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

Minimização Proximal Alternada e Métodos de Projeção para Problemas Não-Convexos: Uma Abordagem Baseada na Desigualdade de Kurdyka-Łojasiewicz

Edvaldo Elias de Almeida Batista

Teresina - 2012

Edvaldo Elias de Almeida Batista

Dissertação de Mestrado:

Minimização Proximal Alternada e Métodos de Projeção para Problemas Não-Convexos: Uma Abordagem Baseada na Desigualdade 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

Batista, E A B.

xxxx Minimização Proximal Alternada e Métodos
de Projeção para Problemas Não-Convexos: Uma Abordagem
Baseada na Desigualdade de Kurdyka-Łojasiewicz.

Edvaldo Elias de Almeida Batista – Teresina: 2012.

Orientador: Prof. Dr. João Xavier da Cruz Neto.

1. Otimização Matemática

CDD 516.36

Dedico este trabalho a meus pais, Carlos(In memoriam) e Gleice, minha avó Dalva, meus irmãos, Lívia e Adriano, e a minha noiva Luama.

Agradecimentos

Agradeço a Deus pelo enorme privilégio de nascer e crescer em Corrente-PI no seio de uma família muito rica em amor, carinho, respeito, honestidade, solidariedade, humildade, união, alegria e fé, cercado por pessoas maravilhosas com quem dividi experiências memoráveis. A partir desta origem modesta que me tornei um homem simples, objetivo, sonhador, corajoso, perseverante, otimista e bem-humorado. Agradeço ainda a Deus por guiar meus passos e conduzir-me até Teresina-PI, onde pude iniciar minha vida acadêmica, fazer novos amigos e conheçer a mulher com quem hoje sonho em dividir cada um dos meus dias.

A minha avó Dalva e a minha mãe Gleice pela educação, amor, carinho, paciência, respeito, companheirismo e dedicação que me prestaram por todos esses anos.

Ao meu pai Carlos, que mesmo morando distante se fez presente em muitos momentos da minha vida, por todo amor e admiração. Espero que continue se orgulhando do seu filho e zelando por ele daí de cima.

Aos meus irmãos Lívia e Adriano e a minha noiva Luama pelo amor, carinho, companheirismo, admiração, tolerância, cumplicidade e respeito.

Aos velhos e bons amigos de Corrente, em especial Edson Filho, Heitor, Ítalo, Marcos dos Santos, Pedro Henrique e Raimundo Vítor, pelo companheirismo, afeto, camaradagem, cumplicidade e lealdade. Aos professores do Colégio São José, em especial a Marcelo Coelho, principal responsável por eu ter escolhido essa carreira, pela amizade, motivação e dedicação. Vocês foram fundamentais pra formação das minhas ideologias e do meu caráter.

Aos colegas e amigos da UFPI, em especial Alex, Dênis, Diogo, Edilson, George, Jenilson, Kelson, Kim, Ramon, Valdinês e Yuri, pela consideração, afeto e ajuda em todos momentos, quer sejam de estudos ou não. Aos professores da UFPI, em especial Gilvan (obrigado pela motivação e confiança depositada!), João Xavier, Jurandir, Paulo Alexandre e Paulo Sérgio, por todo apoio moral e matemático necessários para a concretização deste trabalho e da minha formação, assim como pela amizade e solidariedade (valeu por comprar a camisa P. A.!).

Ao meu orientador, professor Xavier, pela paciência, confiança, críticas, motivação e amizade. Aos professores Paulo Sérgio e Glaydston por aceitar o convite de participar da minha banca.

A CAPES pelo apoio financeiro.

"Ninguém pode descobrir novos caminhos
até que tenha coragem de perder de vista
a terra firme.".

André Gide.

Resumo

O trabalho aqui desenvolvido baseia-se no artigo [1]. Estudamos as propriedades de convergência do algoritmo de minimização proximal alternado para funções estruturadas não-convexas do tipo: $L(x, y) = f(x) + Q(x, y) + g(y)$, onde f e g são funções próprias semicontínuas inferiormente, definidas sobre espaços Euclideanos, e Q é uma função suave que acopla as variáveis x e y. O algoritmo pode ser visto como uma regularização proximal do usual método de Gauss-Seidel para minimizar L.

Trabalhamos em um cenário não-convexo, assumindo apenas que L satisfaz a desigualdade de Kurdyka-Łojasiewicz. Exibimos várias estruturas onde uma função possui esta desigualdade válida, além de fornecer uma prova para o caso analítico (veja Teorema 6).

Nosso resultado principal pode ser assim enunciado: Se L satisfaz a propriedade Kurdyka-Łojasiewicz, então cada sequência gerada pelo algoritmo converge a um ponto crítico de L. Este resultado é completado pelo estudo da taxa de convergência do algoritmo, que depende das propriedades geométricas do função L perto de seu ponto crítico. Quando especializado para $Q(x, y) = \| x - y \|^2$, e para f, g funções indicadoras, o algoritmo é um método de projeção alternado (uma variação de von Neumann's) que converge para uma ampla classe de conjuntos.

Abstract

The work developed here is based on article [1]. We study the convergence properties of an alternating proximal minimization algorithm for nonconvex structured functions of the type: $L(x, y) = f(x) + Q(x, y) + g(y)$, where f and g are proper lower semicontinuous functions, defined on Euclidean spaces, and Q is a smooth function that couples the variables x and y. The algorithm can be viewed as a proximal regularization of the usual Gauss-Seidel method to minimize L.

We work in a nonconvex setting, just assuming that the function L satisfies the Kurdyka-£ojasiewicz inequality. We display several structures where this function has a valid inequality, and provides a proof for the analytic case (see Theorem 6).

Our main result can be stated as follows: If L has the Kurdyka-£ojasiewicz property, then each bounded sequence generated by the algorithm converges to a critical point of L. This result is completed by the study of the convergence rate of the algorithm, which depends on the geometrical properties of the function L around its critical points. When specialized to $Q(x, y) = \| x - y \|^2$, and to f, g indicator functions, the algorithm is an alternating projection method (a variant of von Neumann's) that converges for a wide class of sets.

Sumário

Resumo v

Abstract vi

1 Noções Preliminares 4

1.1 Fatos elementares de análise não-suave 4

1.2 Conjuntos algébricos e semi-algébricos 8

1.3 Conjuntos analíticos, semi-analíticos e subanalíticos 12

1.4 Funções KL 14

1.4.1 Estrutura o-minimal 16

2 Algoritmos de minimização proximal alternada 23

2.1 Convergência a um valor crítico 23

2.2 Convergência a um ponto crítico e outros resultados de convergência 27

2.3 Convergência de métodos de projeção alternados 33

Referências Bibliográficas 34

Introdução

1. Apresentação do Algoritmo. Nesta dissertação analisaremos a convergência do algoritmo de minimização alternada para funções (não-convexas) $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ do seguinte tipo:

$$(H)\begin{cases} L(x,y) = f(x) + Q(x,y) + g(y), \\ f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}, g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\} \text{ são próprias semicontínuas inferiormente} \\ Q : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \text{ é uma função de classe } C^1. \\ \nabla Q \text{ é Lipschitz contínuo sobre subconjuntos limitados de } \mathbb{R}^n \times \mathbb{R}^m. \end{cases}$$

A hipótese (H) será necessária ao longo da dissertação.

Queremos encontrar os pontos críticos de

$$L(x,y) = f(x) + Q(x,y) + g(y) \tag{1}$$

e possivelmente resolver o problema de minimização correspondente.

A estrutura específica de (L) permite, em particular, atacar problemas da forma

$$\min\{f(z) + g(z) : z \in \mathbb{R}^n\} \tag{2}$$

Com efeito, basta definir $L_p(x,y) = f(x) + (\rho/2) \parallel x - y \parallel^2 + g(y)$, ρ sendo um parâmetro positivo de penalização (ou relaxamento), e minimizar L_p sobre $\mathbb{R}^n \times \mathbb{R}^m$. Problemas de viabilidade envolvendo dois conjuntos fechados são casos particulares: basta tomar f e g funções indicadoras.

Minimizar a soma de funções ou encontrar um ponto comum a uma coleção de conjuntos fechados é um campo de pesquisa muito ativo, com aplicações em teoria da aproximação (von Neumann [23]), reconstrução de imagem (Combettes e Wajs [12], Donoho [15]), estatística (Csiszár e Tusnády [13], Grubisié e Pietersz [16]), equações diferencias parciais e controle ótimo (Lions [19], Widrow e Wallach [24]). Uma boa referência para problemas envolvendo casos convexos é Combettes e Wajs [12]: muitos exemplos vindos de problemas de processamento de sinal são mostrados regraváveis.

A especificidade da nossa abordagem é dupla. Primeiro, trabalhamos em um cenário não-convexo, assumindo apenas que a função L satisfaz a desigualdade de Kurdyka-Łojasiewicz. Mais precisamente, para cada $x^* \in \operatorname{dom} \partial L$ 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 $s \in (0, \eta)$, $\varphi'(s) > 0$,

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

$$\varphi'(L(x) - L(x^*))\operatorname{dist}(0, \partial f(x)) \geqslant 1, \tag{3}$$

onde

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

S. Łojasiewicz provou em 1963 (Łojasiewicz [20]) que funções analíticas reais satisfazem uma desigualdade do tipo acima com $\varphi(s) = s^{1-\theta}$ onde $\theta \in [\frac{1}{2}, 1)$. Uma boa prova deste resultado pode ser encontrada na monografia (Denkowska and Stasica [14]). Em um artigo recente, Kurdyka [17] estendeu este resultado a funções diferenciáveis definíveis em estruturas o-minimais. Mais recentemente Bolte et al. [7]-[9] estendeu a desigualdade de Kurdyka-Łojasiewicz a funções não-suaves no cenário subanalítico e o-minimal; veja também Bolte et al. [10].

