UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

ABIEL COSTA MACEDO

Método de Ponto Proximal para Otimização

Goiânia
2009

Abiel Costa Macedo

Método de Ponto Proximal para Otimização

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

Área de concentração: Matemática Aplicada.

Orientador: Prof. Dr. Orizon Pereira Ferreira

Goiânia
2009

Abiel Costa Macedo

Método de Ponto Proximal para Otimização

Dissertação defendida no Programa de Pós–Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás como requisito parcial para obtenção do título de Mestre em Matemática, aprovada em 26 de fevereiro de 2009, pela Banca Examinadora constituída pelos professores:

Prof. Dr. Orizon Pereira Ferreira
Instituto de Matemática e Estatística – UFG
Presidente da Banca

Profa. Dra. Elizabeth Wegner Karas
Departamento de Matemática – UFPR

Prof. Dr. Luis Román Lucambio Pérez
Instituto de Matemática e Estatística – UFG

Abiel Costa Macedo

Graduou–se em Bacharel em Matemática pela UFG - Universidade Federal de Goiás. Durante sua graduação, foi monitor. Durante o Mestrado foi bolsista do CNPq.

Dedico este trabalho à minha
Família

Agradecimentos

A Deus.

Ao professor Orizon, meu orientador, pelos ensinamentos, pela paciência e pelo tempo dedicado.

Ao professor Glaydston pelo constante apoio e pelas dicas que me ajudaram muito.

A todos os professores, funcionários e amigos do Instituto de Matemática e Estatística da UFG.

Ao CNPq, pela ajuda financeira durante o curso.

E a todos que não citei, mas que de alguma forma contribuíram para a conclusão de mais uma etapa em minha vida.

Resumo

Macedo, Abiel Costa. **Método de Ponto Proximal para Otimização**. Goiânia, 2009. 60p. Dissertação de Mestrado. Instituto de Matemática e Estatística, Universidade Federal de Goiás.

O Método de Ponto Proximal é um método bastante estudado na Otimização contínua, em especial na Otimização convexa. Nosso objetivo é usá-lo para minimizar funções que não sejam necessariamente convexas e assim ampliar o alcance deste método. O trabalho será elaborado sobre funções dadas como o supremo de funções diferenciáveis. Com relação às funções diferenciáveis será exigido que o gradiente seja Lipschitz. Finalmente, mostraremos que o método, aplicado a esta classe de funções, fica bem definido e a sequência gerada converge para um ponto de mínimo.

Palavras–chave

Otimização Contínua, convergência local e Método de Ponto Proximal.

Abstract

Macedo, Abiel Costa. **Proximal Point Method for Optimization**. Goiânia, 2009. 60p. MSc. Dissertation. Instituto de Matemática e Estatística, Universidade Federal de Goiás.

The Proximal Point Method is quite studied in continuous Optimization, specially in the convex Optimization. Our goal is to use the method to minimize functions that are not necessarily convex, in order to expand the range of application of this method. The work will be presented for objective functions given by supreme of differentiable functions. With respect to these differentiable functions are required that its gradients are Lipschitz. For concluding, will be shown that the proximal point method is well-defined and the generated sequence converges for a minimizer of the objetive function.

Keywords

Continuous Optimization, local convergence and Proximal Point Method.

Sumário

Notações

Seguem aqui uma lista de notações utilizadas durante o texto.

1. $\mathbb{R}$ é o conjunto dos números reais.
2. $\mathbb{R}^n$ é o conjunto dos vetores reais da forma $x = (x_1, x_2, \ldots, x_n)$.
3. $||x||$ é a norma euclidiana de $x \in \mathbb{R}^n$.
4. $||x||_1$ é a norma da soma de $x \in \mathbb{R}^n$.
5. $\langle x, y \rangle$ é o produto interno euclidiano entre $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^n$.
6. e_i é um vetor de zeros, com o valor 1 na posição i.
7. $f'(x,d)$ é a derivada direcional de uma função convexa calculado em $x \in \mathbb{R}^n$ numa direção $d \in \mathbb{R}^n$.
8. $f^o(x,d)$ é a derivada direcional generalizada de uma função localmente Lipschitz calculado em $x \in \mathbb{R}^n$ numa direção $d \in \mathbb{R}^n$.
9. $\partial f(x)$ é o subdiferencial de uma função convexa calculado em $x \in \mathbb{R}^n$.
10. $\partial_c f(x)$ é o subdiferencial generalizado de Clarke de uma função localmente Lipschitz calculado em $x \in \mathbb{R}^n$.
11. $B_\delta(x)$ é a bola aberta com raio δ e centro $x \in \mathbb{R}^n$.
12. $B_\delta[x]$ é a bola fechada com raio δ e centro $x \in \mathbb{R}^n$.
13. $\text{Lip}_L(V)$ é o conjunto das função $g : \mathbb{R}^n \to \mathbb{R}^n$ Lipschitz em $V \subset \mathbb{R}^n$ com constante $L \geq 0$.
14. $\text{dom} f$ é o dominio efetivo, ou apenas domínio, de $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.
15. $L_f(c)$ é o cojunto de nível $c \in \mathbb{R}$ de f.
16. $\text{int}(V)$ é o conjunto dos pontos interiores de $V \subset \mathbb{R}^n$.

17. $\mathrm{conv}(V)$ é o fecho convexo do conjunto $V \subset \mathbb{R}^n$.

18. $f^* = \inf\{f(x) : x \in \mathbb{R}^n\}$.

19. $\arg\min_{x\in\mathbb{R}^n} \{g(x)\}$ são os pontos $x' \in \mathbb{R}^n$ onde g assume o valor mínimo.

20. $\mathcal{P}(\mathbb{R}^n)$ é o conjunto das partes de $\mathbb{R}^n$.

Introdução

Nesta dissertação estudaremos o *método de Ponto Proximal*, introduzido por Martinet [10] e Rockafellar [14], que se destina a resolver o problema

$$\min_{x \in \mathbb{R}^n} f(x). \tag{0-1}$$

O método de ponto proximal é reconhecidamente uma ferramenta eficiente na resolução de problemas de Otimização, através de aplicações direta ou no desenvolvimento de métodos para Lagrangeanos Aumentados clássicos ou generalizados. Os trabalhos [5] e [6] fazem um estudo detalhado do método de ponto proximal e diversas aplicações.

Muitos trabalhos foram produzidos com o objetivo de estender o domínio de aplicação deste método a outros contextos. Extensões particularmente interessantes são aquelas que tratam do estudo de aplicações locais deste método, entre os quais citamos Kaplan e Tichatschke [8], o qual é a base desta dissertação, Hare e Sagastizábal [3], que aplica o *método de Ponto Proximal* a uma classe de funções conhecida como "$Lower - C^2$". Ainda em um contexto mais amplo citamos os trabalhos de Pennanen [11], [12] e [13] que tratam da extensão do método (local) a operadores monótonos e a certas classes de operadores não-monótonos.

Nosso objetivo principal é aplicar o método para minimizar funções do tipo $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, que não sejam necessariamente convexas. Estas funções, sobre as quais é aplicado o método, são dadas como o supremo de funções diferenciáveis. Mais especificamente, sejam $T \subset \mathbb{R}^m$, $\Omega \subset \mathbb{R}^n$ um conjunto aberto e convexo e $\varphi : \mathbb{R}^n \times T \to \mathbb{R} \cup \{+\infty\}$. Suponhamos que a função

$$\begin{array}{rcl} \varphi_\tau : \mathbb{R}^n & \to & \mathbb{R} \cup \{+\infty\} \\ x & \mapsto & \varphi_\tau(x) := \varphi(x, \tau), \end{array}$$

seja continuamente diferenciável em Ω para todo $\tau \in T$. Defina $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ como

$$f(x) = \sup_{\tau \in T} \varphi_\tau(x).$$

A sequência produzida pelo *método de Ponto Proximal* é formalmente descrita do seguinte modo. Dados a sequência $\{\lambda_k\}$ de números reais positivos e um ponto de partida $x_0 \in \mathrm{dom}(f)$, as iteradas são produzidas pela seguinte relação:

$$x^{k+1} := \arg\min_{x\in\mathbb{R}^n} \left\{ f(x) + \frac{\lambda_k}{2}\|x - x^k\|^2 \right\}, \qquad k = 0,1,\ldots.$$

A boa definição da sequência gerada pelo método vem ao exigirmos, entre outras coisas, que o gradiente das funções φ_τ sejam Lipschitz e que f seja semicontínua inferiormente. Mais ainda, impondo algumas condições sobre f e seus subdiferenciais, em um determinado conjunto de nível, podemos garantir a convergência desta sequência para uma solução do problema (0-1).

Esta dissertação está dividida da seguinte forma. No Capítulo 1 é feita uma boa explanação sobre análise convexa onde são apresentados conceitos importantes como conjuntos convexos, funções convexas, *derivada direcional* e *subdiferencial* de funções convexas, com o intuito principal de não deixar dúvidas com respeito aos resultados utilizados nos Capítulos 4 e para tornar o trabalho auto contido. Quando é dito "não deixar dúvidas", faz-se referência a forma da função objetivo $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ pois é pouco comum na literatura trabalhar-se com este tipo de função, a qual acarreta maiores detalhes nas provas de alguns resultados clássicos para funções a valores reais, isto é, funções do tipo $f : \mathbb{R}^n \to \mathbb{R}$. Um boa referência que trabalha com este tipo de função é o livro de Hiriart-Urruty, J.-B e Lemaréchal, C. [4].

No Capítulo 2 fazemos uma análise de convergência do método de ponto proximal para problemas de Otimização convexa.

No Capítulo 3 fazemos uma introdução à Teoria Diferencial Generalizada, onde é estendido os conceitos de diferenciabilidade de funções convexas para funções localmente Lipschitz. Esta extensão é devida a Clarke, F. H. [2].

Por fim no Capítulo 4 definimos formalmente as funções às quais aplicamos o método de ponto proximal e ainda é enunciamos e demonstramos o teorema de convergência do método, que é o resultado principal desta dissertação.

CAPÍTULO 1

Análise Convexa

Neste capítulo apresentamos alguns conceitos de análise e análise convexa, como conjuntos convexos, funções convexas, *derivada direcional* e *subdiferencial* de funções convexas.

1.1 Elementos de Análise

Nesta seção apresentamos alguns conceitos básicos de análise que terão uma importância crucial neste trabalho. Neste seção fizemos uso das seguintes referências [1], [7].

Definição 1.1 *Seja $W \subset \mathbb{R}^n$ um subespaço linear. Dizemos que $p : W \to \mathbb{R}$ é um funcional subaditivo se para todo $x, y \in W$ tivermos*

$$p(x+y) \leq p(x) + p(y).$$

Dizemos também que p é positivamente homogêneo se para todo $x \in \mathbb{R}^n$ e $t \in \mathbb{R}_+ \setminus \{0\}$

$$p(tx) = tp(x).$$

Proposição 1.2 *Sejam $W \subset \mathbb{R}^n$ um subespaço linear e $p : W \to \mathbb{R}$ um funcional subaditivo, positivamente homogêneo. Então temos as seguintes propriedades.*

i) $p(0) = 0$,

ii) $p(-x) \geq -p(x)$,

iii) para todo $t \in \mathbb{R}$ temos $p(tx) \geq tp(x)$.

