduz ao método do gradiente*

$$x^{k+1} = x^k - \frac{\theta}{p}\nabla f(x^k) + \varepsilon^k,$$

e gera um sequência bem definida. Além disso, essa sequência tem a propriedade de comprimento finito e converge para um ponto viável $\overline{x}$*, isto é,*

$$\overline{x} \in \bigcap_{i=1}^{p} F_i.$$

Demonstração. Como foi mostrado no Exemplo 12, a função f definida em (4.7) é semi-algébrica. Assim pelo Teorema 8, f é uma função KL.

Tome um ponto $x^* \in \bigcap_{i=1}^{p} F_i$. Pelo Teorema 2, existe $\delta > 0$ tal que, para cada $i = 1, ..., p$,

(a) a projeção P_{F_i} é um único valor na $B(x^*, \delta)$;

(b) a função $g_i = \frac{1}{2}\text{dist}(\cdot, F_i)^2$ é C^1 na $B(x^*, \delta)$ e $\nabla g_i(x) = x - P_{F_i}(x)$;

(c) a aplicação gradiente ∇g_i é 1-Lipschitz contínua na $B(x^*, \delta)$.

Uma vez que f tem a propriedade KL em torno de x^*, existem φ, U e η como na Definição 25. Diminuindo δ se necessário, podemos assumir que $B(x^*, \delta) \subset U$. Fixe $\rho \in (0, \delta)$ e escolha $\eta > 0$ tal que

$$\eta < \frac{1 - 2\alpha}{2p}(\delta - \rho)^2. \tag{4.12}$$

Escolhamos x^0 tal que: $0 = f(x^*) \leqslant f(x^0) < \eta$ e

$$\|x^* - x^0\| + 2\sqrt{\frac{f(x^0)}{a}} + \frac{b}{a}\varphi(f(x^0)) < \rho, \tag{4.13}$$

onde $a = p(\frac{1-\alpha}{\theta} - \frac{1}{2}) > 0$ (ver (4.8)) e $b = p(1 + \frac{1+M}{\theta})$.

Vamos provar por indução que $x^k \in B(x^*, \delta)$ para todo inteiro $k \geqslant 0$.

O caso $k = 0$ segue de (4.13). Note que se $x \in B(x^*, \delta)$, temos

$$\nabla f(x) = \sum_{i=1}^{p}(x - P_{F_i}(x)).$$

Daí,

$$\|\nabla f(x)\| \leqslant \sum_{i=1}^{p} \|x - P_{F_i}\|,$$

e usando a desigualdade de Cauchy-Schwarz, obtemos

$$\begin{aligned}
\|\nabla f(x)\|^2 &\leqslant \left(\sum_{i=1}^{p} \|x - P_{F_i}(x)\|\right)^2 \\
&\leqslant p\sum_{i=1}^{p} \|x - P_{F_i}(x)\|^2 \\
&\leqslant p\sum_{i=1}^{p} \operatorname{dist}(x, F_i)^2 \\
&= 2pf(x),
\end{aligned}$$

isto é,

$$\|\nabla f(x)\|^2 \leqslant 2pf(x). \tag{4.14}$$

Seja $k \geqslant 0$. Assuma agora que $x^j \in B(x^*, \rho)$ para $j \in \{0, 1, ..., k\}$. Pelo item (i) do Teorema 2, $P_{F_i}(x^k)$ é um único valor. Assim, segue de (4.9) que

$$\begin{aligned}
x^{k+1} &= (1-\theta)x^k + \theta\left(\frac{1}{p}\sum_{i=1}^{p} P_{F_i}(x^k)\right) + \varepsilon^k \\
&= (1-\theta)x^k + \theta\left(\frac{1}{p}(x^k p - \nabla f(x^k))\right) + \varepsilon^k \\
&= (1-\theta)x^k + \theta x^k - \frac{\theta}{p}\nabla f(x^k) + \varepsilon^k \\
&= x^k - \frac{\theta}{p}\nabla f(x^k) + \varepsilon^k,
\end{aligned}$$

isto é,

$$x^{k+1} = x^k - \frac{\theta}{p}\nabla f(x^k) + \varepsilon^k. \tag{4.15}$$

Portanto, x^{k+1} é unicamente definido. Vamos verificar que a propriedade $\mathcal{H}1$ é satisfeita para todo x^j, $j \in \{0, 1, ..., k+1\}$. Usando (4.10), obtemos

$$\begin{aligned}
\langle \nabla f(x^k), x^{k+1} - x^k \rangle &= \langle \frac{p}{\theta}(x^k - x^{k+1} + \varepsilon^k), x^{k+1} - x^k \rangle \\
&= \frac{p}{\theta}\Big[\langle \varepsilon^k, x^{k+1} - x^k \rangle - \|x^{k+1} - x^k\|^2\Big] \\
&\leqslant -\frac{p}{\theta}(1-\alpha)\|x^{k+1} - x^k\|^2.
\end{aligned}$$