Uma função própria semicontínua inferiormente $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ tem a propriedade Kurdyka Łojasiewicz em qualquer ponto não-crítico $\bar{x} \in \mathbb{R}^n$, veja Lema 1, seção 1.1.

Quando uma função tem a propriedade Kurdyka-Łojasiewicz, é importante ter estimativas de η, U e φ. Veremos, por exemplo, que muitas funções convexas satisfazem a propriedade acima com $U = \mathbb{R}^n$ e $\eta = +\infty$. A determinação de limites apertados para o caso não-convexo é muito mais envolvido.

Segundo, contamos com uma nova classe de algoritmos de minimização alternada com custos para mover que tem sido introduzidos recentemente em Attouch et al. [3] e que provou ser uma ferramenta flexível permitindo lidar com funções gerais de acoplamento $Q(x, y)$ (por exemplo, $Q(x, y) = \| Ax - By \|^2$, com A, B operadores lineares):

$(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$ dados, $(x_k, y_k) \to (x_{k+1}, y_k) \to (x_{k+1}, y_{k+1})$

$$\begin{cases} x_{k+1} \in \operatorname{argmin}\left\{L(u, y_k) + \frac{1}{2\lambda_k} \parallel u - x_k \parallel^2 : u \in \mathbb{R}^n\right\}, \\ y_{k+1} \in \operatorname{argmin}\left\{L(x_{k+1}, v) + \frac{1}{2\mu_k} \parallel v - y_k \parallel^2 : v \in \mathbb{R}^m\right\}. \end{cases} \tag{4}$$

O algoritmo acima pode ser visto como uma regularização proximal de um método bloco-duplo Gauss-Seidel para minimizar L:

$$\begin{cases} x_{k+1} \in \operatorname{argmin}\{L(u, y_k) : u \in \mathbb{R}^n\}, \\ y_{k+1} \in \operatorname{argmin}\{L(x_{k+1}, v) : v \in \mathbb{R}^m\}. \end{cases}$$

Alguns resultados gerais para o método de Gauss-Seidel, também conhecido como método de descida coordenada, pode ser encontrado por exemplo em Auslender [5], Bertsekas [6]; método bloco coordenado para funções não suaves e não convexas tem sido investigados por muitos autores (veja Tseng [22]). Entretanto, poucos resultados gerais asseguram que a sequência (x_k, y_k) converge a um mínimo global, mesmo para funções estritamente convexas. Um importante fato sobre nossa abordagem é que a convergência do algoritmo (4) funciona para quaisquer λ_k, μ_k maiores que um parâmetro positivo fixado, o qual pode ser escolhido arbitrariamente grande.

Capítulo 1

Noções Preliminares

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

1.1 Fatos elementares de análise não-suave

O produto escalar euclidiano do $\mathbb{R}^n$ e sua norma correspondente são, respectivamente, denotados por $\langle .,. \rangle$ e $\|.\|$. Uma boa referência sobre análise não-suave é Rockafellar e Wets [21].

Se $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ é uma aplicação ponto-conjunto seu gráfico é definido por

$$\mathrm{Graf}(F) := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m : y \in F(x)\}.$$

Analogamente, o gráfico de uma função $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ é definido por

$$\mathrm{Graf}(f) := \{(x,s) \in \mathbb{R}^n \times \mathbb{R} : s = f(x)\}.$$

Recordemos algumas definições sobre o cálculo subdiferencial.

Definição 1. *(Rockafellar e Wets [21]). Seja* $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *uma função própria semicontínua inferiormente (salvo menção contrária consideraremos* f *sempre nessa configuração).*

(i) O domínio de f *é definido e denotado por* $\mathrm{dom} f := \{x \in \mathbb{R}^n : f(x) < +\infty\}$.

(ii) Para cada $x \in \mathrm{dom} f$*, o subdiferencial de Fréchet de* f *em* x*, denotado por* $\widehat{\partial} f(x)$*, é o conjunto dos vetores* $x^* \in \mathbb{R}^n$ *tais que*

$$f(y) \geqslant f(x) + \langle x^*, y - x \rangle + o(\| y - x \|)$$

Se $x \notin \mathrm{dom}(f)$, *então* $\widehat{\partial} f(x) = \emptyset$.

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

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

Observação 1. *(a) A definição acima implica que* $\widehat{\partial} f(x) \subset \partial f(x)$ *para cada* $x \in \mathbb{R}^n$, *onde o primeiro conjunto é convexo e fechado enquanto o segundo é fechado. Com efeito, se* $x^* \in \widehat{\partial} f(x)$ *então tomando* $x_n = x$, $\forall n \in \mathbb{N}$, *obtemos:*

i) $x_n \to x$,

ii) $f(x_n) \to f(x)$,

iii) $x_n^* = x^* \in \widehat{\partial} f(x_n)$, $\forall n \in \mathbb{N}, \text{logo } x_n^* \to x^*$.

Portanto, $x^* \in \partial f(x) \Rightarrow \widehat{\partial} f(x) \subset \partial f(x)$.

Quanto a convexidade, dados ${x_1}^*, {x_2}^* \in \widehat{\partial} f(x)$ *definamos* ${x_t}^* := (1-t){x_1}^* + t{x_2}^*$, *onde* $t \in [0,1]$.

$f(y) - f(x) - \langle {x_t}^*, y - x\rangle - o(\| y - x \|) =$

$f(y) - f(x) - \langle (1-t){x_1}^* + t{x_2}^*, y - x\rangle - o(\| y - x \|) =$

$f(y) - f(x) - (1-t)\langle {x_1}^*, y - x\rangle - t\langle {x_2}^*, y - x\rangle - o(\| y - x \|) =$

$$(1-t)[\underbrace{(f(y) - f(x)) - \langle {x_1}^*, y - x\rangle - o(\| y - x \|)}_{\geqslant 0, \text{ pois } {x_1}^* \in \widehat{\partial} f(x)}] +$$

$$+t[\underbrace{(f(y) - f(x)) - \langle {x_2}^*, y - x\rangle - o(\| y - x \|)}_{\geqslant 0, \text{ pois } {x_2}^* \in \widehat{\partial} f(x)}] \geqslant 0 \Rightarrow {x_t}^* \in \widehat{\partial} f(x)$$

Portanto, $\widehat{\partial} f(x)$ *é convexo.*

Seja $\{x_n^*\}$ *uma sequência de pontos de* $\widehat{\partial} f(x)$ *tal que* $x_n^* \to x^*$. *Então:*

$f(y) \geqslant f(x) + \langle {x_n}^*, y - x\rangle + o(\| y - x \|)$, $\forall n \in \mathbb{N}$

Ao limite, com $n \to +\infty$ *obtemos*

$f(y) \geqslant f(x) + \langle x^*, y - x\rangle + o(\| y - x \|) \Rightarrow x^* \in \widehat{\partial} f(x)$

Portanto, $\widehat{\partial} f(x)$ *é fechado.*

Seja $x^{*j} \to x^*$ *uma sequência de pontos de* $\partial f(x)$. *Então,* $\forall j \in \mathbb{N}\ \exists\ {x^*}_k^j \to x^{*j}$ *tal que* $\exists\ x_k^j \to x, f(x_k^j) \to f(x)$ *e* ${x^*}_k^j \in \widehat{\partial} f(x_k^j)$. *Para concluir que* $x^* \in \partial f(x)$ *basta definir* $u_j := {x^*}_{k_j}^j$, *onde* k_j *é tal que*

(i) $\| {x^*}_{k_j}^j - x^{*j} \| < \varepsilon/2$, *o que é possível pois* ${x^*}_k^j \to x^{*j}$

(ii) $\| x_{k_j}^j - x \| < \varepsilon$, *o que é possível pois* $x_k^j \to x$

(iii) $\| f(x^j_{k_j}) - f(x) \| < \varepsilon$, *o que é possível pois* $f(x^j_k) \to f(x)$

(iv) $k_j < k_{j+1}, \forall j \in \mathbb{N}$

Com efeito, $u_j = x^j_{k_j} \in \widehat{\partial} f(x^j_k)$ $\forall j \in \mathbb{N}$ *e* $u_j \to x^*$, *pois dado* $\varepsilon > 0$ $\exists$ $j_0 \in \mathbb{N}$ *tal que* $j > j_0 \Rightarrow \| x^{*j} - x^* \| < \varepsilon/2$; *logo,*

$\| u_j - x^* \| = \| x^{*j}_{k_j} - x^* \| \leqslant \| x^{*j}_{k_j} - x^{*j} \| + \| x^{*j} - x^* \| < \varepsilon/2 + \varepsilon/2 = \varepsilon$

Portanto, $\partial f(x)$ *é fechado.*

(b) ($\mathrm{Graf}\partial f$ *é fechado) Seja* $(x_k, x^*_k)_{k\in\mathbb{N}} \in \mathrm{Graf}\partial f$ *uma sequência que converge a* (x, x^*). *Pela definição de* $\partial f(x)$, *se* $f(x_k)$ *converge a* $f(x)$ *então* $(x, x^*) \in \mathrm{Graf}\partial f$.

(c) No caso em que f *é diferenciável temos* $\{\nabla f(x)\} = \widehat{\partial} f(x) \subset \partial f(x)$

(d) Uma condição necessária (mas não suficiente) para $x \in \mathbb{R}^n$ *ser um mínimo de* f *é*

$$\partial f(x) \ni 0. \tag{1.1}$$

Com efeito, se x *é mínimo de* f *temos que*

$$f(y) \geqslant f(x),\ \forall y \in \mathbb{R}^n \Rightarrow f(y) \geqslant f(x) + \langle 0, y - x\rangle \Rightarrow 0 \in \widehat{\partial} f(x).$$

Não é suficiente, pois basta tomar $f : \mathbb{R} \to \mathbb{R}, f(x) = x^3$.