Prova. (i) Tome $x = 0$. Como p é positivamente homogêneo temos que $p(0) = tp(0)$, qualquer que seja $t > 0$. Logo $p(0) = 0$. (ii) Agora, de i, juntamente com a subaditividade de p, obtemos que $0 = p(x-x) \leq p(x) + p(-x)$ então $p(-x) \geq -p(x)$. (iii) Combinando os itens i e ii com p ser positivamente homogêneo o resultado segue. $\square$

Definição 1.3 *(Funcional Linear) Seja $W \subset \mathbb{R}^n$ um subespaço linear. Dizemos que $h : W \to \mathbb{R}$ é um funcional linear se para todo $x, y \in W$ e $t, r \in \mathbb{R}$ tivermos*

$$h(tx + ry) = th(x) + rh(y).$$

Teorema 1.4 *(Hahn-Banach) Sejam $W \subset \mathbb{R}^n$ um subespaço linear próprio não vazio, $p : \mathbb{R}^n \to \mathbb{R}$ um funcional subaditivo, positivamente homogêneo e $h : W \to \mathbb{R}$ um funcional linear tais que*

$$h(x) \leq p(x), \qquad \forall x \in W.$$

Então existe um funcional linear $H : \mathbb{R}^n \to \mathbb{R}$ tal que

$$H(x) = h(x), \quad \forall x \in W \qquad e \qquad H(y) \leq p(y), \quad \forall y \in \mathbb{R}^n. \tag{1-1}$$

Prova. Suponhamos inicialmente que $m < n$ seja a dimensão de W. Assim tome $z \in \mathbb{R}^n \setminus W$ e W' o subespaço gerado por z e W. Desta forma, se $w \in W'$ então $w = tz + x$, para algum $x \in W$ e $t \in \mathbb{R}$. Agora defina h' como

$$h'(tz + x) = th'(z) + h(x), \qquad x \in W, \quad t \in \mathbb{R},$$

onde $h'(z) = c$ será um número constante tal que

$$tc + h(x) \leq p(tz + x), \qquad \forall x \in W, \quad t \in \mathbb{R}. \tag{1-2}$$

Com h' assim definido, h' é um funcional linear sobre W' satisfazendo

$$h'(x) = h(x), \quad \forall x \in W \qquad e \qquad h'(y) \leq p(y), \quad \forall y \in W'.$$

Assim falta apenas mostrar que sempre é possível escolher este número c. Para isso, observe que (1-2) é equivalente a

$$c \leq p(z + \frac{x}{t}) - h(\frac{x}{t}),$$

para $t > 0$. Para $t < 0$, pela Proposição 1.2 item *ii*, (1-2) é equivalente a

$$c \geq -p(-z - \frac{x}{t}) - h(\frac{x}{t}).$$

Para o caso em que $t = 0$ qualquer numero c serve. Então, dados $y', y'' \in W$, pela Proposição 1.2 item *ii* teremos

$$h(y'') - h(y') \leq p(y'' - y') = p(y'' + z - y' - z) \leq p(y'' + z) + p(-y' - z),$$

que implica

$$-h(y'') + p(y'' + z) \geq -h(y') - p(-y' - z).$$

Fazendo $c'' = \inf_{y'' \in W}(-h(y'') + p(y'' + z))$ e $c' = \sup_{y' \in W}(-h(y') - p(-y' - z))$ teremos que $c'' \geq c'$. Assim tomando $c \in [c'', c']$, (1-2) é satisfeita. Repetindo este processo um número finito de vezes encontramos um funcional linear H que satisfaz (1-1). $\square$

Teorema 1.5 *(Riesz) Seja $h : \mathbb{R}^n \to \mathbb{R}$ um funcional linear. Então existe $s \in \mathbb{R}^n$ tal que*

$$h(x) = \langle s, x \rangle, \qquad \forall x \in \mathbb{R}^n.$$

Prova. Seja $e_1, e_2, \ldots, e_n$ a base canônica de $\mathbb{R}^n$. Assim qualquer que seja $x \in \mathbb{R}^n$ existem $x_1, x_2, \ldots, x_n$ tais que $x = x_1e_1 + x_2e_2 + \cdots + x_ne_n$. Portanto tomando o vetor $s = (h(e_1), h(e_2), \ldots, h(e_n))$ o resultado segue. $\square$

Definição 1.6 *Sejam $g : \mathbb{R}^n \to \mathbb{R}^n$, $V \subset \mathbb{R}^n$ e $L \geq 0$ uma constante. Dizemos que g é Lipschitz em V, com constante $L \geq 0$, se*

$$\|g(x) - g(y)\| \leq L\|x - y\|, \qquad \forall\, x, y \in V.$$

Denotamos o conjunto de todas as funções Lipschitz em V, com constante L, por $\mathrm{Lip}_L(V)$. É importante ressaltar que nesta dissertação tabalhamos com funções definida em $\mathbb{R}^n$ que tomam valores em $\mathbb{R} \cup \{+\infty\}$, isto é,

$$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}.$$

Chamamos de *domínio efetivo*, ou apenas *domínio*, de f o conjunto

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

Vejamos agora alguns conceitos e resultados para este tipo de função.

Definição 1.7 *Sejam $\Omega \subset \mathbb{R}^n$ e $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Dizemos que $\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ é a restrição de f ao conjunto Ω se*

$$\tilde{f}(x) = \begin{cases} f(x), & x \in \Omega, \\ +\infty, & \textit{caso contrário}. \end{cases}$$

Usamos a seguinte notação $\tilde{f} = f_{|\Omega}$ para a restrição de f ao conjunto Ω.

Com relação à diferenciabilidade de $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ em um ponto consideraremos o conceito usual de diferenciabilidade, notando apenas que este ponto deve pertencer ao interior do domínio de f. Naturalmente se define a diferenciabilidade sobre um conjunto $V \subset \mathbb{R}^n$, novamente, notando apenas que $V \subset \operatorname{int}(\operatorname{dom} f)$.

Definição 1.8 *Sejam $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, $V \subset \mathbb{R}^n$ e $L \geq 0$ uma constante. Dizemos que f é Lipschitz em V, com constante $L \geq 0$, se $V \subset dom(f)$ e*

$$|f(x) - f(y)| \leq L\|x - y\|, \qquad \forall\, x, y \in V.$$

A definição seguinte é dada para caracterizar a notação usada.

Definição 1.9 *Defina bola aberta e bola fechada com centro em $x \in \mathbb{R}^n$ e raio δ como sendo, respectivamente,*

$$B_\delta(x) = \{y \in \mathbb{R}^n : \|x - y\| < \delta\}, \qquad B_\delta[x] = \{y \in \mathbb{R}^n : \|x - y\| \leq \delta\}.$$

Definição 1.10 *A função $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ é dita Lipschitz em $x \in \mathbb{R}^n$, com constante $L_x \geq 0$, se existe $\delta > 0$ tal que $f \in Lip_{L_x}(B_\delta(x))$. E ainda, f é dita localmente Lipschitz em V se ela é Lipschitz em cada ponto $x \in V$ e localmente Lipschitz se ela é localmente Lipschitz em* $\operatorname{dom} f$.

Observação 1.11 *Note que o domínio de uma função localmente Lipschitz é aberto.*

Definição 1.12 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Dizemos que f é semicontínua inferiormente em $x \in \mathbb{R}^n$ quando*

$$\liminf_{y \to x} f(y) \geq f(x).$$

Dizemos ainda que f é semicotínua inferiormente se é semicotínua inferiormente $\forall x \in \mathbb{R}^n$.

Definição 1.13 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Dizemos ainda que f é semicontínua superiormente em $x \in \mathbb{R}^n$ quando*

$$\limsup_{y \to x} f(y) \leq f(x).$$

Dizemos ainda que f é semicotínua superiormente se é semicotínua superiormente $\forall x \in \mathbb{R}^n$.

1.2 Conjuntos Convexos

Nesta seção definiremos o que vem a ser conjuntos convexos e daremos algumas propriedades. Neste seção fizemos uso das seguintes referências [15], [14] e [4].

Definição 1.14 *Um conjunto $C \subset \mathbb{R}^n$ é dito convexo se, para todo $\alpha \in [0,1]$ e $x, x' \in C$, $\alpha x + (1-\alpha)x' \in C$.*

Definição 1.15 *Seja $X \in \mathbb{R}^n$ um conjunto não vazio. Chamamos de fecho convexo de X o conjunto*

$$\operatorname{conv}\{X\} = \{\sum_{i=1}^{r} \alpha_i x_i : \alpha_i \in [0,1],\ x_i \in X,\ i = 1,\ 2, \ldots,\ r, \quad \sum_{i=1}^{r} \alpha_i = 1\}.$$

Proposição 1.16 *Seja $X \in \mathbb{R}^n$ um conjunto compacto. Então o fecho convexo de X é compacto.*

Prova. Primeiro mostraremos que, se $x \in \operatorname{conv}\{X\}$ então existem $\{y^1, y^2, \ldots, y^{n+1}\} \subset X$ e $\lambda_i \in [0,1]$, $i = 1, 2, \ldots, n, n+1$, tais que

$$x = \sum_{i=1}^{n+1} \lambda_i y^i, \qquad \sum_{i=1}^{n+1} \lambda_i = 1.$$

Mostremos, pela definição de fecho convexo existem $\{x^1, x^2, \ldots, x^r\} \subset X$ e $\lambda_i \in [0,1]$, $i = 1, 2, \ldots, r$, tais que

$$x = \sum_{i=1}^{r} \lambda_i x^i, \qquad \sum_{i=1}^{r} \lambda_i = 1.$$

De fato, para $r \leq n+1$ o resultado é obvio. Suponhamos então que $r > n+1$, onde $\lambda_i \neq 0$, $i = 1, 2, \ldots, r$. Assim $\{x^1 - x^r, x^2 - x^r, \ldots, x^{r-1} - x^r\} \subset \mathbb{R}^n$ é um conjunto linearmente dependente, isto é, existe $\{\gamma_1, \gamma_2, \ldots, \gamma_{r-1}\}$, $\gamma_i \neq 0$ para algum i, tal que

$$\sum_{i=1}^{r-1} \gamma_i (x^i - x^r) = 0. \tag{1-3}$$

Agora, tome $\gamma_r = -\sum_{i=1}^{r-1} \gamma_i$, faça $(\gamma_{i_m}/\lambda_{i_m}) = \max_{i=1,2,\ldots,r}(\gamma_i/\lambda_i)$ e defina

$$\beta_i = \lambda_i \left(1 - \frac{\lambda_{i_m}}{\gamma_{i_m}} \frac{\gamma_i}{\lambda_i}\right) = \lambda_i - \frac{\lambda_{i_m}}{\gamma_{i_m}} \gamma_i.$$

Consequentemente $\beta_i \geq \beta_{i_m} = 0$, e ainda

$$\sum_{i=1}^{r} \beta_i = \sum_{i=1}^{r} \lambda_i - \frac{\lambda_{i_m}}{\gamma_{i_m}} \sum_{i=1}^{r} \gamma_i = \sum_{i=1}^{r} \lambda_i = 1,$$

que, juntamente com (1-3), nos dá

$$\sum_{i=1}^{r} \beta_i x^i = \sum_{i=1}^{r} \lambda_i x^i - \frac{\lambda_{i_m}}{\gamma_{i_m}} \sum_{i=1}^{r} \gamma_i x^i = \sum_{i=1}^{r} \lambda_i x^i = x.$$