Afirmamos que ∇f é p-Lipschitz contínuo. De fato, usando que ∇g_i é 1-Lipschitz contínuo, obtemos

$$\begin{aligned}\|\nabla f(x)-\nabla f(y)\| &= \|\sum_{i=1}^{p}(\nabla g_i(x)-\nabla g_i(y))\| \\ &\leqslant \sum_{i=1}^{p}\|\nabla g_i(x)-\nabla g_i(y)\| \\ &\leqslant \sum_{i=1}^{p}\|x-y\| = p\|x-y\|.\end{aligned}$$

Fazendo $x = x^{k+1}$ e $y = x^k$ no Lema 13, obtemos

$$f(x^{k+1}) \leqslant f(x^k) + \langle \nabla f(x^k), x^{k+1}-x^k\rangle + \frac{p}{2}\|x^{k+1}-x^k\|^2,$$

isto é,

$$f(x^{k+1}) \leqslant f(x^k) - \frac{p}{\theta}(1-\alpha)\|x^{k+1}-x^k\|^2 + \frac{p}{2}\|x^{k+1}-x^k\|^2.$$

Assim,

$$f(x^{k+1}) + p\Big(\frac{1-\alpha}{\theta}-\frac{1}{2}\Big)\|x^{k+1}-x^k\|^2 \leqslant f(x^k),$$

e portanto,

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

com $a = p(\frac{1-\alpha}{\theta}-\frac{1}{2})$.

Mostraremos agora que a condição $\mathcal{H}2$ é valida para todo x^j, $j \in \{0, 1, ..., k+1\}$. Com efeito, segue de (4.11) e (4.15) que

$$\begin{aligned}\|\nabla f(x^{k+1})\| &\leqslant \|\nabla f(x^{k+1}) - \nabla f(x^k)\| + \|\nabla f(x^k)\| \\ &\leqslant p\|x^{k+1}-x^k\| + \frac{p}{\theta}(\|x^{k+1}-x^k\| + \|\varepsilon^k\|) \\ &\leqslant p\Big(1+\frac{1+M}{\theta}\Big)\|x^{k+1}-x^k\| \\ &= b\|x^{k+1}-x^k\|,\end{aligned}$$

onde $b = p(1+\frac{1+M}{\theta})$.

Por outro lado,

$$\begin{aligned}\|x^{k+1}-x^k\|^2 - 2\langle x^{k+1}-x^k, \varepsilon^k\rangle + \|\varepsilon^k\|^2 &= \|x^{k+1}-x^k-\varepsilon^k\|^2 \\ &= \|-\frac{\theta}{p}\nabla f(x^k)\|^2 \\ &\leqslant \|\nabla f(x^k)\|^2,\end{aligned}$$

uma vez que, $\theta \in (0,1)$ e $p \geqslant 1$. Segue de (4.10) e (4.14) que

$$\begin{aligned} \|x^{k+1} - x^k\|^2 \leqslant \|x^{k+1} - x^k\|^2 + \|\varepsilon^k\|^2 &\leqslant \|\nabla f(x^k)\|^2 + 2\langle x^{k+1} - x^k, \varepsilon^k\rangle \\ &\leqslant 2pf(x^k) + 2\alpha\|x^{k+1} - x^k\|^2. \end{aligned}$$

Assim,

$$\|x^{k+1} - x^k\|^2 \leqslant \frac{2pf(x^k)}{(1-2\alpha)}.$$

Como $f(x^0) < \eta < \dfrac{1-2\alpha}{2p}(\delta - \rho)^2$, concluímos que

$$\|x^{k+1} - x^k\|^2 \leqslant \frac{2pf(x^k)}{(1-2\alpha)} < (\delta - \rho)^2. \tag{4.16}$$

Uma vez que

$$\|x^{k+1} - x^*\| \leqslant \|x^{k+1} - x^k\| + \|x^k - x^*\|,$$

segue de (4.16) e da hipótese de indução, que $x^{k+1} \in B(x^*, \delta)$.

Aplicando o Corolário 8, obtemos que $x^{k+1} \in B(x^*, \rho)$, o que completa a prova por indução.

Como consequência, o algoritmo define uma sequência única que satisfaz as hipóteses do Teorema 13, e portanto gera uma sequência de comprimento finito que converge para um ponto $\overline{x}$ tal que $f(\overline{x}) = f(x^*) = 0$. Como os conjuntos $F_1, ..., F_p$ são fechados existem $\overline{y}_i \in F_i$, tal que $\|\overline{x} - \overline{y}_i\| = 0$, para todo $i = 1, ..., p$. Logo $\overline{x} \in \bigcap_{i=1}^p F_i$. □

4.2 Algoritmo proximal inexato

Vamos recordar a versão exata do algoritmo proximal para funções não-suaves.

Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função semi-contínua inferiormente, limitada inferiormente, e λ uma parâmetro positivo. Definamos a correspondência proximal como a aplicação ponto-conjunto $\text{prox}_{\lambda f} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ definida através da fórmula

$$\text{prox}_{\lambda f}(x) := \arg\min\Big\{f(y) + \frac{1}{2\lambda}\|y - x\|^2 : y \in \mathbb{R}^n\Big\}. \tag{4.17}$$

Observação 7.