Um ponto que satisfaz (1.1) é chamado limite-crítico ou simplesmente crítico. O conjunto dos pontos críticos de f *é denotado por crit* f.

Se K é um subconjunto de $\mathbb{R}^n$ *e x é qualquer ponto em* $\mathbb{R}^n$, *definimos*

$$\mathrm{dist}(x, K) = \inf\{\| x - z \| : z \in K\}$$

Lembre-se que se K é vazio, temos $\mathrm{dist}(x, K) = +\infty$ *para todo* $x \in \mathbb{R}^n$. *Note também que para qualquer função real estendida* f *sobre* $\mathbb{R}^n$ *e qualquer* $x \in \mathbb{R}^n$,

$$\mathrm{dist}(0, \partial f(x)) = \inf\{\| x^* \| : x^* \in \partial f(x)\}.$$

Lema 1. *Seja* $\overline{x} \in \mathrm{dom} f$ *um ponto não-crítico de* f. *Então, existe* $c > 0$, *tal que*

$$\| x - \overline{x} \| + \| f(x) - f(\overline{x}) \| < c \Longrightarrow \mathrm{dist}(0, \partial f(x)) \geqslant c.$$

Demonstração. Caso contrário, existiria uma sequência (c_k) com $c_k > 0$, $c_k \to 0$ e uma sequência (x_k), com $\| x_k - \bar{x} \| + \| f(x_k) - f(\bar{x}) \| < c_k$ e $\mathrm{dist}(0, \partial f(x_k)) < c_k$. A última desigualdade implica a existência de algum ${x_k}^* \in \partial f(x_k)$ com $\| {x_k}^* \| < c_k$.

Como $c_k \to 0$ temos que $\| x_k - \bar{x} \| + \| f(x_k) - f(\bar{x}) \| \to 0$ e $\| {x_k}^* \| \to 0$, logo $x_k \to \bar{x}, f(x_k) \to f(\bar{x})$ e ${x_k}^* \to 0$. Portanto, como $\mathrm{Graf}\partial f$ é fechado temos que $(\bar{x}, 0) \in \mathrm{Graf}\,\partial f \Rightarrow 0 \in \partial f(\bar{x})$. Isto é uma contradição! $\square$

Subdiferenciação parcial. Seja L: $\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ uma função semicontínua inferiormente. Quando fixado $y \in \mathbb{R}^m$, o subdiferencial da função $L(., y)$ sobre u é denotado por $\partial_x L(u, y)$. Analogamente, quando fixado $x \in \mathbb{R}^n$, pode-se definir a subdiferenciação parcial em relação à variável y. O operador correspondente é denotado por $\partial_y L(x, .)$.

Lema 2. *Se* $f = g + f_0$ *com* g *finito sobre* $\bar{x}$ *e* $f_0 \in C^1$ *sobre uma vizinhança de* $\bar{x}$ *então* $\partial f(\bar{x}) = \partial g(\bar{x}) + \nabla f_0(\bar{x})$.

Demonstração. Veja Rockafellar e Wets [21] p.304. □

O seguinte resultado, embora elementar, é central para este trabalho.

Proposição 1. *Seja L satisfazendo* (H)*. Então para todo* $(x, y) \in \text{dom}(L) = \text{dom}(f) \times \text{dom}(g)$, *temos*

$$\partial L(x, y) = \{\partial f(x) + \nabla_x Q(x, y)\} \times \{\partial g(y) + \nabla_y Q(x, y)\} = \partial_x L(x, y) \times \partial_y L(x, y).$$

Demonstração. Observe primeiro que temos $\partial L(x, y) = \partial(f(x) + g(y)) + \nabla Q(x, y)$, pois Q é continuamente diferenciável (Lema 2). Além disso, o cálculo subdiferencial para funções separáveis nos dá $\partial(f(x) + g(y)) = \partial f(x) \times \partial g(y)$ (Rockafellar e Wets [21] Proposição 10.5, p.426]). Daí, segue a primeira igualdade.

Invocando mais uma vez o Lema 2 obtemos a segunda igualdade. □

Cones normais, funções indicadoras e projeções. Se C é um conjunto fechado de $\mathbb{R}^n$ denotamos por δ_C sua função indicadora, i.e., para todo $x \in \mathbb{R}^n$, temos

$$\delta_C(x) = \begin{cases} 0 & \text{se} \quad x \in C, \\ +\infty & \text{se} \quad x \notin C. \end{cases}$$

A projeção em C, denotada por P_C, é a seguinte aplicação ponto-conjunto

$$\begin{cases} P_C : \mathbb{R}^n \rightrightarrows \mathbb{R}^m, \\ P_C(x) := \text{argmin}\{\|x - z\| : z \in C.\} \end{cases}$$

Quando C é não-vazio, o fechamento de C implica que $P_C(x)$ é não vazio para todo x em $\mathbb{R}^n$.

Definição 2. *(Cone normal). Seja* C *um subconjunto fechado e não-vazio de* $\mathbb{R}^n$.

(i) Para qualquer $x \in C$ *o Fréchet cone normal de* C *em* x *é definido por*

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

Quando $x \notin C$, *temos* $\widehat{N}_C(x) = \emptyset$.

(ii) O cone normal de C *em* $x \in C$ *é denotada por* $N_C(x)$ *e é definido por*

$$v \in N_C(x) \Leftrightarrow \exists\ x_k \in C,\ x_k \to x,\ \exists\ v_k \in \widehat{N}_C(x_k),\ v_k \to v.$$

1.2 Conjuntos algébricos e semi-algébricos

Definição 3. *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 1. *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 2. *Se* A, B *são 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. □

Proposição 3. *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 $\mathbb{R}^m$. □

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

Exemplo 2. *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 1 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 é 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.*

Este exemplo motiva a definição de uma classe mais geral que a dos conjuntos algébricos, a saber, a classe dos conjuntos semialgébricos.

Definição 4. *Um subconjunto* A *de* $\mathbb{R}^n$ *é dito semi-algébrico se existe um número finito de funções polinômiais reais* $P_j, Q_{ij} : \mathbb{R}^n \to \mathbb{R}$ *tais que*

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

Definição 5. *Uma função* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *(resp. uma aplicação ponto-conjunto* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$*) é chamada semi-algébrica se o gráfico* $\{(x, \lambda) \in \mathbb{R}^{n+1} : f(x) = \lambda\}$ *(resp.* $\{(x, y) \in \mathbb{R}^{n+m} : y \in F(x)\}$*) é um subconjunto semi-algébrico de* $\mathbb{R}^{n+1}$ *(resp.* $\mathbb{R}^{n+m}$*).*

Definição 6. *Seja* $X \subset \mathbb{R}^m$ *e* $Y \subset \mathbb{R}^n$ *conjuntos semi-algébricos. Uma aplicação* $f : X \to Y$ *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

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

Proposição 4. *Na reta, um conjunto semi-algébrico é a reunião finita de intervalos abertos, fechados, semi-abertos e pontos.*

Demonstração. Veja página 25 de [11]. □

Teorema 1. (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 qualquer 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 [11]. □

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

Demonstração. Sejam $F : \mathbb{R}^n \to \mathbb{R}^m$ uma aplicação polinomial, e A um subconjunto semi-algébrico de $\mathbb{R}^n$. O gráfico de $F|_A : A \to \mathbb{R}^m$ é dado por Graf $F|_A = \{(x, F(x)) : x \in A\}$. Mostraremos inicialmente que Graf $F|_A$ é um conjunto semi-algébrico.

Sejam $x = (x_1, ..., x_n) \in \mathbb{R}^n$ e $y = (y_1, ..., y_m) \in \mathbb{R}^m$. Daí,

$$\begin{aligned}
\text{Graf } F &= \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : y_1 = F_1(x), ..., y_m = F_m(x)\} \\
&= \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : (F_1(x) - y_1)^2 + ... + (F_m(x) - y_m)^2 = 0\} \\
&= \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : \varphi(x, y) = 0\},
\end{aligned}$$

onde $\varphi(x, y) = (F_1(x) - y_1)^2 + ... + (F_m(x) - y_m)^2$. Logo

$$\text{Graf } F|_A = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : x \in A, \varphi(x, y) = 0\}.$$

Como A é um conjunto semi-algébrico, existe um número finito de funções polinômiais reais $P_j, Q_{ij} : \mathbb{R}^n \to \mathbb{R}$ tais que

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

Defina as funções polinomiais, $\widetilde{f}_j(x, y) = P_j^2(x) + \varphi^2(x, y)$ e $\widetilde{g}_{ij}(x, y) = Q_{ij}(x)$. Daí

$$\begin{aligned}
\text{Graf } F|_A &= \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : x \in A, \varphi(x, y) = 0\} \\
&= \bigcup_{j=1}^{p} \bigcap_{i=1}^{q_j} \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : \widetilde{f}_j(x, y) = 0, \widetilde{g}_{ij}(x, y) < 0\}.
\end{aligned}$$

Isso mostra que Graf $F|_A$ é um subconjunto semi-algébrico de $\mathbb{R}^n \times \mathbb{R}^m$. Considere a projeção $\pi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ dada por $\pi(x, y) = y$. Assim, $F(A) = \pi(\text{Graf } F|_A)$, e pelo Teorema de Tarski-Seidenberg, o conjunto $F(A)$ é semi-algébrico. □

Corolário 2. *A imagem de um conjunto semi-algébrico por uma aplicação semi-algébrica é um conjunto semi-algébrico.*