Assim, por $\beta_{i_m} = 0$, podemos reindexar a sequência $\{\beta_i\}$ tal que

$$\sum_{i=1}^{r-1} \beta_i x^i = x, \qquad \sum_{i=1}^{r-1} \beta_i = 1.$$

Repetindo este processo um número finito de vezes obteremos $\lambda_i > 0$, para $i = 1, 2, \ldots, n+1$, e um conjunto $\{y^1, y^2, \ldots, y^{n+1}\} \subset \{x^1, x^2, \ldots, x^r\} \subset X$, tais que

$$\sum_{i=1}^{n+1} \lambda_i y^i = x, \qquad \sum_{i=1}^{n+1} \lambda_i = 1.$$

Sendo assim, tome agora $\{x^k\} \subset \text{conv}\{X\}$ uma sequência qualquer. Então existem $\lambda_i^k > 0$ e $x_i^k \in X$ para $i = 1, 2, \cdots, n+1$, tais que

$$\sum_{i=1}^{n+1} \lambda_i^k x_i^k = x^k, \qquad \sum_{i=1}^{n+1} \lambda_i^k = 1.$$

Como X é compacto e $(\lambda_1^k, \lambda_2^k, \ldots, \lambda_{n+1}^k) \in B_1[0]$ para todo k, na norma da soma, podemos obter subsequências $\{x^{k_j}\}$ e $\{\lambda_i^{k_j}\}$ tais que

$$\lim_{j\to\infty} x_i^{k_j} = \bar{x}^i \in X, \qquad \lim_{j\to\infty} \lambda_i^{k_j} = \bar{\lambda}_i \geq 0,$$

para cada i, com $\sum_{i=1}^{n+1} \bar{\lambda}_i = 1$. Então

$$\lim_{j\to\infty} x^{k_j} = \lim_{j\to\infty} \sum_{i=1}^{n+1} \lambda_i^{k_j} x_i^{k_j} = \sum_{i=1}^{n+1} \bar{\lambda}_i \bar{x}^i = \bar{x},$$

implicando que $\bar{x} \in \text{conv}X$. Portanto, toda sequência contida em $\text{conv}X$ possui uma subsequência convergente, isto é, $\text{conv}X$ é compacto. $\square$

Definição 1.17 *(Funcional de Minkowski de um conjunto convexo) Seja $C \subset \mathbb{R}^n$ um conjunto convexo e aberto tal que $0 \in C$. Para todo $x \in \mathbb{R}^n$ defina $p : \mathbb{R}^n \to \mathbb{R}$ por*

$$p(x) = \inf\{\alpha > 0 : \alpha^{-1} x \in C\}.$$

O funcinal p é denominado *funcional de Minkowski* de C. A proposição seguinte item i implica que o funcional de Minkowski está bem definido.

Proposição 1.18 *Sejam $C \subset \mathbb{R}^n$ um conjunto convexo e aberto, tal que $0 \in C$, e p o funcinal de Minkowski de C. Então temos as seguintes propriedades*

i) existe $M > 0$ tal que $0 \leq p(x) \leq M\|x\|$ qualquer seja $x \in \mathbb{R}^n$;

ii) $C = \{x \in \mathbb{R}^n : p(x) < 1\}$*;*

iii) *p é um funcional subaditivo e positivamente homogêneo.*

Prova. (i) Como C é aberto e $0 \in C$, existe $\delta_0 > 0$ tal que $B_{\delta_0}(0) \subset C$. Assim, para todo $x \in \mathbb{R}^n \setminus \{0\}$, temos que $\delta_0 x/(2\|x\|) \in C$ implicando que

$$p(x) \leq \frac{2}{\delta_0}\|x\|, \qquad \forall x \in \mathbb{R}^n.$$

Portanto tome $M = 2/\delta_0$.
(ii) Se $x \in C$, como C é aberto, $(1+\varepsilon)x \in C$, para $\varepsilon > 0$ suficientemente pequeno. Assim $p(x) \leq 1/(1+\varepsilon) < 1$, isto é,

$$C \subset \{x \in \mathbb{R}^n : p(x) < 1\}.$$

Agora, se $p(x) < 1$, existe $\alpha \in (p(x), 1)$ tal que $\alpha^{-1}x \in C$. Por outo lado, como C é convexo, $x = \alpha(\alpha^{-1}x) + (1-\alpha)0 \in C$. O que mostra a igualdade.
(iii) Dado $t > 0$ e $x \in \mathbb{R}^n$ faça $y = tx$. Então

$$p(tx) = p(y) = \inf\{\alpha > 0 : \frac{y}{\alpha} \in C\} = \inf\{t\alpha' > 0 : \frac{x}{\alpha'} \in C\} = tp(x), \qquad \alpha' = \frac{\alpha}{t}.$$

Agora tome $x, y \in \mathbb{R}^n$ e $\varepsilon > 0$ arbitrários. Pela definição de p existem $r', r'' > 0$ tais que $p(x) < r' < p(x) + \varepsilon$, $p(y) < r'' < p(y) + \varepsilon$ e $x/r', y/r'' \in C$. Para $r = r' + r''$, a convexidade de C implica que

$$\frac{x+y}{r} = \frac{r'}{r}\frac{x}{r'} + \frac{r''}{r}\frac{y}{r''} \in C,$$

Assim $p(x+y) \leq r = r' + r'' < p(x) + p(y) + 2\varepsilon$. Como ε é arbitrário o resultado segue-se.
□
O próximo teorema também é conhecido por Teorema da Separação.

Teorema 1.19 *(Primeira forma geométrica de Hahn-Banach) Sejam* $C_1 \subset \mathbb{R}^n$ *e* $C_2 \subset \mathbb{R}^n$ *conjuntos convexos não vazio e disjuntos. Suponhamos também que* C_2 *seja aberto. Então existem* $a \in \mathbb{R}^n \setminus \{0\}$ *e* $c \in \mathbb{R}$ *tais que*

$$\langle a, x \rangle \leq c \leq \langle a, y \rangle \quad \forall x \in C_1,\ \forall y \in C_2.$$

Para demonstrar este teorema usaremos o seguinte resultado.

Lema 1.20 *Sejam* $C \subset \mathbb{R}^n$ *um conjunto convexo, aberto e não vazio e* $x_0 \in \mathbb{R}^n \setminus C$*. Então existe* $a \in \mathbb{R}^n \setminus \{0\}$ *tal que* $\langle a, x \rangle < \langle a, x_0 \rangle$ *para todo* $x \in C$*.*

Prova. Inicialmente tome $y_0 \in C$ e faça $C' = C - y_0 = \{x - y_0 :\ x \in C\}$. Deste modo, temos que $x' = x_0 - y_0 \notin C'$ e C' é convexo, aberto e não vazio. Assim tomando $W = \{tx' : t \in \mathbb{R}\}$,

$h : W \to \mathbb{R}$ um funcional linear, definido por $h(x) = h(tx') = t$, e p o funcional de Minkowski de C' temos, pela Proposição 1.2 item *iii*, que

$$p(x) = p(tx') \geq t\,p(x') \geq t = h(x).$$

Então, pelo Teorema de Hahn-Banach (1.4), existe $H : \mathbb{R}^n \to \mathbb{R}$, funcional lienar, tal que

$$H(x) \leq p(x), \qquad \forall x \in \mathbb{R}^n, \quad H(x) = h(x) \qquad \forall x \in W.$$

Assim, pela Proposição 1.18 item *ii*, $H(x) < 1 = H(x')$ para todo $x \in C'$, isto é, $H(x) < H(x_0)$ para todo $x \in C$. Portanto usando o Teorema de Riesz 1.5 o resultado segue-se. □

Prova.(Teorema 1.19) Primeiro façamos $C = C_2 - C_1 = \{x - y :\ x \in C_2,\ x \in C_1\}$. Assim $0 \notin C$ e C é aberto, convexo, pois C_1 e C_2 são convexos, disjuntos e C_2 é aberto ($C = \bigcup_{y \in C_1} C_2 - y$). Então, pelo Lema 1.20 existe $a \in \mathbb{R}^n \setminus \{0\}$ tal que

$$\langle a, x\rangle < 0, \qquad \forall x \in C,$$

isto é,

$$\langle a, x\rangle < \langle a, y\rangle \quad \forall x \in C_1,\ \forall y \in C_2.$$

Tomando $c \in [\sup_{x \in C_1} \langle a, x\rangle, \inf_{y \in C_2} \langle a, y\rangle]$ obtemos o resultado. □

O próximo teorema também é conhecido por Teorema da Separação Estrita.

Teorema 1.21 *(Segunda forma geométrica de Hahn Banach) Sejam $C_1 \subset \mathbb{R}^n$ e $C_2 \subset \mathbb{R}^n$ conjuntos convexos não vazio e disjuntos. Suponhamos também que C_1 seja fechado e C_2 compacto. Então existem $a \in \mathbb{R}^n \setminus \{0\}$ e $c \in \mathbb{R}$ tais que*

$$\langle a, x\rangle < c < \langle a, y\rangle \quad \forall x \in C_1,\ \forall y \in C_2.$$

Prova. Primeiramente observe que, por C_1 é fechado, C_2 é compacto e $C_1 \cap C_2 = \emptyset$, existe $\varepsilon > 0$ tal que $C_1' = C_1 + B_\varepsilon(0)$ e $C_2' = C_2 + B_\varepsilon(0)$ são disjuntos. Pois, se não, existiriam sequencias $x^k \in C_1$ e $y^k \in C_2$ tais que $\|x^k - y^k\| < 1/k$, como C_2 é compacto, passando a uma subsequência temos que $\lim_{i \to \infty} y^{k_i} = \bar{y}$, então $x^{k_i} \to \bar{y}$ e assim, por C_1 e C_2 serem fechados, $\bar{y} \in C_1 \cap C_2$ que não pode ser. Assim C_1' e C_2' são abertos, convexos e disjunto. Então, pelo o Teorema 1.19, existem $a \in \mathbb{R}^n$ e $c \in \mathbb{R}$ tais que

$$\langle a, x\rangle \leq c \leq \langle a, y\rangle \quad \forall x \in C_1',\ \forall y \in C_2',$$

então

$$\left\langle a, x + \frac{\varepsilon}{2} z \right\rangle \leq c \leq \left\langle a, y + \frac{\varepsilon}{2} w \right\rangle, \quad \forall x \in C_1, \quad \forall y \in C_2, \quad \forall z, w \in \partial B_1(0),$$

onde $\partial B_1(0)$ denota a fronteira da bola. Observando que $a \in \mathbb{R}^n \setminus \{0\}$, faça $z = a/\|a\|$ e $w = -a/\|a\|$ o que nos dá

$$\langle a, x \rangle + \frac{\varepsilon}{2}\|a\| \leq c \leq \langle a, y \rangle - \frac{\varepsilon}{2}\|a\| \quad \forall x \in C_1, \, \forall y \in C_2.$$

Assim o teorema fica demonstrado. $\square$

Definição 1.22 *Seja $C \subset \mathbb{R}^n$ um conjunto convexo. Dizemos que $v \in \mathbb{R}^n$ é uma direção de recessão de C quando*

$$x + tv \in C, \qquad \forall x \in C, \, \forall t \in \mathbb{R}_+.$$

Proposição 1.23 *Seja $C \subset \mathbb{R}^n$ um conjunto convexo, fechado, não vazio. Então C é ilimitado se, somente se, existe uma direção de recessão $v \in \mathbb{R}^n$.*

Prova. Se existe uma direção de recessão $v \in \mathbb{R}^n$, C é ilimitado, pois C conterá pelo menos uma semi reta.

Suponhamos que C é ilimitado. Assim tome $x \in C$ fixo e $\{x^k\} \subset C \setminus \{x\}$, uma sequência tal que $\|x^k - x\|$ tende para infinito. Tome ainda $\{v^k\}$, definida por $v^k = (x^k - x)/\|x^k - x\|$, e v um ponto de acumulação de $\{v^k\}$. Observe que, pela convexidade de C,

$$x + tv^k = \left(1 - \frac{t}{\|x^k - x\|}\right) x + \frac{t}{\|x^k - x\|} x^k \in C,$$

para k suficientemente grande. Como C é fechado, $x + tv \in C$.