(a) Note que para algum $\mu > 0$, temos que, $\text{prox}_{\lambda(\mu f)}(x) = \text{prox}_{\mu(\lambda f)}(x)$. De fato, $z \in \text{prox}_{\lambda(\mu f)}(x)$ se, e somente se, $\forall y \in \mathbb{R}^n$

$$\mu f(z) + \frac{1}{2\lambda}\|z - x\|^2 \leqslant \mu f(y) + \frac{1}{2\lambda}\|y - x\|^2 \Leftrightarrow$$

$$\mu\Big(f(z)+\frac{1}{2\lambda\mu}\|z-x\|^2\Big)\leqslant\mu\Big(f(y)+\frac{1}{2\lambda\mu}\|y-x\|^2\Big)\Leftrightarrow$$
$$f(z)+\frac{1}{2\lambda\mu}\|z-x\|^2\leqslant f(y)+\frac{1}{2\lambda\mu}\|y-x\|^2\Leftrightarrow$$
$$\lambda f(z)+\frac{1}{2\mu}\|z-x\|^2\leqslant\lambda f(y)+\frac{1}{2\mu}\|y-x\|^2\Leftrightarrow$$

$z\in\operatorname{prox}_{\mu(\lambda f)}(x)$. Dessa forma denotaremos $\operatorname{prox}_{\lambda(\mu f)}(x)$ ou $(\operatorname{prox}_{\mu(\lambda f)}(x))$, por $\operatorname{prox}_{\lambda\mu f}(x)$.

(b) Uma vez que f é limitada inferiormente, isto é, existe $C\in\mathbb{R}$ tal que $C<f(x)$, o conjunto $\operatorname{prox}_{\lambda f}(x)$ é não-vazio. De fato, se $\|x^k\|\to+\infty$ e usando a desigualdade de Cauchy-Schwarz, obtemos

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

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

Seja $x^0\in\mathbb{R}^n$. O clássico algoritmo proximal se escreve

$$x^{k+1}\in\operatorname{prox}_{\lambda_k f}(x^k), \tag{4.18}$$

onde λ_k é uma sequência de passos tal que $\lambda_k\in[\underline{\lambda},\overline{\lambda}]\subset(0,+\infty)$. Afirmamos que a sequência $(x^k)_{k\in\mathbb{N}}$ gerada pelo algoritmo (4.18) satisfaz as seguintes condições:

(i) Para cada $k\in\mathbb{N}$,
$$f(x^{k+1})+\frac{1}{2\lambda_k}\|x^{k+1}-x^k\|^2\leqslant f(x^k). \tag{4.19}$$

(ii) Para cada $k\in\mathbb{N}$, existe $w^{k+1}\in\partial f(x^{k+1})$ tal que
$$\lambda_k w^{k+1}+x^{k+1}-x^k=0. \tag{4.20}$$

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

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

Em particular se $y=x^k$, obtemos (4.19).

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 3 que

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

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

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

4.2.1 Convergência de um algoritmo proximal inexato para funções KL

Vamos introduzir uma versão do método proximal inexato.

Considere uma sequência $(x^k)_{k\in\mathbb{N}}$ gerada pelo seguinte algoritmo:

Algoritmo 2. *Fixe* $x^0 \in \mathbb{R}^n$, $0 < \underline{\lambda} \leqslant \overline{\lambda} < \infty$, $0 \leqslant \sigma < 1$, $0 < \theta \leqslant 1$. *Para cada* $k = 0, 1, ...,$ *escolha* $\lambda_k \in [\underline{\lambda}, \overline{\lambda}]$, *e encontre* $x^{k+1} \in \mathbb{R}^n$, $w^{k+1} \in \mathbb{R}^n$ *tal que*

$$f(x^{k+1}) + \frac{\theta}{2\lambda_k}\|x^{k+1} - x^k\|^2 \leqslant f(x^k), \tag{4.21}$$

$$w^{k+1} \in \partial f(x^{k+1}) \tag{4.22}$$

e

$$\|\lambda_k w^{k+1} + x^{k+1} - x^k\|^2 \leqslant \sigma(\|\lambda_k w^{k+1}\|^2 + \|x^{k+1} - x^k\|^2). \tag{4.23}$$

Como foi visto acima, o algoritmo definido em (4.18) é uma caso particular do Algoritmo 2.

Os lemas seguintes serão úteis para a análise de convergência do Algoritmo 2.

Lema 14. *Seja* $\sigma \in (0, 1]$. *Se* $x, y \in \mathbb{R}^n$, *e*

$$\|x + y\|^2 \leqslant \sigma(\|x\|^2 + \|y\|^2), \tag{4.24}$$

então

$$\frac{1-\sigma}{2}(\|x\|^2 + \|y\|^2) \leqslant -\langle x, y\rangle. \tag{4.25}$$

Se $\sigma \in (0, 1)$, *então*

$$\frac{1 - \sqrt{1 - (1-\sigma)^2}}{1-\sigma}\|y\| \leqslant \|x\| \leqslant \frac{1 + \sqrt{1 - (1-\sigma)^2}}{1-\sigma}\|y\|. \tag{4.26}$$

Demonstração. Por (4.24), se $x, y \in \mathbb{R}^n$, então

$$\|x\|^2 + 2\langle x, y\rangle + \|y\|^2 = \|x + y\|^2 \leqslant \sigma(\|x\|^2 + \|y\|^2).$$