Corolário 3. *Sejam* $f : A \subset \mathbb{R}^p \to \mathbb{R}^n$ *e* $g : B \subset \mathbb{R}^n \to \mathbb{R}^m$ *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, $\{(x, f(x)) \in \mathbb{R}^p \times \mathbb{R}^n : x \in A\}$ e $\{(y, g(y)) \in \mathbb{R}^n \times \mathbb{R}^m : y \in B\}$ são conjuntos semi-algébricos, e como $f(A) \subset B$ segue que o conjunto $\{(f(x), g(f(x))) \in \mathbb{R}^n \times \mathbb{R}^m : x \in A\}$ também é semi-algébrico. Daí o conjunto

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

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 5. *Sejam* $S \subset \mathbb{R}^m$ *semi-algébrico e não-vazio,* $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ *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 1, 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 4. *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 3. *Se* S *é um subconjunto semi-algébrico não-vazio de* $\mathbb{R}^m$*, então a função*

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

é semi-algébrica (basta aplicarmos a Corolário 4 à função polinomial $g(x, y) = \|x - y\|^2$*). Daí,* $\psi(x) = \inf\{\|x - y\|^2 : y \in S\}$ *é uma função semi-algébrica.*

Exemplo 4. *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*

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

é semi-algébrica. Com efeito, pelo Exemplo 3, a aplicação $\xi(x) = (\mathrm{dist}^2(x, F_1), ..., \mathrm{dist}^2(x, F_i))$ *é semi-algébrica. Logo* $f(x) = (h \circ \xi)(x)$*, onde* $h(x) = \dfrac{x_1 + \cdots + x_n}{2}$*. Aplicando o Corolário 3, obtemos o resultado.*

1.3 Conjuntos analíticos, semi-analíticos e subanalíticos

Definição 7. *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 5. *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 8. *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 9. *Um conjunto semi-analítico é a reunião finita de conjuntos semi-analíticos básicos.*

Observação 2. *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 10. *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 uma conjunto semi-analítico.*

Definição 11. *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 uma conjunto semi-analítico.*

Definição 12. *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.

Como antes temos as seguintes definições:

Definição 13. *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.*

Exemplo 6. *Seja* $f : U \to \mathbb{R}^k$ *uma aplicação subanalítica diferenciável em um aberto* $U \subset \mathbb{R}^n$. *Então* $\partial f / \partial x_j$, $j = 1, ..., n$ *são funções subanalíticas, e daí* **grad** f *é uma aplicação subanalítica. Veja [17].*

Definição 14. *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;

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

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.

1.4 Funções KL

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 15. (Propriedade Kurdyka-Łojasiewicz)

(a) Diz-se que uma função $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *tem a propriedade de 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 $s \in (0, \eta)$*,* $\varphi'(s) > 0$*,*

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

$$\varphi'(f(x) - f(x^*))\operatorname{dist}(0, \partial f(x)) \geqslant 1. \tag{1.2}$$

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

Observação 3. *(a) S. Łojasiewicz provou em 1963 (Łojasiewicz [20]) que funções analíticas reais satisfazem uma desigualdade do tipo acima com* $\varphi(s) = s^{1-\theta}$ *onde* $\theta \in [\frac{1}{2}, 1)$. *Uma boa prova deste resultado pode ser encontrada na monografia (Denkowska e Stasica [14]). Em um artigo recente, Kurdyka [17] estendeu este resultado a funções diferenciáveis definíveis em estruturas o-minimais (veja §3.3). Mais recentemente Bolte et al. [7]-[9] estendeu a desigualdade Kurdyka Łojasiewicz a funções não-suaves no cenário subanalítico e o-minimal; veja também Bolte et al. [10].*

(b) Uma função própria semicontínua inferiormente $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *tem a propriedade de Kurdyka Łojasiewicz em qualquer ponto não-crítico* $\bar{x} \in \mathbb{R}^n$. *Com efeito, o lema 1 garante a existência de* $c > 0$, *tal que*

$$\operatorname{dist}(0, \partial f(x)) \geqslant c > 0$$

sempre que $x \in B(\bar{x}, c/2) \cap [f(\bar{x}) < f < f(\bar{x}) + c/2]$; $\varphi(s) = c^{-1}s$ *é então uma função côncava aceitável.*

(c) Quando uma função tem a propriedade Kurdyka Łojasiewicz, é importante ter estimativas de η, U *e* φ. *Veremos, por exemplo, que muitas funções convexas satisfazem a propriedade acima com* $U = \mathbb{R}^n$ *e* $\eta = +\infty$.

Vejamos então alguns exemplos de funções KL.

O próximo teorema nos diz que funções convexas não verificam necessariamente a propriedade de Kurdyka-Łojasiewicz.

Teorema 2. *Existe uma função 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 [10]. $\square$

Definição 16. *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 $\overline{x}$. Para isso utilizaremos o

Lema 3. *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 $\|Hx\| \geqslant c\|x\|$. □

Vamos então à prova da afirmação.

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 grad 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 3 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 $\|x - \overline{x}\| < \delta_2$, $\|\widetilde{\rho}(x)\| < \dfrac{c}{2}$. 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 15, tome $U = B(\overline{x}, \widetilde{\delta})$, $\eta = \widetilde{\delta}$ e $\varphi(s) = 2\dfrac{\sqrt{c_1 s}}{c_2}$. Logo, para todo $x \in U \cap [0 < f(x) - f(\overline{x}) < \eta]$,

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

1.4.1 Estrutura o-minimal

Definição 17. *Seja* $\mathcal{M} = \bigcup_{n \in \mathbb{N}} \mathcal{M}_n$*, 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 18. *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 19. *Dizemos que* $f : A \to \mathbb{R}^m$*, onde* $A \subset \mathbb{R}^n$*, é uma* $\mathcal{M}$*-função se o gráfico de* f *é um* $\mathcal{M}$*-conjunto.*

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

A seguir enunciaremos alguns resultados básicos da estrutura o-minimal, cujas respectivas demonstrações podem ser encontradas em [17].

Lema 4. *Seja* $f : A \to \mathbb{R}$ *uma* $\mathcal{M}$*-função tal que* $f(x) \geqslant 0$ *para todo* $x \in A$*. Seja* $G : A \to \mathbb{R}^m$ *uma* $\mathcal{M}$*-aplicação e defina a função* $\varphi : G(A) \to \mathbb{R}$ *por*

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

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

Corolário 5. *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 6. *Seja* A *uma* $\mathcal{M}$*-conjunto em* $\mathbb{R}^n$*. Então* $\overline{A}$ *e* $\operatorname{int} A$ *são* $\mathcal{M}$*-conjuntos.*

Lema 5. (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 6. *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 7. (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* $\gamma((0, \varepsilon)) \subset A\backslash\{a\}$*.*

Apresentaremos agora o Teorema de Pouiseux, que será muito útil nesta seção.

Teorema 3. *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 [18]. □

O principal resultado desta seção é o seguinte teorema.

Teorema 4. *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*

$$\|\nabla(\Psi \circ f)(x)\| \geqslant c, \tag{1.3}$$

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

Demonstração. Segue do Lema 6 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 4, 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. 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 5, 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,

$$\xi(1) - \xi(0) = \int_0^1 \xi'(t)dt \Rightarrow h(s) = \int_0^1 h'(ts)s dt.$$

Daí, para $0 < t < 1$ e $s \in (0, \delta)$, temos $ts < s$, e consequentemente, $h'(ts) < h'(s)$. Logo

$$h(s) = \int_0^1 h'(ts)s dt \leqslant \int_0^1 sh'(s)dt = sh'(s).$$

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 $\widetilde{\eta} : (-\delta, \delta) \to \mathbb{R}^{n+1}$ 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 5 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, $\|\nabla f(\eta(s))\| \geqslant \frac{1}{B}(f \circ \eta)'(s)$. Logo,

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

□

Teorema 5. *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 4 para o caso subanalítico, podemos aplicar o Teorema 3 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,

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

Derivando Ψ, obtemos

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

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 $|\Psi'(t)| \leqslant Mt^{\frac{1}{k}-1}$. 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. $\square$

A seguir, mostraremos que se f é uma função analítica real em uma vizinhança da origem, então f goza da Propriedade Kurdyca-Łojasiewicz em $\overline{x} = 0$.

Teorema 6. (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 $c > 0, \theta$ tais que $0 < \theta < 1$ e

$$\|\ \text{grad } f(x)\ \| \geqslant c|f(x)|^{\theta} \tag{1.4}$$

em alguma vizinhança de 0.

Demonstração. Como f analítica, em particular, pelo Teorema 3 é subanalítica. Pelo Teorema 5 segue o resultado. $\square$

Exemplo 7. *Seja $f : \mathbb{R} \to \mathbb{R}$ definida por*

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

Afirmamos que f é uma função C^∞ que não satisfaz (1.4). 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 $x \neq 0$,

$f^{(n)}(x) = p\left(\frac{1}{x}\right)e^{-1/x^2}$, *onde* p *é um polinômio. Ponha* $y = 1/x$, *e suponha por indução que para todo* $f^{(n-1)}(0) = 0$. *Segue que,*

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

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

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

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

Assim,

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

o que é uma absurdo.

Capítulo 2

Algoritmos de minimização proximal alternada

2.1 Convergência a um valor crítico

Seja L satisfazendo (H). Seja dado $(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$, lembre-se que o sistema dinâmico discreto alternado que vamos estudar é da forma $(x_k, y_k) \to (x_{k+1}, y_k) \to (x_{k+1}, y_{k+1})$:

$$x_{k+1} \in \operatorname{argmin} \left\{ \mathsf{L}(u, y_k) + \frac{1}{2\lambda_k} \parallel u - x_k \parallel^2 : u \in \mathbb{R}^n \right\}, \tag{2.1}$$

$$y_{k+1} \in \operatorname{argmin} \left\{ \mathsf{L}(x_{k+1}, v) + \frac{1}{2\mu_k} \parallel v - y_k \parallel^2 : v \in \mathbb{R}^m \right\}, \tag{2.2}$$

onde $(\lambda_k)_{k\in\mathbb{N}}, (\mu_k)_{k\in\mathbb{N}}$ são sequências positivas.