Agora mostraremos que $y + tv \in C$ qualquer que seja $y \in C$ e $t \geq 0$. Assim tome $y \in C$ e $t > 0$, se $y = x + t'v$ para algum $t' \geq 0$ o resultado é imediato. Então, supondo que $y \neq x + t'v$ qualquer que seja $t' \geq 0$, faça $\{y^k\} \subset C$ tal que $y^k = x + ktv$ para todo k e

$$v^k = \frac{\|v\|}{\|y^k - x\|}(y^k - y) = v + \frac{\|v\|}{\|y^k - x\|}(x - y).$$

Claramente $\lim_{k\to\infty} v^k = v$. Então, por C é convexo, fechado e

$$y + tv^k = \left(1 - \frac{t\|v\|}{\|y^k - x\|}\right) y + \frac{t\|v\|}{\|y^k - x\|} y^k \in C,$$

para k maior que algum k', segue-se que $y + tv = \lim_{k\to\infty} y + tv^k \in C$. O que conclui a prova. $\square$

1.3 Funções Convexas

Nesta seção definiremos funções convexas e apresentamos alguns resultados importantes sobre as mesmas. Neste seção fizemos uso das seguintes referências [15], [14] e [4].

Definição 1.24 *Dada $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ temos que*

i) f é convexa se para todo $x,\ y \in \mathbb{R}^n$

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y), \qquad \forall \alpha \in [0,1].$$

Se a desigualdade for estrita para todo $\alpha \in (0,1)$ e x,y distintos dizemos que f é estritamente convexa.

ii) f é fortemente convexa, com módulo $\mu > 0$, se para todo $x,\ y \in \mathbb{R}^n$

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y) - \mu\alpha(1-\alpha)\|x-y\|^2, \quad \forall \alpha \in [0,1]. \quad \text{(1-4)}$$

Observação 1.25 *Se uma função f é convexa o $\operatorname{dom} f$ é um conjunto convexo. Pois se f é convexa e $x,y \in \operatorname{dom} f$ então $f(\alpha x + (1-\alpha)y) \in \operatorname{dom} f$ para $\alpha \in [0,1]$.*

Observação 1.26 *Sejam $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convexa e $C \subset \operatorname{dom} f$ um conjunto convexo. Então a função $\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ definida por*

$$\tilde{f}(x) = \begin{cases} f(x), & x \in C, \\ +\infty, & \textit{caso contrário}, \end{cases}$$

que é a restrição de f a C, é convexa. Esta afirmação é imediatamente verificada. O mesmo vale para funções estritamente convexas e fortemente convexas.

Teorema 1.27 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função convexa. Então, para toda coleção de pontos $\{x_1, x_2, \ldots, x_k\}$ em $\mathbb{R}^n$ e toda coleção $\{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ em $\mathbb{R}$ satisfazendo $\alpha_i \geq 0$, $i = 1, 2, \ldots, k$, e*

$$\sum_{i=1}^{k} \alpha_i = 1,$$

vale a desigualdade de Jensen, isto é,

$$f(\sum_{i=1}^{k} \alpha_i x_i) \leq \sum_{i=1}^{k} \alpha_i f(x_i).$$

Prova. Primeiro para $k = 2$ é imediato, sai da definição. Agora, suponha indutivamente que a relação é verdadeira para $k-1$. Tome agora as coleções $\{x_i\}$ e $\{\alpha_i\}$, com $\alpha_i \geq 0$

para $i = 1,2,...,k$ e $\sum_{i=1}^{k} \alpha_i = 1$. Se α_k é 0 ou 1, então nada há para provar. Se não, fixe

$$\alpha := \sum_{i=1}^{k-1} \alpha_i,$$

e faça $\overline{\alpha_i} := \alpha_i/\alpha$ para $i = 1,...,k-1$. Observe que $\overline{\alpha}_i \geq 0$ e $\sum_{i=1}^{k-1} \overline{\alpha}_i = 1$. Assim,

$$\sum_{i=1}^{k} \alpha_i x_i = \alpha \sum_{i=1}^{k-1} \overline{\alpha}_i x_i + (1-\alpha) x_k.$$

Aplicando f no somatório e usando a convexidade de f e a hipótese de indução chegamos

$$f(\sum_{i=1}^{k} \alpha_i x_i) \leq \alpha f(\sum_{i=1}^{k-1} \overline{\alpha}_i x_i) + \alpha_k f(x_k) \leq \sum_{i=1}^{k-1} \alpha_i f(x_i) + \alpha_k f(x_k) = \sum_{i=1}^{k} \alpha_i f(x_i),$$

o que conclui a prova. □

Definição 1.28 *Diz-se que uma função* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *é localmente limitada em* $V \subset \operatorname{dom} f$ *quando para todo* $x_0 \in V$ *existem* $c > 0$ *e* $\delta > 0$ *tais que* $|f(x)| \leq c$*, para todo* $x \in B_\delta(x_0) \subset V$.

Teorema 1.29 *Se* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *é convexa, então* f *é localmente limitada no interior do domínio de* f.

Prova. Dado $x_0 \in \operatorname{int}(\operatorname{dom} f)$, assim é possível tomar um "cubo" $C \subset \operatorname{dom} f$, de vértices $x_1, x_2, ..., x_m$, $(m = 2^n)$ e com centro em x_0. Assim para todo ponto $x \in C$ podemos encontrar escalares $\alpha_1, \alpha_2, ..., \alpha_m$, tais que

$$x = \sum_{i=1}^{m} \alpha_i x_i, \qquad \alpha_i \geq 0, \quad \sum_{i=1}^{m} \alpha_i = 1,$$

pois $C = \operatorname{conv}\{x_1, x_2, ..., x_m\}$. Sendo f uma função convexa vale a desigualdade de *Jensen*

$$f(x) \leq \sum_{i=1}^{m} \alpha_i f(x_i) \leq \max_{1 \leq i \leq n} f(x_i) := \tilde{c}.$$

Por outro lado, para qualquer $x \in C$, podemos escolher $y \in C$, tal que $x_0 = (1/2)x + (1/2)y$. Então

$$f(x_0) \leq \frac{1}{2} f(x) + \frac{1}{2} f(y).$$

Mas isso implica

$$f(x) \geq 2(f(x_0) - \frac{1}{2} f(y)) \geq 2 f(x_0) - \tilde{c}.$$

Portanto, qualquer que seja $x \in \operatorname{int}(\operatorname{dom} f)$ é possível tomar $c > 0$ e $\delta > 0$ tais que $|f(x)| \leq c$, para todo $x \in B_\delta(x_0) \subset \operatorname{int}(\operatorname{dom} f)$, isto é, f é localmente limitada em $\operatorname{int}(\operatorname{dom} f)$. $\square$

Teorema 1.30 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função convexa. Então f é localmente lipschitz no interior do domínio de f.*

Prova. Dado $x \in \operatorname{int}(\operatorname{dom} f)$ pelo teorema anterior existem $c > 0$ e $\delta > 0$, tais que $|f(x)| \leq c$, para todo $x \in B_{2\delta}(x)$. Assim tome $y, y' \in B_\delta(x_0)$, com $y \neq y'$, e faça

$$y'' := y' + \delta \frac{y' - y}{\|y' - y\|} \in B_{2\delta}(x_0).$$

Observe que $y' \in [y, y'']$, segmento ligando y e y'', mais precisamente,

$$y' = \frac{\|y' - y\|}{\delta + \|y' - y\|} y'' + \frac{\delta}{\delta + \|y' - y\|} y.$$

Aplicando a convexidade de f e fazendo algumas manipulações, obtemos

$$f(y') - f(y) \leq \frac{\|y' - y\|}{\delta + \|y' - y\|} [f(y'') - f(y)].$$

Como $|f(y'')| \leq c$ e $|f(y)| \leq c$, temos que

$$\frac{\|y' - y\|}{\delta + \|y' - y\|} [f(y'') - f(y)] \leq \frac{\|y' - y\|}{\delta} 2c.$$

Assim,

$$f(y') - f(y) \leq \frac{\|y' - y\|}{\delta} 2c.$$

Como y e y' são quaisquer obtemos a desigualdade análoga trocando y por y'. Portanto,

$$|f(y') - f(y)| \leq \frac{2c}{\delta} \|y' - y\|.$$

Como $x \in \operatorname{int}(\operatorname{dom} f)$ é qualquer, f é localmente lipschitz em $\operatorname{int}(\operatorname{dom} f)$. $\square$

Proposição 1.31 *Sejam $T \subset \mathbb{R}^l$ e $\psi : \mathbb{R}^n \times T \to \mathbb{R} \cup \{+\infty\}$. Defina $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ como*

$$g(x) = \sup_{\tau \in T} \psi(x, \tau) = \sup_{\tau \in T} \psi_\tau(x).$$

Se $\psi_\tau : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ é fortemente convexa com módulo $\mu \geq 0$, para todo $\tau \in T$ (respectivamente, convexa) então g também é fortemente convexa com módulo $\mu \geq 0$ (respectivamente, convexa).

Prova. Seja $x, y \in \mathbb{R}^n$, pela definição,

$$\psi_\tau(\alpha x + (1-\alpha)y) \leq \alpha\psi_\tau(x) + (1-\alpha)\psi_\tau(y) + \mu\alpha(1-\alpha)\|x-y\|^2, \qquad \forall \tau \in T.$$

Assim, pelas propriedades de supremo e definição de g temos

$$g(\alpha x + (1-\alpha)y) \leq \alpha g(x) + (1-\alpha)g(y) + \mu\alpha(1-\alpha)\|x-y\|^2.$$

Portanto g é fortemente convexa com módulo μ. O caso considerando apenas convexidade é imediato, $\mu = 0$. □

Definição 1.32 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Será chamado de conjunto de nível da função f associado a $d \in \mathbb{R}$ o conjunto $L_f(d)$ definido por*

$$L_f(d) = \{x \in \mathbb{R}^n : f(x) \leq d\}.$$

Proposição 1.33 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Então*

i) Se f é convexa, $L_f(d)$ é um conjunto convexo qualquer que seja $d \in \mathbb{R}$;

ii) Se f é semicontínua inferiormente, $L_f(d)$ é fechado qualquer que seja $d \in \mathbb{R}$.