Ou seja,

$$\frac{1-\sigma}{2}(\|x\|^2 + \|y\|^2) \leqslant -\langle x, y\rangle.$$

Suponha agora que $\sigma \in (0, 1)$. Usando (4.25) e a desigualdade de Cauchy-Schwarz, obtemos

$$\frac{1-\sigma}{2}(\|x\|^2 + \|y\|^2) \leqslant -\langle x, y\rangle \leqslant \|x\|\|y\|,$$

isto é,

$$(1-\sigma)\|x\|^2 - 2\|y\|\|x\| + (1-\sigma)\|y\|^2 \leqslant 0.$$

Resolvendo a inequação quadrática acima com variável $\|x\|$, obtemos (4.26). □

Lema 15. *Nas condições do Algoritmo 2, existe* $b > 0$ *tal que*

$$\|w^{k+1}\| \leqslant b\|x^{k+1} - x^k\|. \tag{4.27}$$

Demonstração. De (4.23), obtemos

$$\|\lambda_k w^{k+1} + x^{k+1} - x^k\|^2 \leqslant \sigma(\|\lambda_k w^{k+1}\|^2 + \|x^{k+1} - x^k\|^2).$$

Assim, segue de (4.26) que

$$\underline{\lambda}\|w^{k+1}\| \leqslant \|\lambda_k w^{k+1}\| \leqslant \frac{1 + \sqrt{1-(1-\sigma)^2}}{1-\sigma}\|x^{k+1} - x^k\|.$$

Portanto,

$$\|w^{k+1}\| \leqslant b\|x^{k+1} - x^k\|,$$

com $b = \dfrac{1 + \sqrt{1-(1-\sigma)^2}}{\underline{\lambda}(1-\sigma)}$. □

O principal resultado desta seção é o seguinte teorema.

Teorema 15. (Algoritmo proximal inexato)

Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *uma função KL limitada inferiormente. Assuma que a restrição de* f *ao seu domínio efetivo é uma função contínua. Se a sequência* $(x^k)_{k\in\mathbb{N}}$ *gerada pelo Algoritmo 2 (ou por* (4.21), (4.22) *e* (4.27)) *é limitada, então converge para algum ponto crítico* $\bar{x}$ *de* f. *Além disso, a sequência* $(x^k)_{k\in\mathbb{N}}$ *tem comprimento finito, isto é,* $\sum_k \|x^{k+1} - x^k\| < +\infty$.

Demonstração. Se $(x^k)_{k\in\mathbb{N}}$ é uma sequência gerada pelo Algoritmo 2 (ou por (4.21), (4.22) e (4.27)), então $(x^k)_{k\in\mathbb{N}}$ satisfaz as condições $\mathcal{H}1$ e $\mathcal{H}2$ (a condição $\mathcal{H}2$ segue do Lema 15). Como $(x^k)_{k\in\mathbb{N}}$ é limitada, existem $(x^{k_j})_{j\in\mathbb{N}}$ e $\bar{x}$ tal que

$$x^{k_j} \to \bar{x} \quad \text{quando} \quad j \to \infty.$$

Uma vez que f é contínua no seu domínio efetivo e $f(x^{k_j}) \leqslant f(x^0) < +\infty$ para todo j, concluimos que

$$f(x^{k_j}) \to f(\bar{x}) \quad \text{quando} \quad j \to \infty.$$

O resultado segue aplicando o Teorema 10. □

Quando a função em questão é convexa e satisfaz a propriedade Kurdyka-Lojasiewicz, o Algoritmo 2 pode ser simplificado, enquanto suas propriedades de convergência são mantidas.

Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convexa, semi-contínua inferiormente. Considere a sequência $(x^k)_{k\in\mathbb{N}}$ gerada pelo seguinte algoritmo.

Algoritmo 2bis. Fixe $x^0 \in \mathbb{R}^n$, $0 < \underline{\lambda} \leqslant \bar{\lambda} < \infty$, $0 \leqslant \sigma < 1$.

Para cada $k = 0, 1, ...$, escolha $\lambda_k \in [\underline{\lambda}, \bar{\lambda}]$ e encontre $x^{k+1} \in \mathbb{R}^n$, $w^{k+1} \in \mathbb{R}^n$ tal que

$$w^{k+1} \in \partial f(x^{k+1}), \tag{4.28}$$

$$\|\lambda_k w^{k+1} + x^{k+1} - x^k\|^2 \leqslant \sigma(\|\lambda_k w^{k+1}\|^2 + \|x^{k+1} - x^k\|^2). \tag{4.29}$$

Antes de provarmos o principal resultado desta seção, estabeleceremos a seguinte lema.