Fazemos as hipóteses seguintes, relativas a (2.1), (2.2):

$$(\mathsf{H}_1) \begin{cases} \mathrm{L} \quad \text{é} \quad \text{limitada} \quad \text{inferiormente,} \\ \mathrm{L}(., \mathrm{y}_0) \quad \text{é} \quad \text{própria} \\ \text{para} \quad 0 < \mathrm{r}_- < \mathrm{r}_+ \quad \lambda_k, \mu_k \quad \text{pertencem} \quad \text{a} \quad (\mathrm{r}_-, \mathrm{r}_+) \quad \text{para} \quad \text{todo} \quad \mathrm{k} \geqslant 0. \end{cases}$$

O lema a seguir, especialmente o ponto (iii), é de uso constante no que se segue.

Lema 8. *Sob as hipóteses* (H), (H_1), *as sequências* $(x_k), (y_k)$ *estão bem definidas. Mais ainda, o seguinte é válido:*

(i) Para todo $k \geqslant 0$

$$\mathsf{L}(x_{k+1}, y_{k+1}) + \frac{1}{2\lambda_k} \parallel x_{k+1} - x_k \parallel^2 + \frac{1}{2\mu_k} \parallel y_{k+1} - y_k \parallel^2 \leqslant \mathsf{L}(x_k, y_k); \tag{2.3}$$

segue que $\{L(x_k, y_k)\}$ *é não-crescente.*

(ii)

$$\sum_{k=0}^{\infty}(\| x_{k+1} - x_k \|^2 + \| y_{k+1} - y_k \|^2) < +\infty;$$

logo, $\lim(\| x_{k+1} - x_k \| + \| y_{k+1} - y_k \|) = 0$.

(iii) Defina $(x^*_{k+1}, y^*_{k+1}) =$

$$(\nabla_x Q(x_{k+1}, y_{k+1}) - \nabla_x Q(x_{k+1}, y_k), 0) - ((1/\lambda_k)(x_{k+1} - x_k), (1/\mu_k)(y_{k+1} - y_k)),$$

para $k \geqslant 0$; *temos*

$$(x^*_{k+1}, y^*_{k+1}) \in \partial L(x_{k+1}, y_{k+1}). \tag{2.4}$$

Para toda subsequência limitada (x_{k_j}, y_{k_j}) de (x_k, y_k), *temos* $(x^*_{k_j}, y^*_{k_j}) \to 0$; *portanto* $\text{dist}(0, \partial L(x_{k_j}, y_{k_j})) \to 0$.

Demonstração. Por hipótese temos que L é limitada inferiormente e que (λ_k) é limitada. Além disso, $\xi : \mathbb{R}^n \to \mathbb{R}, u \to \xi(u) = \| u - x_k \|^2$ é coerciva, logo $= L(., y_k) + \frac{1}{2\lambda_k}\xi(.)$ é coerciva e como $L(., y_0)$ é propria existe $x_1 \in \text{argmin}\left\{L(u, y_k) + \frac{1}{2\lambda_k} \| u - x_k \|^2 : u \in \mathbb{R}^n\right\}$

Segue-se que $L(x_1 .,)$ é própria, pois $L(x_1, y_0) < +\infty$. Analogamente para y_1 e por indução obtemos que $(x_k), (y_k)$ estão bem definidas.

Como $x_{k+1} \in \text{argmin}\left\{L(u, y_k) + \frac{1}{2\lambda_k} \| u - x_k \|^2 : u \in \mathbb{R}^n\right\}$ temos que $L(x_{k+1}, y_k) + \frac{1}{2\lambda_k} \| x_{k+1} - x_k \|^2 \leqslant L(x_k, y_k) + \frac{1}{2\lambda_k} \| x_k - x_k \|^2 = L(x_k, y_k)$

Por outro lado, $y_{k+1} \in \text{argmin}\left\{L(x_{k+1}, v) + \frac{1}{2\mu_k} \| v - y_k \|^2 : v \in \mathbb{R}^m\right\} \Rightarrow$

$L(x_{k+1}, y_{k+1}) + \frac{1}{2\mu_k} \| y_{k+1} - y_k \|^2 \leqslant L(x_{k+1}, y_k) + \frac{1}{2\mu_k} \| y_k - y_k \|^2 = L(x_{k+1}, y_k)$

$\therefore L(x_{k+1}, y_{k+1}) + \frac{1}{2\mu_k} \| y_{k+1} - y_k \|^2 + \frac{1}{2\lambda_k} \| x_{k+1} - x_k \|^2 \leqslant$

$L(x_{k+1}, y_k) + \frac{1}{2\lambda_k} \| x_{k+1} - x_k \|^2 \leqslant L(x_k, y_k)$, isto demonstra (i).

Além disso, como $0 < \lambda_k, \mu_k < r_+$ temos

$\frac{1}{2r_+} \| y_{k+1} - y_k \|^2 + \frac{1}{2r_+} \| x_{k+1} - x_k \|^2 \leqslant L(x_k, y_k) - L(x_{k+1}, y_{k+1})$

$\frac{1}{2r_+}(\| y_{k+1} - y_k \|^2 + \| x_{k+1} - x_k \|^2) \leqslant L(x_k, y_k) - L(x_{k+1}, y_{k+1})$

$\frac{1}{2r_+}\sum_{k=0}^{n}(\| x_{k+1} - x_k \|^2 + \| y_{k+1} - y_k \|^2) \leqslant L(x_0, y_0) - L(x_n, y_n) \leqslant L(x_0, y_0) - \inf L$

$\Rightarrow \sum_{k=0}^{n}(\| x_{k+1} - x_k \|^2 + \| y_{k+1} - y_k \|^2) \leqslant 2r_+(L(x_0, y_0) - \inf L), \forall n \in \mathbb{N}$. Portanto,

$\sum_{k=0}^{\infty}(\| x_{k+1} - x_k \|^2 + \| y_{k+1} - y_k \|^2) < +\infty \Rightarrow \lim_{k\to\infty}(\| x_{k+1} - x_k \|^2 + \| y_{k+1} - y_k \|^2) = 0$

$\Rightarrow \lim \| x_{k+1} - x_k \|^2 = 0.$

Segue-se que $\lim \| x_{k+1} - x_k \| = \lim(\| x_{k+1} - x_k \|^2)^{1/2} = (\lim \| x_{k+1} - x_k \|^2)^{1/2} = 0$. Analogamente, $\lim \| y_{k+1} - y_k \| = 0$, portanto (ii) está provado.

Como $\inf L > -\infty$, (H) implica que para todo $r > 0$, $(\bar{u}, \bar{v}) \in \mathbb{R}^n \times \mathbb{R}^m$ as funções $u \to L(u, \bar{v}) + (1/(2r)) \| u - \bar{u} \|^2$ e $v \to L(\bar{u}, v) + (1/(2r)) \| v - \bar{v} \|^2$ são coercivas.

Pela definição de x_{k+1}, e pela obsevação 1(c), 0 deve pertencer ao subdiferencial, no ponto x_{k+1}, da função $u \mapsto L(u, y_k) + (1/2\lambda_k) \| u - x_k \|^2$, o qual é igual a $\partial_x L(x_{k+1}, y_k) + (1/\lambda_k)(x_{k+1} - x_k)$ pois a função $u \mapsto (1/2\lambda_k) \| u - x_k \|^2$, é diferenciável. Daí,

$$0 \in \partial_x L(x_{k+1}, y_k) + \frac{1}{\lambda_k}(x_{k+1} - x_k), \forall k \geqslant 0; \tag{2.5}$$

analogamente,

$$0 \in \partial_y L(x_{k+1}, y_{k+1}) + \frac{1}{\mu_k}(y_{k+1} - y_k), \forall k \geqslant 0. \tag{2.6}$$

Pela estrutura de L, temos $\partial_x L(x_{k+1}, y_k) = \partial f(x_{k+1}) + \nabla_x Q(x_{k+1}, y_k)$ e

$\partial_y L(x_{k+1}, y_{k+1}) = \partial g(y_{k+1}) + \nabla_y Q(x_{k+1}, y_{k+1})$. Logo, por (2.5) e (2.6) temos

$$-\frac{1}{\lambda_k}(x_{k+1} - x_k) - (\nabla_x Q(x_{k+1}, y_k) - \nabla_x Q(x_{k+1}, y_{k+1}), 0) \in \partial f(x_{k+1}) + \nabla_x Q(x_{k+1}, y_{k+1});$$

$$-\frac{1}{\mu_k}(y_{k+1} - y_k) \in \partial g(y_{k+1}) + \nabla_y Q(x_{k+1}, y_{k+1}).$$

Basta então usar a Proposição 1 para concluir (2.4).

Se $\{(x_{k_j}, y_{k_j})\}$ é uma sequência limitada, então $\{(x_{k_j}, y_{k_j - 1})\}$ também o é, e por (ii), $(x_{k_j}, y_{k_j}) - (x_{k_j}, y_{k_j - 1}) \to 0, \quad j \to +\infty$. O fato de $\nabla_x Q$ ser Lipschitz sobre conjuntos limitados assegura a última afirmação de (iii).

□

Observação 4. *Sem hipóteses adicionais (por exemplo, a convexidade de L) a sequência* (x_k, y_k), *a priori, não é única, pois cada subproblema pode ter várias soluções.*

A próxima proposição, especialmente os itens (i),(ii), fornece o primeiro resultado de convergência sobre sequências geradas por (2.1),(2.2). Os teoremas 1 e 2 tornarão as propriedades de convergência muito mais precisas.