Prova. (i)Dado $d \in \mathbb{R}$, tome $x, y \in L_f(d)$ e $\alpha \in [0,1]$. Como f é convexa temos que $\alpha x + (1-\alpha)y \in \operatorname{dom} f$ e

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) \leq \alpha d + (1-\alpha)d = d,$$

então $\alpha x + (1-\alpha)y \in L_f(d)$. Portanto $L_f(d)$ é convexo qualquer que seja $d \in \mathbb{R}$.
(ii)Dado $d \in \mathbb{R}$, seja $\{y^k\} \subset L_f(d)$ uma sequência que converge para $\bar{y} \in \mathbb{R}^n$ então

$$f(\bar{y}) \leq \liminf_{k\to\infty} f(y^k) \leq d.$$

Portanto $\bar{y} \in L_f(d)$, isto é, $L_f(d)$ é fechado qualquer que seja $d \in \mathbb{R}$. □

Proposição 1.34 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função convexa semicontínua inferiormente. Se existe $d \in \mathbb{R}$ tal que $L_f(d)$ é limitado, então todo conjunto de nível de f é limitado.*

Prova. Para demonstrarmos esta proposição consideramos o problema equivalente. Se existe $d \in \mathbb{R}$ tal que $L_f(d)$ é ilimitado provaremos que para qualquer $c \in \mathbb{R}$, tal que $L_f(c) \neq \emptyset$, o conjunto $L_f(c)$ também é ilimitado. Assim sendo, se $c \geq d$ temos que $L_f(c) \supset L_f(d)$

é ilimitado. Então considere $c < d$. Podemos observar que, pela Proposição 1.23, $L_f(d)$ possui uma direção de recessão $v \in \mathbb{R}^n$. Agora como $L_f(c) \subset L_f(d)$ a semi reta

$$x + tv, \qquad x \in L_f(c), \quad t > 0,$$

está toda contida em $L_f(d)$ e pela convexidade de f temos que

$$f(x+tv) = f\left(\left(1 - \frac{1}{\alpha}\right)x + \frac{1}{\alpha}(x + \alpha tv)\right) \leq \left(1 - \frac{1}{\alpha}\right)f(x) + \frac{1}{\alpha}f(x + \alpha tv),$$

qualquer que seja $\alpha \geq 1$. Fazendo α tender para infinito, temos

$$f(x+tv) \leq f(x) \leq c \qquad \forall t > 0.$$

Portanto $x + tv \in L_f(c)$ para todo $t > 0$, o que prova que $L_f(c)$ é ilimitado. □

Definição 1.35 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função convexa e $d \in \mathbb{R}^n$. A derivada direcional de f é a função $f'(\cdot, d) : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, +\infty\}$ que é definida por*

$$f'(x,d) = \begin{cases} \lim\limits_{t \downarrow 0} \dfrac{f(x+td) - f(x)}{t} & x \in \operatorname{dom} f, \\ -\infty & x \notin \operatorname{dom} f. \end{cases}$$

Dizemos que $f'(x,d)$ é a derivada direcional de f no ponto x na direção de d.

Observação 1.36 *Note que, se $x \in (\partial \operatorname{dom} f) \cap \operatorname{dom} f$ então existe $d \in \mathbb{R}^n$ em que $x + td \notin \operatorname{dom} f$ para todo $t > 0$. Neste caso, $f'(x,d) = +\infty$.*

Teorema 1.37 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função convexa. Então, para $x \in \operatorname{dom} f$ e $d \in \mathbb{R}^n$, temos que a derivada direcional de f em x na direção d satisfaz*

$$f'(x,d) = \inf_{t>0} \frac{f(x+td) - f(x)}{t}. \tag{1-5}$$

Se ainda, $x \in \operatorname{int}(\operatorname{dom} f)$ a derivada existe, isto é, $f'(x,d)$ é finita.

Prova. Inicialmente seja $\Phi : \mathbb{R} \setminus \{0\} \to \mathbb{R} \cup \{+\infty\}$ definida por

$$\Phi(t) = \frac{f(x+td) - f(x)}{t}.$$

Se $x \in (\partial \operatorname{dom} f) \cap \operatorname{dom} f$ e d for uma direção para a qual $f(x+td) = +\infty$, qualquer que seja $t > 0$, teremos que $\Phi(t) = +\infty$, para $t > 0$, assim 1-5 fica estabelecida. Agora se d não for uma destas direções ou $x \in \operatorname{int}(\operatorname{dom} f)$ então mostraremos que Φ é crescente para

$t > 0$ e limitada inferiormente no caso $x \in \operatorname{int}(\operatorname{dom} f)$. Para isso tome t_1, t_2, $\varepsilon \in \mathbb{R}$ tais que $0 < t_1 < t_2 < \varepsilon$ e $x + td \in \operatorname{dom} f$ qualquer que seja $t \in (0, \varepsilon)$. Então temos que

$$\Phi(t_2) - \Phi(t_1) = \frac{1}{t_1 t_2}\left(t_1 f(x + t_2 d) - t_2 f(x + t_1 d) + (t_2 - t_1) f(x)\right),$$

distribuindo $(1/t_2)$ e usando a convexidade de f teremos

$$\Phi(t_2) - \Phi(t_1) = \frac{1}{t_1}\left(\frac{t_1}{t_2} f(x + t_2 d) + (1 - \frac{t_1}{t_2}) f(x) - f\left(\frac{t_1}{t_2}(x + t_2 d) + (1 - \frac{t_1}{t_2}) x\right)\right) \geq 0.$$

Deste modo, da última expressão e da definida Φ, Φ é crescente para $t > 0$. Então

$$f'(x, d) = \lim_{t \downarrow 0} \Phi(t) = \inf_{t > 0} \Phi(t) = \inf_{t > 0} \frac{f(x + td) - f(x)}{t}.$$

Se $x \in \operatorname{int}(\operatorname{dom} f)$ escolha $\varepsilon > 0$ tal que $x + td \in \operatorname{dom} f$ qualquer que seja $t \in (-\varepsilon, \varepsilon)$, assim usando novamente a convexidade de f

$$\Phi(t) - \Phi\left(\frac{-\varepsilon}{2}\right) = \frac{\frac{1}{2} f(x + td) + \frac{1}{2} f(x) + \frac{t}{\varepsilon} f\left(x - \frac{\varepsilon}{2} d\right) + \left(1 - \frac{t}{\varepsilon}\right) f(x) - 2 f(x)}{t/2},$$

$$\Phi(t) - \Phi\left(\frac{-\varepsilon}{2}\right) \geq \frac{\frac{1}{2} f\left(x + \frac{t}{2} d\right) + \frac{1}{2} f\left(x - \frac{t}{2} d\right) - f(x)}{t/4} \geq \frac{f(x) - f(x)}{t/4} = 0.$$

Portanto $f'(x, d) > -\infty$. É bom ressaltar que ε depende de x e de d. $\square$

Teorema 1.38 *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *convexa e* $x \in \operatorname{int}(\operatorname{dom} f)$. *Então* $f'(x, \cdot)$ *é subaditiva, positivamente homogêneo e*

$$|f'(x, d)| \leq K \|d\|.$$

Definição 1.39 *(Subdiferencial) Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *uma função convexa. O subdiferencial de* f *em* $x \in \mathbb{R}^n$ *é o conjunto* $\partial f(x)$ *definido por*

$$\partial f(x) := \begin{cases} \{s \in \mathbb{R}^n : \langle s, d \rangle \leq f'(x, d) \;\; \forall d \in \mathbb{R}^n\}, & x \in \operatorname{dom} f \\ \emptyset. & x \notin \operatorname{dom} f \end{cases}$$

Teorema 1.40 *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *uma função convexa e* $x \in \operatorname{int}(\operatorname{dom} f)$. *Então*

i) $\partial f(x)$ *é não vazio, convexo e compacto,*

ii) $f'(x, d) = \max\{\langle s, d \rangle : s \in \partial f(x)\}$.

Os Teoramas 1.38 e 1.40 não serão demonstrados agora suas provas serão feitas em um contexto mais amplo no Capítulo 3 (Teoremas 3.2 e 3.5 combinados com o Teorema 3.9).

Teorema 1.41 *Sejam* $f,g:\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ *funções convexas e* $x \in \operatorname{dom} f \cap \operatorname{dom} g$ *tais que* $\partial f(x)$ *e* $\partial g(x)$ *sejam não vazios. Então*

$$\partial(f+g)(x) = \partial f(x) + \partial g(x).$$

Prova. Tome $s \in \partial f(x)$ e $w \in \partial g(x)$, então pela definição temos

$$\langle s,d\rangle \leq f'(x,d), \qquad \langle w,d\rangle \leq g'(x,d) \qquad \forall d \in \mathbb{R}^n.$$

Em consequência das propriedade de limite temos que

$$(f+g)'(x,d) = \lim_{t\downarrow 0} \frac{f(x+td)-f(x)+g(x+td)-g(x)}{t} = f'(x,d)+g'(x,d).$$

Então

$$\langle s+w,d\rangle \leq f'(x,d)+g'(x,d) = (f+g)'(x,d), \qquad \forall d \in \mathbb{R}^n,$$

portanto $s+w \in \partial(f+g)(x)$.
Suponhamos agora que exista $z \in \partial(f+g)(x) \setminus (\partial f(x)+\partial g(x))$. Pela compacidade de $\partial f(x)+\partial g(x)$ e pelo Teorema da Separação Estrita 1.21 existe $a \in \mathbb{R}^n$ tal que

$$\langle a,w+s\rangle < c \leq \langle a,z\rangle \leq (f+g)'(x,a) = f'(x,a)+g'(x,a) \qquad \forall w \in \partial f(x), \quad \forall s \in \partial g(x),$$

que é um absurdo pois, pelo Teorema 1.40, exite um $w_0+s_0 \in (\partial f(x)+\partial g(x))$ tal que

$$\langle a,s_0\rangle = f'(x,a), \qquad \langle a,w_0\rangle = g'(x,a).$$

Portanto $\partial(f+g)(x) = \partial f(x)+\partial g(x)$. $\square$

Proposição 1.42 *Sejam* $f:\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ *e* $x \in \operatorname{dom} f$.

i) se f *é convexa, então* $s \in \partial f(x)$ *se, e somente se,*

$$f(y) \geq f(x) + \langle s, y-x\rangle, \qquad \forall\, y \in \mathbb{R}^n; \tag{1-6}$$

ii) se f *é fortemente convexa com módulo* μ*, então* $s \in \partial f(x)$ *se, e somente se,*

$$f(y) \geq f(x) + \langle s, y-x\rangle + \mu\|x-y\|^2, \qquad \forall\, y \in \mathbb{R}^n. \tag{1-7}$$

Prova. (i)Primeiro suponhamos que $s \in \partial f(x)$ então, pela definição de subdiferencial e pelo Teorema 1.37, temos que

$$\langle s, y-x\rangle \leq \frac{f(x+t(y-x))-f(x)}{t}, \qquad \forall\, y \in \mathbb{R}^n, \quad \forall t \in (0,1].$$

Assim tomando $t = 1$ obtemos o que desejávamos. Suponhamos agora que, para $s \in \mathbb{R}^n$, (1-6) seja válida. Então, para todo $d \in \mathbb{R}^n$ e $t \in \mathbb{R}+$.

$$f(x+td)-f(x) \geq \langle s, td\rangle\,.$$

Portanto

$$f'(x,d) = \lim_{t\downarrow 0} \frac{f(x+td)-f(x)}{t} \geq \lim_{t\downarrow 0} \langle s, d\rangle = \langle s, d\rangle\,.$$

que implica que $s \in \partial f(x)$.