Lema 16. *Nas condições do Algoritmo 2bis, para cada k,*

$$f(x^{k+1}) + \frac{1-\sigma}{2\underline{\lambda}}\|x^{k+1} - x^k\|^2 + \frac{(1-\sigma)\bar{\lambda}}{2}\|w^{k+1}\|^2 \leqslant f(x^k). \tag{4.30}$$

Demonstração. Como $w^{k+1} \in \partial f(x^{k+1})$, então

$$f(z) \geqslant f(x^{k+1}) + \langle w^{k+1}, z - x^{k+1}\rangle \quad \forall z \in \mathbb{R}^n.$$

Em particular se $z = x^k$, obtemos

$$f(x^k) \geqslant f(x^{k+1}) + \langle w^{k+1}, x^k - x^{k+1}\rangle.$$

Note que

$$\langle w^{k+1}, x^k - x^{k+1}\rangle = \frac{1}{2\lambda_k}\Big[\|\lambda_k w^{k+1}\|^2 + \|x^{k+1} - x^k\|^2 - \|\lambda_k w^{k+1} + x^{k+1} - x^k\|^2\Big].$$

Como $\lambda_k \in [\underline{\lambda}, \bar{\lambda}]$ e usando que (4.29), obtemos

$$\begin{aligned}
f(x^k) - f(x^{k+1}) &\geqslant \langle w^{k+1}, x^k - x^{k+1}\rangle \\
&= \frac{1}{2\lambda_k}\Big[\|\lambda_k w^{k+1}\|^2 + \|x^{k+1} - x^k\|^2 - \|\lambda_k w^{k+1} + x^{k+1} - x^k\|^2\Big] \\
&\geqslant \frac{1}{2\lambda_k}\Big[\|\lambda_k w^{k+1}\|^2 + \|x^{k+1} - x^k\|^2 - \sigma(\|\lambda_k w^{k+1}\|^2 + \|x^{k+1} - x^k\|^2)\Big] \\
&= \frac{(1-\sigma)}{2}\Big[\lambda_k\|w^{k+1}\|^2 + \frac{\|x^{k+1} - x^k\|^2}{\lambda_k}\Big] \\
&\geqslant \frac{(1-\sigma)}{2}\Big[\underline{\lambda}\|w^{k+1}\|^2 + \frac{\|x^{k+1} - x^k\|^2}{\bar{\lambda}}\Big].
\end{aligned}$$

Portanto,

$$\mathsf{f}(x^{k+1}) + \frac{1-\sigma}{2\underline{\lambda}}\|x^{k+1} - x^k\|^2 + \frac{(1-\sigma)\overline{\lambda}}{2}\|w^{k+1}\|^2 \leqslant \mathsf{f}(x^k).$$

□

Teorema 16. *Seja* $\mathsf{f} : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ *uma função semi-contínua inferiormente. Assuma que* f *é uma função KL o qual é limitada inferiormente e seja* $(x^k)_{k\in\mathbb{N}}$ *uma sequência gerada pelo Algoritmo 2bis.*

Se $(x^k)_{k\in\mathbb{N}}$ *é limitada, então converge para um minimizador de* f *e a sequência de valores* $\mathsf{f}(x^k)$ *converge para o valor* $\min \mathrm{f}$. *Além disso, a sequência tem comprimento finito, isto é,* $\sum_k \|x^{k+1} - x^k\| < +\infty$.

Demonstração. Provaremos inicialmente que

$$\sum_{k=0}^{+\infty} \|w^k\|^2 < +\infty. \tag{4.31}$$

De fato, como f é limitada inferiormente, existe $C \in \mathbb{R}$ tal que $C < \mathsf{f}(x)$ para todo $x \in \mathbb{R}^n$. Por outro lado segue de (4.30) que

$$\begin{aligned} \mathsf{f}(x^{k+1}) + \frac{(1-\sigma)\overline{\lambda}}{2}\|w^{k+1}\|^2 &\leqslant \mathsf{f}(x^{k+1}) + \frac{1-\sigma}{2\underline{\lambda}}\|x^{k+1} - x^k\|^2 + \frac{(1-\sigma)\overline{\lambda}}{2}\|w^{k+1}\|^2 \\ &\leqslant \mathsf{f}(x^k). \end{aligned}$$

Logo,

$$\frac{(1-\sigma)\overline{\lambda}}{2}\|w^{k+1}\|^2 \leqslant \mathsf{f}(x^k) - \mathsf{f}(x^{k+1}),$$

e consequentemente,

$$\frac{(1-\sigma)\overline{\lambda}}{2}\sum_{j=0}^{k} \|w^{j+1}\|^2 \leqslant \sum_{j=0}^{k}(\mathsf{f}(x^j) - \mathsf{f}(x^{j+1})) = \mathsf{f}(x^0) - \mathsf{f}(x^{k+1}),$$

ou seja,

$$C + \frac{(1-\sigma)\overline{\lambda}}{2}\sum_{j=0}^{k} \|w^{j+1}\|^2 \leqslant \mathsf{f}(x^{k+1}) + \frac{(1-\sigma)\overline{\lambda}}{2}\sum_{j=0}^{k} \|w^{j+1}\|^2 \leqslant \mathsf{f}(x^0),$$

isto é,

$$\sum_{k=0}^{+\infty} \|w^k\|^2 < +\infty.$$

Portanto segue de (4.31) que

$$w^k \to 0 \ \text{ quando } \ k \to +\infty.$$

Uma vez que (x^k) é limitada, existe uma subsequência $(x^{k_j})_{j\in\mathbb{N}}$ e um ponto $\overline{x}$ tal que, (x^{k_j}) converge para $\overline{x}$, quando $j \to +\infty$. Por (4.30), segue que a sequência $\mathsf{f}(x^k)_{k\in\mathbb{N}}$ é não-crescente. Por outro lado, como f é semi-contínua inferiormente, temos

$$f(\overline{x}) \leqslant \liminf_{j\to\infty} f(x^{k_j}) = \lim_{k\to\infty} f(x^k).$$