Proposição 6. *Assuma que* $(H), (H_1)$ *são válidas. Seja* $\{(x_k, y_k)\}$ *uma sequência satisfazendo (2.1) e (2.2). Denote por* $\omega(x_0, y_0)$ *o conjunto (possivelmente vazio) de seus valores de aderência. Então:*

(i) $\omega(x_0, y_0) \subset \mathrm{crit} L$;

(ii) L é finito e constante sobre $\omega(x_0, y_0)$, *igual a* $\inf_{k \in \mathbb{N}} L(x_k, y_k) = \lim_{k \to \infty} L(x_k, y_k)$.

Demonstração. (i) Pela definição de $(x_{k+1}, y_{k+1})(k \geqslant 0)$, nós temos

$$L(x_{k+1}, y_k) + \frac{1}{2\lambda_k} \| x_{k+1} - x_k \|^2 \leqslant L(u, y_k) + \frac{1}{2\lambda_k} \| u - x_k \|^2, \quad \forall u \in \mathbb{R}^n,$$

$$L(x_{k+1}, y_{k+1}) + \frac{1}{2\mu_k} \| y_{k+1} - y_k \|^2 \leqslant L(x_{k+1}, v) + \frac{1}{2\mu_k} \| y_k - v \|^2, \quad \forall v \in \mathbb{R}^m.$$

Devido à forma especial de L, a $0 < r_- \leqslant \lambda_k \leqslant r_+$ e a $0 < r_- \leqslant \mu_k \leqslant r_+$, segue que, $\forall u \in \mathbb{R}^n, \quad \forall v \in \mathbb{R}^m$,

$$f(x_{k+1}) + Q(x_{k+1}, y_k) + \frac{1}{2r_+} \| x_{k+1} - x_k \|^2 \leqslant f(u) + Q(u, y_k) + \frac{1}{2r_-} \| u - x_k \|^2, \tag{2.7}$$

$$g(y_{k+1}) + Q(x_{k+1}, y_{k+1}) + \frac{1}{2r_+} \| y_{k+1} - y_k \|^2 \leqslant g(v) + Q(x_{k+1}, v) + \frac{1}{2r_-} \| y_k - v \|^2 . \tag{2.8}$$

Seja $(\bar{x}, \bar{y})$ um ponto em $\omega(x_0, y_0)$, então existe uma subsequência $\{(x_{k_j}, y_{k_j})\}$ de $\{(x_k, y_k)\}$ convergindo a $(\bar{x}, \bar{y})$. Como $\| x_{k+1} - x_k \| + \| y_{k+1} - y_k \| \to 0$, deduzimos de (2.7)

$$\text{lim inf } f(x_{k_j}) + Q(\bar{x}, \bar{y}) \leqslant f(u) + Q(u, \bar{y}) + \frac{1}{2r_-} \| u - \bar{x} \|^2 \quad \forall u \in \mathbb{R}^n.$$

Em particular, para $u = \bar{x}$, obtemos

$$\text{lim inf } f(x_{k_j}) \leqslant f(\bar{x}).$$

Como f é semicontínua inferiormente, conclui-se

$$\text{lim inf } f(x_{k_j}) = f(\bar{x}).$$

Sem perda de generalidade podemos assumir que $\{f(x_{k_j})\}$ converge a $f(\bar{x})$:

$$\lim f(x_{k_j}) = f(\bar{x}). \tag{2.9}$$

Analogamente, usando (2.8), podemos assumir que lim $g(y_{k_j}) = g(\bar{y})$. Agora, como Q é contínua, temos $Q(x_{k_j}, y_{k_j}) \to Q(\bar{x}, \bar{y})$, e daí $L(x_{k_j}, y_{k_j}) \to L(\bar{x}, \bar{y})$. Mas pelo Lema 9(iii), usando a mesma notação, temos $(x^*_{k_j}, y^*_{k_j}) \in \partial L x_{k_j}, y_{k_j})$ e $(x^*_{k_j}, y^*_{k_j}) \to 0$. Devido ∂L ser fechado, finalmente obtemos $0 \in \partial L(\bar{x}, \bar{y})$.

(ii) Para qualquer ponto $(\bar{x}, \bar{y}) \in \omega(x_0, y_0)$, acabamos de ver que existe uma subsequência $\{(x_{k_j}, y_{k_j})\}$ com $L(x_{k_j}, y_{k_j}) \to \{L(\bar{x}, \bar{y})\}$. Como a sequência $L(x_k, y_k)$ é não-crescente, obtemos $L(\bar{x}, \bar{y}) = \inf L(x_k, y_k)$ independente de $(\bar{x}, \bar{y})$. □

2.2 Convergência a um ponto crítico e outros resultados de convergência

Esta seção é devotada à análise de convergência do algoritmo de minimização proximal alternada introduzido em § 2.1. Ele fornece os principais resultados matemáticos deste trabalho.

Em Attouch e Bolte [2] a desigualdade Łojasiewicz é usada para derivar a convergência do usual método proximal, mas algoritmos alternados não são considerados.

Introduzamos as notações: $z_k = (x_k, y_k), l_k = L(z_k), \bar{z} = (\bar{x}, \bar{y}), \bar{l} = L(\bar{z})$.

Teorema 7. (Resultado de pré-convergência.)**.** *Assuma que* L *satisfaz* (H), (H_1) *e tem a propriedade Kurdyka-Łojasiewicz sobre* $\bar{z} = (\bar{x}, \bar{y})$. *Denote por* U, η e $\varphi : [0, \eta) \to \mathbb{R}$ *os objetos que aparecem em (1.2) relativos a* L *e* $\bar{z}$. *Seja* $\rho > 0$ *tal que* $B(\bar{z}, \rho) \subset U$.

Seja (z_k) *uma sequência gerada pelo algoritmo alternado (2.1), (2.2), com* z_0 *como ponto inicial. Assuma que*

$$\bar{l} < l_k < \bar{l} + \eta, \; \forall k \geqslant 0. \tag{2.10}$$

e

$$M\varphi(l_0 - \bar{l}) + 2\sqrt{2r_+}\sqrt{l_0 - \bar{l}} + \| z_0 - \bar{z} \| < \rho. \tag{2.11}$$

com $M = 2r_+(C + 1/r_-)$ *onde C é uma constante de Lipschitz para* ∇Q *sobre* $B(\bar{z}, \sqrt{2}\rho)$.

Então, a sequência (z_k) *converge a um ponto crítico de L e vale a seguinte estimativa:* $\forall k \geqslant 0$

$$(i) \qquad z_k \in B(\bar{z}, \rho), \tag{2.12}$$

$$(ii) \qquad \sum_{i=k+1}^{\infty} \| z_{i+1} - z_i \| \leqslant M\varphi(l_k - \bar{l}) + \sqrt{2r_+}\sqrt{l_k - \bar{l}}. \tag{2.13}$$

O significado do teorema é mais ou menos o seguinte: Uma seqüência (z_k) *que começa numa vizinhança de um ponto* $\bar{z}$ *(como dado em (2.11)) e que não melhora* $L(\bar{z})$ *(como dado em (2.10)) converge a um ponto crítico perto de* $\bar{z}$.

Demonstração. Sem perda de generalidade podemos assumir que $L(\bar{z}) = 0$ (substitua L por $L - L(\bar{z})$ se for necessário),

Com (2.3) temos para $i \geqslant 0$:

$$l_i - l_{i+1} \geqslant \frac{1}{2r_+} \| z_{i+1} - z_i \|^2 . \tag{2.14}$$

Entretanto, $\varphi'(l_i)$ faz sentido em vista de (2.10)

$$\varphi'(l_i)(l_i - l_{i+1}) \geqslant \frac{\varphi'(l_i)}{2r_+} \parallel z_{i+1} - z_i \parallel^2 .$$

Devido a φ ser côncava, temos

$$\varphi(l_i) - \varphi(l_{i+1}) \geqslant \frac{\varphi'(l_i)}{2r_+} \parallel z_{i+1} - z_i \parallel^2, \ \forall i \geqslant 0. \tag{2.15}$$

Vamos primeiro checar (i) para k=0 e k=1. Em vista de (2.11), $z_0 \in B(\bar{z}, \rho)$.

Quanto a z_1, (2.14) implica, para i=0,

$$\frac{1}{2r_+} \parallel z_1 - z_0 \parallel^2 \leqslant l_0 - l_1 \leqslant l_0. \tag{2.16}$$

logo,

$$\parallel z_1 - \bar{z} \parallel \leqslant \parallel z_1 - z_0 \parallel + \parallel z_0 - \bar{z} \parallel \leqslant \sqrt{2r_+}\sqrt{l_0} + \parallel z_0 - \bar{z} \parallel . \tag{2.17}$$

Daí $z_1 \in B(\bar{z}, \rho)$ em vista de (2.11).

Vamos agora provar por indução que $(x_k, y_k) \in B(\bar{z}, \rho)$, $\forall \ k \geqslant 0$.

Sendo isto verdade para $k \in \{0, 1\}$, suponhamos que seja válido até algum $k \geqslant 1$.