(ii) Suponhamos que $s \in \partial f(x)$, assim dividindo (1-4), para $x \neq y$ e $\alpha \in (0,1)$, por $\alpha(1-\alpha)$ temos

$$\frac{f(x+\alpha(y-x))-f(x)}{\alpha} \leq \frac{f(y)-f(x)}{1-\alpha} - \mu\|x-y\|^2,$$

fazendo α tender a zero pela direita temos que

$$f(y)-f(x)-\mu\|x-y\|^2 \geq f'(x,y-x) \geq \langle s, y-x\rangle\,, \qquad \forall\, y \in \mathbb{R}^n.$$

Suponhamos agora que para um $s \in \mathbb{R}^n$ seja válida (1-7). Então, para todo $d \in \mathbb{R}^n$ e $t \in \mathbb{R}+$

$$f(x+td)-f(x) \geq \langle s, td\rangle + t\mu\|d\|^2,$$

dividindo a expressão por t e fazendo t tender a 0 obtemos

$$f'(x,d) \geq \langle s, d\rangle + \mu\|d\|^2 \geq \langle s, d\rangle\,, \qquad \forall d \in \mathbb{R}^n.$$

Portanto $s \in \partial f(x)$. □

Proposição 1.43 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convexa e diferenciável em $x \in \operatorname{int}(\operatorname{dom} f)$. Então $\partial f(x) = \{\nabla f(x)\}$ qualquer que seja $x \in \operatorname{int}(\operatorname{dom} f)$.*

Prova. Como f é diferenciável em $x \in \operatorname{int}(\operatorname{dom} f)$ temos que

$$f'(x,d) = \lim_{t\downarrow 0} \frac{f(x+td)-f(x)}{t} = \lim_{t\to 0} \frac{f(x+td)-f(x)}{t} = \langle \nabla f(x), d\rangle, \qquad \forall d \in \mathbb{R}^n.$$

Assim $\nabla f(x) \in \partial f(x)$. Tome agora $s \in \partial f(x)$. Então

$$\langle s, d \rangle \leq \langle \nabla f(x), d \rangle, \qquad \forall d \in \mathbb{R}^n.$$

Fazendo $d = s - \nabla f(x)$ obtemos que $\|s - \nabla f(x)\|^2 \leq 0$, isto é, $s = \nabla f(x)$. Portanto $\{\nabla f(x)\} = \partial f(x)$. □

Proposição 1.44 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função de classe $\mathbb{C}^1$ tal que* $\mathrm{dom}\, f$ *seja convexa. Então f é fortemente convexa com módulo $\mu > 0$ se, e somente se,*

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq \mu \|y - x\|^2, \qquad \forall\, x, y \in \mathrm{dom}\, f. \tag{1-8}$$

Prova. Suponhamos f fortemente convexa, então para quaisquer $x, y \in \mathrm{dom}\, f$ temos

$$f(x + \alpha(y - x)) - f(x) \leq \alpha f(y) - \alpha f(x) - \mu\alpha(1 - \alpha)\|x - y\|^2, \qquad \forall \alpha \in [0, 1].$$

Tomando $\alpha \in (0, 1)$, dividindo a expressão por $\alpha(1 - \alpha)$ e fazendo α tender a zero temos

$$\mu\|x - y\|^2 \leq \langle \nabla f(x), x - y \rangle + f(y) - f(x),$$

e permutando x e y se obtém

$$\mu\|x - y\|^2 \leq \langle \nabla f(y), y - x \rangle + f(x) - f(y).$$

Portanto somando as duas ultimas expressão chegamos a (1-8).
Suponhamos agora que (1-8) seja válida. Dado $x, y \in \mathrm{dom}\, f$, $x \neq y$, faça $w = \alpha y + (1 - \alpha)x \in \mathrm{dom}\, f$, para $\alpha \in (1, 0)$. Pelo Teorema do Valor Médio existem $\xi \in (x, w)$ e $\eta \in (w, y)$ tais que

$$f(w) = f(x) + \langle \nabla f(\xi), w - x \rangle, \qquad f(w) = f(y) + \langle \nabla f(\eta), w - y \rangle.$$

Agora multiplicando a primeira igualdade por $(1 - \alpha)$, a segunda por α e somando-as obtemos

$$f(w) = \alpha f(y) + (1 - \alpha) f(x) + (1 - \alpha) \langle \nabla f(\xi), w - x \rangle + \alpha \langle \nabla f(\eta), w - y \rangle,$$

consequentemente

$$f(w) = \alpha f(y) + (1 - \alpha) f(x) + \alpha(1 - \alpha) \left(\langle \nabla f(\xi), y - x \rangle + \langle \nabla f(\eta), x - y \rangle \right).$$

Dados que $w \in (x,y)$, $\xi \in (x,w)$ e $\eta \in (w,y)$ então existe $\gamma > 0$ tal que $y - x = \gamma(\eta - \xi)$. Logo,

$$f(w) = \alpha f(y) + (1-\alpha)f(x) + \gamma\alpha(1-\alpha)\langle \nabla f(\xi) - \nabla f(\eta), \eta - \xi \rangle,$$

e por (1-8) temos

$$f(w) \leq \alpha f(y) + (1-\alpha)f(x) + \mu\alpha(1-\alpha)\|\gamma(\eta-\xi)\|^2,$$

portanto

$$f(\alpha y + (1-\alpha)x) = \alpha f(y) + (1-\alpha)f(x) + \mu\alpha(1-\alpha)\|y-x\|^2.$$

O que conclui a prova. □

A próxima proposição garante a existência de um único mínimo global para funções semicontínuas inferiormente e fortemente convexas que possuem $\mathrm{int}(\mathrm{dom}\, f) \neq \emptyset$.

Proposição 1.45 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função semicontínua inferiormente e fortemente convexa com módulo $\mu > 0$. Se $\mathrm{int}(\mathrm{dom}\, f) \neq \emptyset$, então f possui um único mínimo global.*

Para demonstrar esta proposição faremos uso do seguinte lema.

Lema 1.46 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função fortemente convexa com módulo $\mu > 0$. Se $\mathrm{int}(\mathrm{dom}\, f) \neq \emptyset$, então $L_f(c)$ é limitado, $\forall c \in \mathbb{R}$.*

Prova. Tome $x \in \mathrm{int}(\mathrm{dom}\, f)$ e $s \in \partial f(x)$, que é possível pois pelo Teorema 1.40 $\partial f(x) \neq \emptyset$. Suponha agora, por absurdo, que $L_f(c)$ é ilimitado, isto é, existe $\{y^k\} \subset L_f(c) \setminus \{x\}$ tal que $\lim_{k\to+\infty} \|y^k - x\| = +\infty$. Pela Proposição 1.42 teremos

$$\frac{f(y^k)}{\|y^k - x\|} \geq \frac{f(x)}{\|y^k - x\|} + \left\langle s, \frac{y^k - x}{\|y^k - x\|} \right\rangle + \mu\|y^k - x\|, \qquad \forall\, k.$$

o que implica

$$\limsup_{k\to\infty} \frac{f(y^k)}{\|y^k - x\|} = +\infty.$$

Que é um absurdo, pois $\{y^k\} \subset L_f(c) \setminus \{x\}$. Logo $L_f(c)$ é limitado. □

Prova.(*Proposição 1.45*) Primeiro observe que, pela Proposição 1.33, $L_f(c)$ é fechado, para $c \in \mathbb{R}$ e, pelo Lema 1.46, $L_f(c)$ é limitado, isto é, $L_f(c)$ é compacto. Então, tomando $c \in \mathbb{R}$ tal que $L_f(c) \neq \emptyset$, f possui um mínimo em $L_f(c)$ que é um mínimo global de f, pois $L_f(c)$ é um conjunto de nível e f é semicontínua inferiormente. Como f é estritamente convexa, possui um o único mínimo global, pois se existir x^*, $y^* \in U^* = \{u \in \mathbb{R}^n : f(u) \leq$

$f(x), \forall x \in \mathbb{R}^n\}$ tais que $x^* \neq y^*$, teremos que

$$f\left(\frac{1}{2}x^* + \frac{1}{2}y^*\right) < \frac{1}{2}f(x^*) + \frac{1}{2}f(y^*) = f(x^*),$$

que é um absurdo. E assim fica provada a proposição. □

Teorema 1.47 *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *uma função convexa. Então* $x \in \operatorname{dom} f$ *é um minimizador de* f *se e somente se* $0 \in \partial f(x)$.

Prova. Suponhamos que x seja um minimizador de f, isso implica que

$$f(x) \leq f(y) = f(y) - \langle 0, y - x \rangle, \qquad \forall y \in \mathbb{R}^n.$$

Portanto, pela Proposição 1.42, $0 \in \partial f(x)$. Reciprocamente, suponhamos que $0 \in \partial f(x)$, pela Proposição 1.42, temos

$$f(x) \leq f(y) - \langle 0, y - x \rangle = f(y), \qquad \forall y \in \mathbb{R}^n.$$

Portanto x é um minimizador de f. □

Teorema 1.48 *Suponhamos que* $T \subset \mathbb{R}^l$ *seja um conjunto compacto,* $\Omega \subset \mathbb{R}^n$ *um conjunto aberto, convexo e* $\psi : \mathbb{R}^n \times T \to \mathbb{R} \cup \{+\infty\}$ *uma função contínua em* $\Omega \times T$. *Dado* $\tau \in T$, *defina*

$$\begin{array}{rcl} \psi_\tau : \mathbb{R}^n & \to & \mathbb{R} \cup \{+\infty\} \\ x & \mapsto & \psi_\tau(x) = \psi(x, \tau). \end{array}$$

Sejam $\tilde{\psi}_\tau = \psi_{\tau|_\Omega}$ *a restrição de* ψ_τ *ao conjunto* Ω *e* $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *definida por*

$$g(x) = \sup_{\tau \in T} \psi_\tau(x).$$

Se ψ_τ *é diferenciável em* Ω *e* $\tilde{\psi}_\tau$ *é convexa, para todo* $\tau \in T$. *Então*

(a) $\Omega \subset \operatorname{dom} g$, *a função* $\tilde{g} = g_{|\Omega}$ *é convexa e tem-se*

$$g'(x, d) = \max_{\bar{\tau} \in T(x)} \psi'_{\bar{\tau}}(x, d), \qquad \forall\, x \in \Omega, \quad \forall\, d \in \mathbb{R}^n, \tag{1-9}$$

onde $T(x) = \{\bar{\tau} \in T : g(x) = \psi(x, \bar{\tau}) = \max_{\tau \in T} \psi(x, \tau)\}$.

(b) se $\nabla_x \psi$ *é contínua em* $\Omega \times T$,

$$\partial g(x) = \operatorname{conv}\{\nabla_x \psi(x, \tau) : \bar{\tau} \in T(x)\}, \quad x \in \Omega.$$

Prova. Como T é compacto e ψ é contínua em $\Omega \times T$ o Teorema de Weierstrass nos dá que $T(x) \neq \emptyset$, para todo $x \in \Omega$, e consequentemente $g(x) < +\infty$. Logo $\Omega \subset \operatorname{dom} g$. Já a convexidade de $\tilde{g}$ é dada pela Proposição 1.31, pois $\tilde{g}(x) = sup_{\tau \in T} \tilde{\psi}_\tau(x)$. Agora tome $x \in \Omega$. Assim para todo $\bar{\tau} \in T(x)$ temos $\tilde{g}(x) = \tilde{\psi}_{\bar{\tau}}(x)$ e $\tilde{g}(x+td) \geq \tilde{\psi}_{\bar{\tau}}(x+td)$. Portanto,

$$\frac{\tilde{g}(x+td) - \tilde{g}(x)}{t} \geq \frac{\tilde{\psi}_{\bar{\tau}}(x+td) - \tilde{\psi}_{\bar{\tau}}(x)}{t},$$

qualquer que seja $d \in \mathbb{R}^n$ e $t > 0$. Visto que $\tilde{g}$ e $\tilde{\psi}$ são convexas, fazendo t tender a 0 pela direita na desigualdade acima obtemos

$$\tilde{g}'(x,d) \geq \tilde{\psi}'_{\bar{\tau}}(x,d) = \langle \nabla \tilde{\psi}_{\bar{\tau}}(x), d \rangle,$$

qualquer que seja $\bar{\tau} \in T(x)$ e $d \in \mathbb{R}^n$. Logo,

$$\tilde{g}'(x,d) \geq \sup_{\bar{\tau} \in T(x)} \tilde{\psi}'_{\bar{\tau}}(x,d) \quad \forall d \in \mathbb{R}^n. \tag{1-10}$$