Como f é convexa e $w^k \in \partial f(x^k)$ para $k \geqslant 1$, obtemos

$$f(z) \geqslant f(x^{k_j}) + \langle w^{k_j}, z - x^{k_j}\rangle \quad \forall z \in \mathbb{R}^n.$$

Em particular se $z = \overline{x}$, temos

$$f(\overline{x}) \geqslant f(x^{k_j}) + \langle w^{k_j}, \overline{x} - x^{k_j}\rangle.$$

Passando o limite quando $j \to \infty$, concluimos que

$$f(\overline{x}) \geqslant \lim_{j\to\infty} f(x^{k_j}) = \lim_{k\to\infty} f(x^k).$$

Portanto,

$$\lim_{k\to\infty} f(x^k) = f(\overline{x}).$$

Usando (4.27) e (4.30), podemos aplicar o Teorema 10 para obter o resultado. □

4.3 Algoritmo forward-backward inexato

Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função semi-contínua inferiormente, limitada inferiormente e que satisfaz a propriedade Kurdyka-Łojasiewicz.

Assumimos que f é uma função estruturada que pode ser dividida como

$$f = g + h, \tag{4.32}$$

onde $h : \mathbb{R}^n \to \mathbb{R}$ é uma função de classe C^1 com ∇h L-Lipschitz contínuo.

Consideraremos sequências geradas de acordo com o seguinte algoritmo:

Algoritmo 3. *Fixe* $a, b > 0$ *com* $a > L$, *e* $x^0 \in \operatorname{dom} g$.
Para $k = 0, 1, ...,$ *encontre* $x^{k+1} \in \mathbb{R}^n$, $v^{k+1} \in \mathbb{R}^n$ *tal que*

$$g(x^{k+1}) + \langle x^{k+1} - x^k, \nabla h(x^k)\rangle + \frac{a}{2}\|x^{k+1} - x^k\|^2 \leqslant g(x^k); \tag{4.33}$$

$$v^{k+1} \in \partial g(x^{k+1}); \tag{4.34}$$

$$\|v^{k+1} + \nabla h(x^k)\| \leqslant b\|x^{k+1} - x^k\|. \tag{4.35}$$

4.3.1 Algoritmo da separação forward-backward para funções não-convexas

Vamos supor que g é limitada inferiormente. Sendo dada uma sequência de parâmetros positivos λ_k que satisfaz

$$0 < \underline{\lambda} < \lambda_k < \overline{\lambda} < \frac{1}{L},$$

onde $\underline{\lambda}$ e $\overline{\lambda}$ são constantes dadas. O algoritmo da separação forward-backward é dado por

$$x^{k+1} \in \text{prox}_{\lambda_k g}(x^k - \lambda_k \nabla h(x^k)). \tag{4.36}$$

Vamos mostrar que esse algoritmo se encaixa na estrutura geral do Algoritmo 3.

De fato, como $x^{k+1} \in \text{prox}_{\lambda_k g}(x^k - \lambda_k \nabla h(x^k))$,

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

Em particular, se $y = x^k$, obtemos

$$g(x^{k+1}) + \frac{1}{2\lambda_k}\|x^{k+1} - x^k + \lambda_k \nabla h(x^k)\|^2 \leqslant g(x^k) + \frac{1}{2\lambda_k}\|\lambda_k \nabla h(x^k)\|^2,$$

isto é,

$$\lambda_k g(x^{k+1}) + \frac{1}{2}\|x^{k+1} - x^k + \lambda_k \nabla h(x^k)\|^2 \leqslant \lambda_k g(x^k) + \frac{1}{2}\|\lambda_k \nabla h(x^k)\|^2. \tag{4.37}$$

Sabemos que

$$\|x^{k+1}-x^k+\lambda_k \nabla h(x^k)\|^2 = \|x^{k+1}-x^k\|^2+\|\lambda_k \nabla h(x^k)\|^2+2\lambda_k\langle x^{k+1}-x^k, \nabla h(x^k)\rangle. \tag{4.38}$$

Logo substituindo (4.38) em (4.37), obtemos

$$\lambda_k g(x^{k+1}) + \frac{1}{2}\|x^{k+1} - x^k\|^2 + \lambda_k\langle x^{k+1} - x^k, \nabla h(x^k)\rangle \leqslant \lambda_k g(x^k),$$

ou seja,

$$g(x^{k+1}) + \frac{1}{2\lambda_k}\|x^{k+1} - x^k\|^2 + \langle x^{k+1} - x^k, \nabla h(x^k)\rangle \leqslant g(x^k).$$

Assim,

$$g(x^{k+1}) + \langle x^{k+1} - x^k, \nabla h(x^k)\rangle + \frac{a}{2}\|x^{k+1} - x^k\|^2 \leqslant g(x^k),$$

com $a = \dfrac{1}{\overline{\lambda}}$.