Para $0 \leqslant i \leqslant k$, como $z_i \in B(\bar{z}, \rho)$ e $0 < l_i < \eta$, podemos escrever a desigualdade Kurdica- Łojasiewicz sobre z_i;

$$\varphi'(l_i)\text{dist}(0, \partial L(z_i)) \geqslant 1.$$

Lembre-se de que, pelo lema 2 (iii), temos

$$(x_i^*, y_i^*) = -(\nabla_x Q(x_i, y_{i-1}) - \nabla_x Q(x_i, y_i), 0) - \left(\frac{1}{\lambda_{i-1}}(x_i - x_{i-1}), \frac{1}{\mu_{i-1}}(y_i - y_{i-1})\right)$$

é um elemento de $\partial L(x_i, y_i)$. Daí temos para $1 \leqslant i \leqslant k$.

$$\varphi'(l_i) \parallel (x_i^*, y_i^*) \parallel \geqslant 1. \tag{2.18}$$

Vamos examinar $\parallel (x_i^*, y_i^*) \parallel$, para $1 \leqslant i \leqslant k$. Por um lado,

$$\parallel ((1/\lambda_{i-1})(x_i - x_{i-1}), (1/\mu_{i-1})(y_i - y_{i-1})) \parallel \leqslant (1/r_-) \parallel z_i - z_{i-1} \parallel$$

Por outro lado, vamos observar

$$\parallel (x_i, y_{i-1}) - (\bar{x}, \bar{y}) \parallel^2 = \parallel x_i - \bar{x} \parallel^2 + \parallel y_{i-1} - \bar{y} \parallel^2 \leqslant \parallel z_i - \bar{z} \parallel^2 + \parallel z_{i-1} - \bar{z} \parallel^2 \leqslant 2\rho^2.$$

Daí $(x_i, y_{i-1}), z_i = (x_i, y_i) \in B(\bar{z}, \sqrt{2}\rho)$, que nos permite aplicar a desigualdade de Lipschitz entre estes pontos:

$$\| \nabla_x Q(x_i, y_i) - \nabla_x Q(x_i, y_{i-1}) \| \leqslant C \| y_i - y_{i-1} \| \leqslant C \| z_i - z_{i-1} \| .$$

Para $1 \leqslant i \leqslant k$,

$$\| (x_i^*, y_i^*) \| \leqslant (C + 1/r_-) \| z_i - z_{i-1} \| . \tag{2.19}$$

Agora, por (2.18), obtemos

$$\varphi'(l_i) \geqslant \frac{1}{(C + 1/r_-)} \| z_i - z_{i-1} \|^{-1}, \ 1 \leqslant i \leqslant k,$$

e por (2.15)

$$\varphi(l_i) - \varphi(l_{i+1}) \geqslant \frac{1}{M} \frac{\| z_{i+1} - z_i \|^2}{\| z_i - z_{i-1} \|}, \ 1 \leqslant i \leqslant k.$$

Reescrevemos a desigualdade anterior da seguinte forma:

$$\| z_i - z_{i-1} \|^{1/2} (M(\varphi(l_i) - \varphi(l_{i+1})))^{1/2} \geqslant \| z_{i+1} - z_i \| .$$

Segue-se (lembre-se de que $ab \leqslant (a^2 + b^2)/2$),

$$\| z_i - z_{i-1} \| + M\varphi(l_i) - \varphi(l_{i+1})) \geqslant 2 \| z_{i+1} - z_i \| . \tag{2.20}$$

Esta desigualdade vale para $1 \leqslant i \leqslant k$; vamos somá-la sobre i:

$$\| z_1 - z_0 \| + M(\varphi(l_1) - \varphi(l_{k+1})) \geqslant \sum_{i=1}^{k} \| z_{i+1} - z_i \| + \| z_{k+1} - z_k \| .$$

Em vista das propriedades de monotonicidade de φ e l_k,

$$\| z_1 - z_0 \| + M(\varphi(l_0) \geqslant \sum_{i=1}^{k} \| z_{i+1} - z_i \| .$$

Finalmente

$$\| z_{i+1} - \bar{z} \| \leqslant \sum_{i=1}^{k} \| z_{i+1} - z_i \| + \| z_1 - \bar{z} \| \leqslant M(\varphi(l_0) + \| z_1 - z_0 \| + \| z_1 - \bar{z} \|,$$

o que implica $z_{k+1} \in B(\bar{z}, \rho)$ em vista de (2.16), (2.17), e (2.11). Isso completa a prova de (i).

Na verdade a desigualdade (2.20) vale para $i \geqslant 1$; vamos somá-la sobre i variando de algum k para algum $K > k$

$$\| z_k - z_{k-1} \| + M(\varphi(l_k) - \varphi(l_{k+1})) \geqslant \sum_{i=k}^{K} \| z_{i+1} - z_i \| + \| z_{K+1} - z_K \| .$$

Logo, como $\varphi(l_{k+1}) > 0$ temos

$$\| z_k - z_{k-1} \| + M\varphi(l_k) \geqslant \sum_{i=k}^{K} \| z_{i+1} - z_i \| .$$

Fazendo $K \to +\infty$, obtemos

$$\sum_{i=k}^{\infty} \| z_{i+1} - z_i \| \leqslant M\varphi(l_k) + \| z_k - z_{k-1} \| \tag{2.21}$$

Concluímos com (2.14), o que prova (ii):

$$\sum_{i=k}^{\infty} \| z_{i+1} - z_i \| \leqslant M(\varphi(l_k) + \sqrt{2r_+}\sqrt{l_{k-1}}.$$

Isto claramente implica que (z_k) é uma sequência de Cauchy. Como uma consequência da Proposição 7, obtemos que seu limite é um ponto crítico de L. □

Este teorema tem duas importantes consequências, como mostraremos nos teoremas seguintes.

Teorema 8. *(Convergência). Assuma que* L *satisfaz* $(H), (H_1)$ *e tem a propriedade Kurdyka-Lojasiewicz sobre cada ponto do domínio de* L*. Então ocorre apenas uma das alternativas abaixo:*

(i) $\| (x_k, y_k) \|$ *tende ao infinito;*

(ii) ou $(x_k - x_{k-1}, y_k - y_{k-1})$ *é* l_1*, i.e.,*

$$\sum_{k=1}^{\infty} \| x_{k+1} - x_k \| + \| y_{k+1} - y_k \| < +\infty,$$

e como consequência, $\{(x_k, y_k)\}$ *converge a um ponto crítico de* L*.*

Demonstração. Assuma que $\| (x_k, y_k) \|$ não tende a infinito, e seja $\bar{z}$ um ponto de acumulação de $\{(x_k, y_k)\}$ para o qual denotamos por ρ, η, φ os objetos associados definidos em (2.9). Note que a Proposição 2 implica que $\bar{z}$ é crítico e que $l_k = L(x_k, y_k)$ converge a $L(\bar{z})$.

Se existe um inteiro k_0 para o qual $L(x_{k_0}, y_{k_0}) = L(\bar{z})$, então é simples checar que (lembre-se de (2.2)) que $(x_k, y_k) = (x_{k_0}, y_{k_0})$ para todo $k \geqslant k_0$, logo que $(x_{k_0}, y_{k_0}) = \bar{z}$. podemos assumir então que $L(x_k, y_k) > L(\bar{z})$.

Como $\max(\varphi(l_k - L(\bar{z})), \| z_k - \bar{z} \|)$ admite 0 como ponto de acumulação, obtemos a existência de $k_0 \geqslant 0$ tal que (2.11) é preenchido com z_{k_0} como novo ponto inicial. A conclusão é então uma consequência do teorema 1. □

Observação 5. *(a) Várias hipóteses-padrão garantem automaticamente a limitação da seqüência* (x_k, y_k)*, daí a sua convergência: (i) Coercividade unilateral implica convergência. Assuma que f (ou g) tem compacto de baixo nível e que* $Q(x,y) = \frac{1}{2} \| x-y \|^2$*, isso implica que a sequência* (x_k, y_k) *é limitada (use o Lema 9).*

(ii) Convexidade implica convergência. Assuma que f, g são convexas e que Q *é da forma* $Q(x,y) = \frac{1}{2} \| Ax - By \|$*, onde* $A : \mathbb{R}^m \to \mathbb{R}^p$ *e* $B : \mathbb{R}^n \to \mathbb{R}^p$ *são aplicações lineares. Se L tem ao menos um mínimo, então a sequência* (x_k, y_k) *é limitada (ver Attouch et al. [3]).*

(b) O Teorema 2 dá novas percepções sobre métodos convexos alternados: a primeira mostra que a propriedade comprimento finito é satisfeita por muitas funções convexas (por exemplo, funções convexas definíveis;), mas também relaxa a suposição de Q ser quadrática, o que é exigido pelo algoritmo de minimização alternada de Attouch et al. [3, 2].

O seguinte resultado não deve ser considerado como um resultado clássico da convergência local; não assumimos aqui quaisquer condições padrão de não-degeneração, por exemplo, unicidade dos minimizadores, condições de segunda ordem, ou condições de transversalidade.

Teorema 9. *(Convergência local ao mínimo global). Assuma que* L satisfaz $(H), (H_1)$ e tem a propriedade Kurdyka-Łojasiewicz sobre $(\bar{x}, \bar{y})$, um mínimo global de L. Então existem ϵ, η tais que

$$\| (x_0, y_0) - (\bar{x}, \bar{y}) \| < \epsilon, \qquad \min L < L(x_0, y_0) < \min L + \eta$$

implicam que a sequência $\{(x_k, y_k)\}$ partindo de (x_0, y_0) tem a propriedade de comprimento finito e converge a (x^*, y^*) com $L(x^*, y^*) = \min L$

Demonstração. Uma aplicação direta do Teorema 1 nos dá a convergência de (x_k, y_k) a algum (x^*, y^*), um ponto crítico de L com $L(x^*, y^*) \in [\min L, \min L + \eta)$. Agora, se $L(x^*, y^*)$ não fosse igual a $L(\bar{x}, \bar{y})$ então a desigualdade Kurdyka-Łojasiewicz implicaria $\varphi'(L(x^*, y^*) - L(\bar{x}, \bar{y}))\text{dist}(0, \partial L(x^*, y^*)) \geqslant 1$, uma contradição pois $0 \in \partial L(x^*, y^*)$. □

Teorema 10. *(Taxa de convergência). Assuma que* L *satisfaz* $(H), (H_1)$*. Além disso, assuma que* $(x_k, y_k) \to (x_\infty, y_\infty)$ *e que* L *tem a propriedade Kurdyka-Łojasiewicz sobre* (x_∞, y_∞) *com* $\varphi(s) = cs^{1-\theta}, \theta \in [0, 1), c > 0$*. Então valem as seguintes estimativas:*

(i) Se $\theta = 0$*, então a sequência* (x_k, y_k) *converge em um número finito de passos.*

(ii) Se $\theta \in (0, 1/2]$*, então existe* $c > 0$ *e* $\tau \in [0, 1)$*, tais que*

$$\| (x_k, y_k) - (x_\infty, y_\infty) \| \leqslant c\tau^k.$$

(iii) Se $\theta \in (1/2, 1)$*, então existe* $c > 0$*, tal que*

$$\| (x_k, y_k) - (x_\infty, y_\infty) \| \leqslant c\tau^{-(1-\theta)/(2\theta-1)}.$$

Demonstração. As notações são aquelas do Teorema 7, e por simplicidade nós assumimos que $(l_k) \to 0$. Então $L(x_\infty, y_\infty) = 0$ (Proposição 7).