Agora dados $d \in \mathbb{R}^n$ e $\{t_k\}$, tal que $\lim_{k\to\infty} t_k = 0$ e $t_k > 0$. Para todo k, definimos $x^k = x + t_k d$ e $\bar{\tau}^k \in T(x^k)$. Como $\{\bar{\tau}^k\} \subset T$ e T é compacto, podemos supor (tomando uma subsequência, se for necessário) que $\lim_{k\to\infty} \bar{\tau}^k = \bar{\tau} \in T$. Assim, temos que

$$\tilde{\psi}_{\bar{\tau}^k}(x^k) = \max_{\tau \in T} \tilde{\psi}_\tau(x^k) \geq \tilde{\psi}_\tau(x^k), \qquad \forall \tau \in T.$$

Passando ao limite quando $k \to \infty$, pela continuidade de $\tilde{\psi}$ em $\Omega \times T$, obtemos que

$$\tilde{\psi}_{\bar{\tau}}(x) \geq \tilde{\psi}_\tau(x), \qquad \forall \tau \in T.$$

Isto implica que $\bar{\tau} \in T(x)$. Assim,

$$\tilde{g}'(x,d) \leq \frac{\tilde{g}(x+t_k d) - \tilde{g}(x)}{t_k} = \frac{\tilde{\psi}_{\bar{\tau}^k}(x+t_k d) - \tilde{\psi}_{\bar{\tau}}(x)}{t_k}$$

trocando $\tilde{\psi}_{\bar{\tau}}(x)$ por $\tilde{\psi}_{\bar{\tau}^k}(x)$, dado que $\bar{\tau} \in T(x)$, obtemos

$$\tilde{g}'(x,d) \leq \frac{\tilde{\psi}_{\bar{\tau}^k}(x+t_k d) - \tilde{\psi}_{\bar{\tau}^k}(x)}{t_k}.$$

Passando ao limite quando $k \to \infty$, obtemos

$$\tilde{g}'(x,d) \leq \tilde{\psi}'_{\bar{\tau}}(x,d),$$

onde $\bar{\tau} \in T(x)$. Combinando essa expressão com (1-10) e com o fato de $x \in \Omega$, Ω aberto, obtemos a igualdade (1-9), o que prova a afirmação a).

Provaremos agora o item b). Para todo $u \in \mathbb{R}^n$ e todo $\bar{\tau} \in T(x)$, temos que

$$\tilde{g}(u) = \max_{\tau \in T} \tilde{\psi}_\tau(u) \geq \tilde{\psi}_{\bar{\tau}}(u),$$

como $\tilde{g}$ e $\tilde{\psi}_{\bar{\tau}}$ são convexas, usando Proposição 1.42, obtemos

$$\tilde{g}(u) \geq \tilde{\psi}_{\bar{\tau}}(u) \geq \tilde{\psi}_{\bar{\tau}}(x) + \langle \nabla \tilde{\psi}_{\bar{\tau}}(x), u - x \rangle = \tilde{g}(x) + \langle \nabla \tilde{\psi}_{\bar{\tau}}(x), u - x \rangle .$$

A relação acima vale para todo $u \in \mathbb{R}^n$, então

$$\nabla_x \tilde{\psi}_{\bar{\tau}}(x) \in \partial \tilde{g}(x), \qquad \forall \bar{\tau} \in T(x).$$

Como $\partial g(x)$ é um conjunto convexo (Teorema 1.40), temos então que

$$\operatorname{conv}\{\nabla_x \tilde{\psi}_{\bar{\tau}}(x) : \bar{\tau} \in T(x)\} \subset \operatorname{conv} \partial \tilde{g}(x) = \partial \tilde{g}(x).$$

uma observação interessante é que $T(x) = T \cap \{\tau \in T : \tilde{\psi}_\tau(x) = \tilde{g}(x)\}$, onde T é compacto e $\{\tau : \tilde{\psi}_\tau(x) = \tilde{g}(x)\}$ é fechado (pela continuidade de $\tilde{\psi}$ em $\Omega \times T$) para todo $x \in \Omega$ fixo. Portanto $T(x)$ é compacto para todo $x \in \Omega$. Pela continuidade de $\nabla_x \tilde{\psi}$ em $\Omega \times T$, temos então que o conjunto $\{\nabla_x \tilde{\psi}_{\bar{\tau}}(x) : \bar{\tau} \in T(x)\}$ é compacto $\forall x \in \Omega$. Assim a Proposição 1.16 nos dá que $\operatorname{conv}\{\nabla_x \tilde{\psi}_{\bar{\tau}}(x) : \bar{\tau} \in T(x)\}$ é compacto.

Agora vamos mostrar que a inclusão não pode ser própria, para isso suponhamos que seja própria, isto é,

$$\exists y \in \partial \tilde{g}(x) \setminus \operatorname{conv}\{\nabla_x \tilde{\psi}_{\bar{\tau}}(x) : \bar{\tau} \in T(x)\}.$$

Pelos Teorema da Separação Estrita, Teorema 1.21, existem $a \in \mathbb{R}^n \setminus \{0\}$ e $c \in \mathbb{R}$ tais que

$$\langle a, y \rangle > c > \langle a, \nabla \tilde{\psi}_{\bar{\tau}}(x) \rangle ,$$

qualquer que seja $\bar{\tau} \in T(x)$. O que implica que

$$\langle a, y \rangle > c \geq \max_{\bar{\tau} \in T(x)} \langle a, \nabla \tilde{\psi}_{\bar{\tau}}(x) \rangle = \tilde{g}'(x, a),$$

onde a igualdade segue do item a) e da diferenciabilidade de $\tilde{\psi}$ em Ω. O que contradiz a definição, pois $y \in \partial \tilde{g}(x)$. Portanto,

$$\partial \tilde{g}(x) = \operatorname{conv}\{\nabla_x \tilde{\psi}_{\bar{\tau}}(x) : \bar{\tau} \in T(x)\}, \qquad \forall x \in \Omega,$$

e assim o item b) segue da expressão acima e do fato de $x \in \Omega$, Ω aberto. $\square$

CAPÍTULO 2

Método de Ponto Proximal Global

Neste capítulo aplicaremos o Método de Ponto Proximal para minimizar funções convexas.

2.1 Método de Ponto Proximal

Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função convexa e semicontínua inferiormente. O método de *ponto proximal* para resolver o problema de otimização

$$\min_{x \in \mathbb{R}^n} f(x), \tag{2-1}$$

é definido formalmente do seguinte modo: Dados a sequência $\{\lambda_k\}$ de números reais positivos e $x_0 \in \mathrm{dom}(f)$ defina

$$x^{k+1} = \arg\min_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{\lambda_k}{2}\|x - x^k\|^2 \right\}, \qquad k = 0, 1, \ldots. \tag{2-2}$$

Vamos supor que $\mathrm{int}(\mathrm{dom}\, f) \neq \emptyset$ e f é limitada inferiormente.

2.2 Análise de Convergência

Mostraremos que o método de ponto proximal está bem definido e converge para uma solução do problema (2-1) sempre que existir solução. Inicialmente definamos

$$f^* = \inf\{f(x) : x \in \mathbb{R}^n\}.$$

Lema 2.1 *A sequência $\{x^k\}$ gerada pelo método de ponto proximal (2-2), com ponto inicial $x^0 \in \mathrm{dom}\, f$, está bem definida. Além disso,*

i) $f(x^{k+1}) + (\lambda_k/2)\|x^{k+1} - x^k\| \leq f(x^k), \quad k = 0, 1, \ldots;$

ii) $\{x^k\} \subset L_f(f(x_0))$;

iii) $0 \in \partial f(x^{k+1}) + (\lambda_k)(x^{k+1} - x^k), \quad k = 0, 1, \ldots$;

iv) *a sequência* $\{f(x^k)\}$ *converge;*

v) $\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$.

Prova. Como $x^0 \in \operatorname{dom} f$ e f é semicontínua inferiormente, a boa definição da seqüência $\{x_k\}$ segue da Proposição 1.45. Os itens *i* e *iii* são imediatos. Basta aplicar os Teoremas 1.47 e 1.41 observando que $x^{k+1} \in \operatorname{dom} f$ minimiza a função

$$\mathbb{R}^n \ni x \mapsto f(x) + \frac{\lambda_k}{2}\|x - x^k\|^2.$$

Do item *i*, segue imediatamente que $f(x^k) \leq f(x^{k-1}) \leq \cdots \leq f(x^0)$, isto é, $\{f(x^k)\}$ é monótona não-crescente. Logo o item *ii* segue-se. Agora, como $\{f(x^k)\}$ é monótona não-crescente e f é limitada inferiormente $\{f(x^k)\}$ converge e o item *iv* fica provado. Finalmente o item *v* segue combinando os itens *i* e *iv*. □

Teorema 2.2 *Seja* $\{x^k\}$ *a sequência gerada pelo método de ponto proximal* (2-2)*, com ponto inicial* $x^0 \in \operatorname{dom} f$*, então*

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

Prova. Pelo Lema 2.1 item *iii* temos que

$$0 \in \partial f(x^{k+1}) + (\lambda_k)(x^{k+1} - x^k), \quad k = 0, 1, \ldots,$$

isto é, $\lambda_k(x^k - x^{k+1}) \in \partial f(x^{k+1})$. Como f é convexa, usando a Proposição 1.42 item i temos, após simples manipulações algébricas, que

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

Por outro lado,

$$\|y - x^k\|^2 = \|y - x^{k+1}\|^2 - 2\langle x^k - x^{k+1}, y - x^{k+1}\rangle + \|x^{k+1} - x^k\|^2.$$

Então, combinando as duas últimas expressões obtemos:

$$\|y - x^{k+1}\|^2 \leq \|y - x^k\|^2 - \|x^{k+1} - x^k\|^2 + \frac{1}{\lambda_k}\left(f(y) - f(x^{k+1})\right).$$

Como $\|x^{k+1} - x^k\|^2$ é maior ou igual a zero o resultado segue. $\square$

Definição 2.3 *Uma sequência $\{z^k\}$ é dita Fejér convergente ao conjunto $U \subset \mathbb{R}^n$ quando, para todo $z \in U$*

$$\|z - z^{k+1}\| \leq \|z - z^k\| \qquad k = 0, 1, \ldots. \tag{2-3}$$

Lema 2.4 *Seja $\{z^k\}$ é uma sequência Fejér convergente a $U \subset \mathbb{R}^n$, conjunto não vazio. Então $\{z^k\}$ é limitada. Se, além disso, U possui um ponto de acumulação da sequência $\{z^k\}$, então a sequência converge.*

Prova. Tome $z \in U$ fixo. Usando a desigualdade triangular em (2-3) temos

$$\|z^{k+1}\| \leq \|z - z^0\| + \|z\| \qquad k = 0, 1, \ldots.$$

Assim, $\{z^k\}$ é limitada. Seja $\bar{z} \in U$ um ponto de acumulação de $\{z^k\}$. Assim, pela definição de Fejér convergência temos que a sequência numérica $\{\|\bar{z} - z^k\|\}$ é monótona não crescente e possui zero como ponto de acumulação, logo converge para zero. Portanto $\{z^k\}$ converge para $\bar{z}$. $\square$

Agora vamos supor que o problema em (2-1) tem solução, isto é,

$$U^* = \{z \in \mathbb{R}^n : f(z) \leq f(y), \quad \forall y \in \mathbb{R}^n\} \neq \emptyset.$$

Teorema 2.5 *Seja $\{x^k\}$ a sequência gerada pelo método de ponto proximal (2-2) aplicado a f com ponto inicial $x^0 \in \operatorname{dom} f$ e $\lambda_k > 0$, tal que $\sum_{k=0}^{\infty} 1/\lambda_k = \infty$. Então $\{x^k\}$ converge para uma solução do problema (2-1).*