Por outro lado, para cada $k \in \mathbb{N}$, defina $v^{k+1} = \dfrac{x^k - x^{k+1} - \lambda_k \nabla h(x^k)}{\lambda_k}$. Afirmamos que $v^{k+1} \in \partial g(x^{k+1})$ e que existe $b > 0$ tal que $\|v^{k+1} + \nabla h(x^k)\| \leqslant b\|x^{k+1} - x^k\|$.

De fato, como $x^{k+1} \in \text{prox}_{\lambda_k g}(x^k - \lambda_k \nabla h(x^k))$, segue da Proposiação 3 que

$$0 \in \partial\Big(g(x^{k+1}) + \frac{1}{2\lambda_k}\|x^{k+1} - x^k + \lambda_k \nabla h(x^k)\|^2\Big) = \partial g(x^{k+1}) + \frac{x^{k+1} - x^k + \lambda_k \nabla h(x^k)}{\lambda_k},$$

ou seja,

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

Portanto, $v^{k+1} \in \partial g(x^{k+1})$.

Além disso, $\lambda_k v^{k+1} + \lambda_k \nabla h(x^k) + x^{k+1} - x^k = 0$. Daí,

$$\|v^{k+1} + \nabla h(x^k)\| = \frac{1}{\lambda_k}\|x^{k+1} - x^k\| \leqslant b\|x^{k+1} - x^k\|,$$

onde $b = \dfrac{1}{\underline{\lambda}}$.

4.3.2 Convergência do algoritmo da separação forward-backward inexato

Vamos mostrar o seguinte resultado de convergência para o Algoritmo 3.

Teorema 17. (Nonconvex nonsmooth forward-backward splitting)
Seja $f = g + h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *uma função KL limitada inferiormente. Suponha ainda que* $h : \mathbb{R}^n \to \mathbb{R}$ *tem valores finitos, é diferenciável, que* ∇h *seja* L*-Lipschitz contínuo, e que a restrição de* g *ao seu domínio efetivo seja contínua.*

Se $(x^k)_{k\in\mathbb{N}}$ *é uma sequência limitada gerada pelo Algoritmo 3, então converge para algum ponto crítico de* $f = g + h$. *Alêm disso, a sequência* $(x^k)_{k\in\mathbb{N}}$ *tem comprimento finito, isto é,* $\sum_k \|x^{k+1} - x^k\| < +\infty$.

Demonstração. Usando o Lema 13 para a função h com $x = x^{k+1}$ e $y = x^k$, e a propriedade (4.33) do Algoritmo 3, obtemos

$$h(x^{k+1}) \leqslant h(x^k) + \langle \nabla h(x^k), x^{k+1} - x^k \rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2$$

e

$$\langle x^{k+1} - x^k, \nabla h(x^k) \rangle \leqslant g(x^k) - g(x^{k+1}) - \frac{a}{2}\|x^{k+1} - x^k\|^2.$$

Daí,

$$g(x^{k+1}) + h(x^{k+1}) + \frac{a - L}{2}\|x^{k+1} - x^k\|^2 \leqslant g(x^k) + h(x^k),$$

e portanto,

$$f(x^{k+1}) + \frac{a - L}{2}\|x^{k+1} - x^k\|^2 \leqslant f(x^k).$$

Para cada $k \in \mathbb{N}$, defina $w^{k+1} = v^{k+1} + \nabla h(x^{k+1})$. Daí,

$$w^{k+1} = v^{k+1} + \nabla h(x^{k+1}) \in \partial g(x^{k+1}) + \nabla h(x^{k+1}) = \partial f(x^{k+1}).$$

Logo, $w^{k+1} \in \partial f(x^{k+1})$. Além disso, pela propriedade (4.35) do Algoritmo 3, pela desigualdade triângular e usando o fato de ∇h ser L-Lipschitz contínuo, obtemos

$$\begin{aligned} \|w^{k+1}\| &= \|v^{k+1} + \nabla h(x^{k+1})\| \\ &\leqslant \|v^{k+1} + \nabla h(x^k)\| + \|\nabla h(x^{k+1}) - \nabla h(x^k)\| \\ &\leqslant b\|x^{k+1} - x^k\| + L\|x^{k+1} - x^k\| \\ &= (b+L)\|x^{k+1} - x^k\|. \end{aligned}$$

Por outro lado, a condição $\mathcal{H}3$ é satisfeita trivialmente, uma vez que, $(x^k)_{k\in\mathbb{N}}$ é limitada, h é diferenciável, e g é contínua restrita ao seu domínio efetivo. Basta então aplicarmos o Teorema 10 para obtermos o resultado. □

Observação 8. *No Teorema 17, a única hipótese sobre a função g não é necessária. Isto é, g não precisa ser contínua em seu domínio efetivo. Mostraremos que é válida a condição* $\mathcal{H}3$ sem essa hipótese. De fato, para todo $y \in \mathbb{R}^n$ temos

$$g(x^{k+1}) + \frac{1}{2\lambda_k}\|x^{k+1} - x^k + \lambda_k \nabla h(x^k)\|^2 \leqslant g(y) + \frac{1}{2\lambda_k}\|y - x^k + \lambda_k \nabla h(x^k)\|^2.$$