(i) Assuma primeiro $\theta = 0$. Se existe n_0 tal que (l_k) é constante para $k > n_0$, então (x_k, y_k) também é constante para $k > n_0$ em vista do Lema 9(i). Caso contrário, então a desigualdade Kurdyka-Łojasiewicz assegura para qualquer k suficientemente grande que $c\text{dist}(0, \partial L(x_k, y_k)) \geqslant 1$, uma contradição em vista do Lema 9(iii).

(ii,iii) Assuma $\theta > 0$. Para qualquer $k \geqslant 0$, seja $\Delta_k = \sum_{i=k}^{\infty} \sqrt{\| x_{i+1} - x_i \|^2 + \| y_{i+1} - y_i \|^2}$, o qual é finito pelo teorema 8. Como $\Delta_k \geqslant \sqrt{\| x_k - x_\infty \|^2 + \| y_k - y_\infty \|^2}$, ele é o suficiente para estimar Δ_k. Reescrevendo (2.21) com estas notações obtemos:

$$\Delta_k \leqslant M\varphi(l_k) + (\Delta_{k-1} - \Delta_k). \tag{2.22}$$

Usando a desigualdade Kurdyka-Łojasiewicz obtemos

$$\varphi'(l_k)\text{dist}(0, \partial L(x_k, y_k)) = c(1-\theta)l_k^{-\theta}\text{dist}(0, \partial L(x_k, y_k)) \geqslant 1.$$

$$l_k^{\theta} \leqslant c(1-\theta)\text{dist}(0, \partial L(x_k, y_k));$$

mas por (2.19) temos

$$\text{dist}(0, \partial L(x_k, y_k)) \leqslant \| (x_k^*, y_k^*) \| \leqslant (C + 1/r_+)(\Delta_{k-1} - \Delta_k).$$

Combinando as duas desigualdades acima nós obtemos para algum K positivo

$$\varphi(l_k) = MK(\Delta_{k-1} - \Delta_k)^{(1-\theta)\theta} + (\Delta_{k-1} - \Delta_k).$$

Sequências satisfazendo tais desigualdades tem sido estudadas em Attouch e Bolte [2], Theorem 2. Os itens (i) e (ii) seguem-se destes resultados. $\square$

2.3 Convergência de métodos de projeção alternados

Nesta seção consideramos o particular, mas importante caso de bifunções do tipo

$$L_{C,D}(x,y) = \delta_C(x) + 1/2 \parallel x - y \parallel^2 + \delta_D(y), (x,y) \in \mathbb{R}^n, \tag{2.23}$$

onde C, D são dois subconjuntos fechados não-vazios de $\mathbb{R}^n$. Note que L satisfaz (H) e (H_1) para qualquer $y_0 \in D$. Neste contexto específico, o algoritmo de minimização proximal alternado (2.1), (2.2) nos dá

$$x_{k+1} \in \operatorname{argmin}\left\{\frac{1}{2} \parallel u - y_k \parallel^2 + \frac{1}{2\lambda_k} \parallel u - x_k \parallel^2 : u \in C\right\},$$

$$y_{k+1} \in \operatorname{argmin}\left\{\frac{1}{2} \parallel v - x_{k+1} \parallel^2 + \frac{1}{2\mu_k} \parallel v - y_k \parallel^2 : v \in D\right\}.$$

Então obtemos o seguinte algoritmo de projeção alternado:

$$\begin{cases} x_{k+1} \in P_C\left(\frac{\lambda_k^{-1}x_k + y_k}{\lambda_k^{-1}+1}\right), \\ y_{k+1} \in P_D\left(\frac{\mu_k^{-1}y_k + x_{k+1}}{\mu_k^{-1}+1}\right). \end{cases}$$

O seguinte resultado ilustra o interesse do algoritmo acima para problemas de viabilidade.

Corolário 7. *(Convergência de Sequências). Assuma que a bifunção* $L_{C,D}$ *tem a propriedade de Kurdyka-Łojasiewicz em cada ponto. Então ou* $\parallel x_k - y_k \parallel \to 0$ *quando* $k \to \infty$ *ou* (x_k, y_k) *converge a um ponto crítico de L.*

(CONVERGÊNCIA LOCAL) Assuma que a bifunção $L_{C,D}$ tem a propriedade de Kurdyka-Łojasiewicz sobre (x^*, y^*) e que $\parallel x^* - y^* \parallel = \min\{\parallel x - y \parallel : x \in C, y \in D\}$. Se (x_0, y_0) está suficientemente próximo de (x^*, y^*), então (x_k, y_k) converge a um ponto (x_∞, y_∞), tal que $\parallel x_\infty - y_\infty \parallel = \min\{\parallel x - y \parallel : x \in C, y \in D\}$.

Demonstração. O primeiro ponto é devido ao Teorema 8, já o segundo segue de uma restrição do Teorema 9 a $L_{C,D}$. □

Referências Bibliográficas

[1] Attouch, H., J. Bolte, P. Redont, A. Soubeyran. 2010. Proximal Alternating Minimization and Projection Methods for Nonconvex Problems: An Approach Based on the Kurdyka-£ojasiewicz Inequality. *Mathematics of Operations Research.* Vol. 35, 438-457.

[2] Attouch, H., J. Bolte. 2009. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Math. Programming Ser. B* **116**(1-2) 5-16.

[3] Attouch, H., P.Redont, A. Soubeyran. 2007. A new class of alternating proximal algorithms with costs to move. *SIAM J. Optim.* **18**(3) 1061-1081.

[4] Attouch, H.,J. Bolte, P.Redont, A. Soubeyran. 2008. Alternating proximal algorithms for weakly coupled convex minimization problems. Applications to dynamical games and PDE's. *J. Convex Anal.* **15**(3) 485-506.

[5] Auslender, A. 1976. *Optimisation, Methodes Numeriques.* Masson, Paris.

[6] Bertsekas, D. 1999. *Nonlinear Optimisation*, 2nd ed. Athena, Belmont, MA.

[7] Bolte, J., A. Daniilidis, A. Lewis. 2006. The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J. Optim.* **17**(4) 1205-1223.

[8] Bolte, J., A. Daniilidis, A. Lewis. 2006. A nonsmooth Morse-Sard theorem for subanalytic functions. *J. Math. Anal. Appl.* **321**(2) 729-740.

[9] Bolte, J., A. Daniilidis, A. Lewis, M. Shiota. 2007. Clarke subgradients of stratifiable functions. SIAM J. Optim. 18(2) 556.572.

[10] Bolte, J., A. Daniilidis, O. Ley, L. Mazet. 2010. Characterizations of .ojasiewicz inequalities and applications: Subgradient flows, talweg, convexity. *Trans. Amer. Math. Soc.* **362** 3319-3363.

[11] Bochnak, J., Coste, M., Roy, M.-F. *Real Algebraic Geometry.* (Springer, 1998).

[12] Combettes, P. L., V. Wajs. 2005. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simulation* **4**(4) 1168-1200.

[13] Csiszár, I., G. Tusnady. 1984. Information geometry and alternating minimization procedures. *Statist. Decisions* (Suppl. 1) 205-237.

[14] Denkowska, Z., J. Stasica. 2008. *Ensembles Sous-Analytiques a la Polonaise, Travaux en Cours*, Vol. 69. Hermann, Paris.

[15] Donoho, D. L. 2006. Compressed sensing. *IEEE Trans. Inform. Theory* **4** 1289-1306.

[16] Grubisié, I., R. Pietersz. 2007. Efficient rank reduction of correlation matrices. *Linear Algebra Its Appl.* **422**(2.3) 629-653.

[17] Kurdyka, K. 1998. On gradients of functions definable in o-minimal structures. *Ann. Inst. Fourier* **48**(3) 769-783.

[18] Kurdyka, K. *Points d'un ensemble sous-analytique. Ann. Inst. Fourier*, 38, 1988, pp. 135-156.

[19] Lions, P. L. 1990. On the Schwarz alternating method. III. A variant for nonoverlapping subdomains. T. F. Chan, R. Glowinski, J. Periaux, O. Widlund, eds. *Third Internat. Sympos. Domain Decomposition Methods for Partial Differential Equations,* SIAM, Philadelphia, 202-231.

[20] Łojasiewicz, S. 1963. Une propriete topologique des sous-ensembles analytiques reels. *Les Equations aux Derivees Partielles.* Editions du centre National de la Recherche Scientifique, *Paris,* 87-89.

[21] Rockafellar, R. T., R. Wets. 1998. Variational analysis. *Grundlehren der Mathematischen Wissenschaften*, Vol. 317. Springer, Berlin.

[22] Tseng, P. 2001. Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.* **109**(3) 475-494.

[23] von Neumann, J. 1950. Functional operators. *Annals of Mathematics Studies,* Vol. 22. Princeton University Press, Princeton, NJ.

[24] Widrow, B., E. Wallach. 1996. *Adaptive Inverse Control.* Prentice-Hall, Englewood Cliffs, NJ.