Usando que $\|x+y\|^2 = \|x\|^2 + \|y\|^2 + \langle x, y\rangle$ para todo $x, y \in \mathbb{R}^n$, obtemos

$$g(x^{k+1}) + \frac{1}{2\lambda_k}\|x^{k+1} - x^k\|^2 + \langle x^{k+1} - x^k, \nabla h(x^k)\rangle \leqslant g(y) + \frac{1}{2\lambda_k}\|y - x^k\| + \langle y - x^k, \nabla h(x^k)\rangle.$$

Por hipótese, existe uma subsequência $(x^{k_j})_{j\in\mathbb{N}}$ de $(x^k)_{k\in\mathbb{N}}$ que converge para algum $\bar{x}$. Tomando $y = \bar{x}$, temos que

$$g(x^{k_j+1}) + \frac{1}{2\lambda_{k_j}}\|x^{k_j+1} - x^{k_j}\|^2 + \langle x^{k_j+1} - \bar{x}, \nabla h(x^{k_j})\rangle \leqslant g(\bar{x}) + \frac{1}{2\lambda_{k_j}}\|\bar{x} - x^{k_j}\|.$$

Fazendo $j \to \infty$, obtemos

$$\limsup_{j\to\infty} g(x^{k_j+1}) \leqslant g(\bar{x}).$$

Além disso, como f é semi-contínua inferiormente

$$\begin{aligned} \liminf_{j\to\infty} g(x^{k_j+1}) &= \liminf_{j\to\infty}(f(x^{k_j+1}) - h(x^{k_j+1})) \\ &\geqslant \liminf_{j\to\infty} f(x^{k_j+1}) + \liminf_{j\to\infty}(-h(x^{k_j+1})) \\ &= \liminf_{j\to\infty} f(x^{k_j+1}) - \limsup_{j\to\infty}(h(x^{k_j+1})) \\ &\geqslant f(\bar{x}) - h(\bar{x}) = g(\bar{x}). \end{aligned}$$

Logo,

$$g(\bar{x}) \leqslant \liminf_{j\to\infty} g(x^{k_j+1}) \leqslant \limsup_{j\to\infty} g(x^{k_j+1}) \leqslant g(\bar{x}),$$

e portanto,

$$\lim_{j \to \infty} g(x^{k_j+1}) = g(\overline{x}).$$

Logo segue a conclusão do Teorema 17.

Referências Bibliográficas

[1] Absil, P.-A., Mahony, R. , Andrews, B., *Convergence of the iterates of descent methods for analytic cost functions.* SIAM J. Optim., 16, no. 2, (2005), 531-547.

[2] Attouch, Hedy ; Bolte, Jérôme ; Svaiter, Benar Fux . Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward backward splitting, and regularized Gauss Seidel methods. Mathematical Programming, p. 1-35, 2011.

[3] Attouch, H., Bolte, J., Redont, P., Soubeyran, A. *Proximal alternating minimization and projection methods for nonconvex problems.* An approach based on the Kurdyka-Lojasiewicz inequality, Mathematics of Operations Research, 35, no. 2, (2010), 438-457.

[4] Bierstone, E., Milman, P.D. *Semianalytic and subanalytic sets.* Publicactions Mathématiques, 67 (1988), 5-42.

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

[6] Bolte, J., Daniilidis, A., Ley, O., Mazet, L. *Characterizations of Lojasiewicz inequalities: Subgradient flows, talweg, convexity.* Trans. Amer. Math. Soc., 362, (2010), 3319-3363.

[7] Chill, R. *On the Lojasiewicz-Simon gradient inequality.* Journal of Functional Analysis 201. (2003). 572-601.

[8] Graña Drummond, L. M., Peterzil, Y. *The Central Path in smooth convex semidefinite programs.* Optimization, 2002, Vol. 51(2), pp. 207-233.

[9] Ismailov, A., Solodov, M. *Otimização - volume 1 - Condções de Otimalidade, Elementos de Análise Convexa e de Dualidade.* IMPA, Rio de Janeiro. 2005.

[10] Kurdyka, K. *Points d'un ensemble sous-analytique. Ann. Inst. Fourier*, 38, 1988, pp. 135-156.

[11] Kurdyka, K. *On gradients of functions definable in o-minimal structures.* Ann. Inst. Fourier, 48, (1998), 769-783.

[12] Lewis, A.S., Luke, D.R., Malick, J. *Local linear convergence for alternating and averaged nonconvex projections.* Found. Comput. Math. 9, (2009), 485-513.

[13] Łojasiewicz, S. *Une propriété topologique des sous-ensembles analytiques réls, in: Les équations aux Dérivées Partielles.* pp. 87-89, éditions du centre National de la Recherche Scientifique, Paris 1963.

[14] Poliquin, R.A., Rockafellar, R.T., Thibault, L. *Local differentiability of distance functions.* Trans, 352, (2000), 5231-5249.

[15] Rockafellar, R.T. , Wets, R. *Variational Analysis.* Grundlehren der Mathematischen Wis- senschaften, 317, Springer, 1998.