Prova. Primeiro mostraremos que $f(x^k)$ converge para f^*. Para mostrar isso suponha, por contradição, que $f(x^k)$ não convirja para f^*. Então existe um $\delta > 0$ e k_0 tal que

$$f(x^k) > f(z) + \delta, \qquad \forall z \in U^*, \quad k \geq k_0.$$

Por outro lado, pelo Teorema 2.2 temos que

$$\|z - x^{k+1}\|^2 < \|z - x^k\|^2 - \frac{\delta}{\lambda_k}, \qquad \forall z \in U^*, \qquad k \geq k_0.$$

Assim somando em k teremos

$$\sum_{k=k_0}^{n} \frac{\delta}{\lambda_k} < \|z - x^{k_0}\|^2 - \|z - x^{n+1}\|^2 \leq \|z - x^{k_0}\|^2, \qquad \forall z \in U^*, \qquad n > k_0,$$

que contradiz a hipótese, logo $f(x^k)$ converge para f^*. Agora como $f(x^*) - f(x^{k+1}) \leq 0$ temos, pelo Teorema 2.2, que

$$\|x^* - x^{k+1}\|^2 \leq \|x^* - x^k\|^2, \qquad k = 0, 1, \ldots$$

Então $\{x^k\}$ é uma sequência Fejér convergente a U^*. Assim, pelo Lema 2.4, a sequência é limitada e consequentemente possui um ponto de acumulação $\bar{x} \in \mathbb{R}^n$. Mas na verdade $\bar{x} \in U^*$ pois f é semicontínua inferiormente, isto é,

$$f^* = \liminf_{k \to \infty} f(x^k) \geq f(\bar{x}).$$

Portanto, pelo Lema 2.4, $\{x^k\}$ converge para $\bar{x} \in U^*$. $\square$

CAPÍTULO 3

Teoria Diferencial Generalizada

Neste capítulo generalizamos os conceitos de diferenciação de funções convexas, definidos no Capítulo 1, para funções localmente Lipschitz. Neste capítulo os principais livros que foram utilizados como referência são [2] e [9].

3.1 Derivada direcional generalizada

Nesta seção vamos apresentar o conceito de derivada direcional generalizada segundo Clarke [2].

Definição 3.1 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ localmente Lipschitz e $d \in \mathbb{R}^n$. A derivada direcional generalizada $f^o(\cdot, d) : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, +\infty\}$ é definida por*

$$f^o(x;d) = \begin{cases} \limsup\limits_{\substack{y \to x \\ t \downarrow 0}} \dfrac{f(y+td) - f(y)}{t}, & x \in \operatorname{dom}(f), \\ -\infty & x \notin \operatorname{dom}(f). \end{cases}$$

Dizemos que $f^o(x;d)$ é a derivada direcional generalizada de f em x na direção de d. Se f for localmente Lipschitz em $x \in \mathbb{R}^n$ podemos calcular a derivada direcional de f em x numa direcão $d \in \mathbb{R}^n$ como

$$f^o(x;d) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y+td) - f(y)}{t}.$$

Teorema 3.2 *Sejam $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ localmente Lipschitz em $x \in \operatorname{dom} f$ e K_x a constante de Lipschitz. Então $f^0(x;\cdot)$ é positivamente homogêneo, subaditiva e*

$$|f^0(x;d)| \leq K_x \|d\|, \qquad \forall\, d \in \mathbb{R}^n.$$

Prova. Primeiro mostraremos que $|f^0(x;d)| \leq K_x \|d\|$. Usando algumas propriedades de lim sup temos

$$|f^0(x;d)| = \left| \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y+td)-f(y)}{t} \right| \leq \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{|f(y+td)-f(y)|}{t} \leq K_x \|d\|.$$

Agora mostraremos que $f^0(x;\cdot)$ é positivamente homogêneo. Dado $\lambda > 0$ temos

$$\begin{aligned} f^0(x;\lambda d) &= \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y+\lambda td)-f(y)}{t} = \limsup_{\substack{y \to x \\ t \downarrow 0}} \lambda \left\{ \frac{f(y+\lambda td)-f(y)}{\lambda t} \right\} \\ &= \lambda \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y+\lambda td)-f(y)}{\lambda t} = \lambda f^0(x;d). \end{aligned}$$

O caso $\lambda = 0$ é imediato. Para mostrar a subaditividade de $f^0(x;\cdot)$ tome $d, w \in \mathbb{R}^n$ então

$$\begin{aligned} f^0(x;d+w) &= \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y+t(d+w))-f(y)}{t} \\ &\leq \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y+tw+td)-f(y+tw)}{t} + \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y+tw)-f(y)}{t}, \end{aligned}$$

e assim $f^0(x;d+w) \leq f^0(x;d) + f^0(x;w)$, o que conclui a prova. □

Teorema 3.3 *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *localmente Lipschitz. Então* f^0 *é semicontínua superiormente em* $\operatorname{dom} f \times \mathbb{R}^n$.

Prova. Tome $\{x^i\} \in \operatorname{dom} f$ e $\{d^i\} \subset \mathbb{R}^n$ tais que $\lim_{i \to \infty} x^i = x \in \operatorname{dom} f$ e $\lim_{i \to \infty} d^i = d$. Assim, pela definição de limite superior, existem $\{y^i\} \subset \mathbb{R}^n$ e $\{t_i\} \subset \mathbb{R}_+ \setminus \{0\}$ tais que

$$f^o(x^i, d^i) \leq \frac{f(y^i + t_i d^i) - f(y^i)}{t_i} + \frac{1}{i}, \qquad \|y^i - x^i\| + t_i < \frac{1}{i}, \qquad \forall i \in \mathbb{N}, \tag{3-1}$$

onde foi usado que

$$f^o(x^i, d^i) = \limsup_{\substack{y \to x^i \\ t \downarrow 0}} \frac{f(y+td^i)-f(y)}{t}.$$

Assim, usando 3-1 e simples manipulações algébricas teremos que

$$f^o(x^i, d^i) - \frac{1}{i} \leq \frac{f(y^i + t_i d^i) - f(y^i)}{t_i} = \frac{f(y^i + t_i d) - f(y^i)}{t_i} + \frac{f(y^i + t_i d^i) - f(y^i + t_i d)}{t_i}, \tag{3-2}$$

para i suficientemente grande. Por outro lado, como f é localmente Lipschitz, para i suficientemente grande

$$\frac{|f(y^i+t_id^i)-f(y^i+t_id|)}{t_i} \leq \frac{K_x\|t_id^i-t_id\|}{t_i} = K_x\|d^i-d\| \overset{i\to\infty}{\to} 0.$$

Tomando o limite superior em (3-2) e utilizando a última expressão teremos

$$\limsup_{i\to\infty} f^o(x^i,d^i) \leq \limsup_{i\to\infty} \frac{f(y^i+t_id)-f(y^i)}{t_i} \leq f^o(x,d),$$

o que conclui a prova. □

3.2 Subdiferencial generalizado

Nesta seção estenderemos o conceito de subdiferencial. Esta extensão também é devida a Clarke [2].

Definição 3.4 *Seja $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ localmente Lipschitz. O subdiferencial generalizado de f em $x\in\mathbb{R}^n$ é o conjunto $\partial_c f(x)$ definido por*

$$\partial_c f(x) := \begin{cases} \{s\in\mathbb{R}^n : \langle s,d\rangle \leq f^o(x,d),\ \forall d\in\mathbb{R}^n\}, & x\in \operatorname{dom} f, \\ \emptyset, & x\notin \operatorname{dom} f. \end{cases}$$

Os elementos $s\in\partial_c f(x)$ são chamados de subgradiente de f em x. De maneira análoga à derivada direcional generalizada, se f for localmente Lipschitz em $x\in\mathbb{R}^n$ podemos calcular o subdiferencial generalizado de f em x como

$$\partial_c f(x) = \{s\in\mathbb{R}^n : \langle s,d\rangle \leq f^o(x,d)\ \ \forall d\in\mathbb{R}^n\}.$$

Vejamos algumas propriedades deste subdiferencial.

Teorema 3.5 *Seja $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ localmente Lipschitz em $x\in\operatorname{dom} f$ com constante K_x. Então*

i) $\partial_c f(x)$ é não vazio, convexo e compacto;

ii) $f^o(x,d)=\max\{\langle s,d\rangle : s\in\partial f(x)\}$.

Prova. (i) Façamos $f_x^o(d) := f^o(x,d)$ para todo $d\in\mathbb{R}^n$. Então dado $d_0\in\mathbb{R}^n$ exite $w\in\mathbb{R}^n$ tal que $f_x^o(d_0)=\langle w,d_0\rangle$. Assim, pelo Teorema 3.2, f_x^o é subaditiva e positivamente homogêneo. Assim, pela Proposição 1.2 item *iii*, temos

$$f_x^o(td_0) \geq t f_x^o(d_0) = \langle w, td_0 \rangle, \qquad \forall t \in \mathbb{R}.$$

Portanto, pelo Teorema de Hahn-Banach 1.4 combinado com o Teorema de Riesz 1.5, existe $\bar{s} \in \mathbb{R}^n$ tal que

$$f^o(x,d) = f_x^o(d) \geq \langle \bar{s}, d \rangle, \quad \forall d \in \mathbb{R}^n, \qquad\qquad \langle \bar{s}, td_0 \rangle = \langle w, td_0 \rangle, \quad \forall t \in \mathbb{R}.$$

Isto é, $\bar{s} \in \partial f(x)$.

Agora vamos mostrar a convexidade. Tome $s, w \in \partial f(x)$ e $\lambda \in [0,1]$. Assim, para todo $d \in \mathbb{R}^n$ temos que

$$\langle \lambda w + (1-\lambda)s, d \rangle = \lambda \langle w, d \rangle + (1-\lambda) \langle s, d \rangle \leq f^o(x,d),$$

implicando que $\lambda w + (1-\lambda)s \in \partial f(x)$, isto é, $\partial f(x)$ é convexo.

Finalmente vamos mostrar a compacidade. Tome $\xi \in \partial f(x)$ e observemos que

$$\|\xi\|^2 = \langle \xi, \xi \rangle \leq f^o(x, \xi) \leq K_x \|\xi\|.$$

Portanto $\partial f(x)$ é limitado. Basta agora provar que $\partial f(x)$ é fechado, para isso tome $\xi_i \in \partial f(x)$ uma sequência que converge para $\xi \in \mathbb{R}^n$. Deste modo para todo $d \in \mathbb{R}^n$ temos que

$$\langle \xi, d \rangle = \lim_{i \to \infty} \langle \xi_i, d \rangle \leq \lim_{i \to \infty} f^o(x,d) = f^o(x,d).$$

Portanto $\partial f(x)$ é fechado e isto conclui a prova do item i.

(ii)Pela definição de subdiferencial

$$f(x,d) \geq \langle s, d \rangle, \qquad \forall s \in \partial f(x), \quad \forall d \in \mathbb{R}^n.$$

Usando argumento análogo à prova que $\partial_c f(x)$ é não vazio, podemos mostrar que para cada $d \in \mathbb{R}^n$ existe $\bar{s} \in \partial f(x)$ tal que

$$f(x,d) = \langle \bar{s}, d \rangle.$$

Portanto o resultado segue combinando as duas últimas espre. □

Teorema 3.6 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função localmente Lipschitz. Então a aplicação $\partial_c f(\cdot) : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ é semi contínua superiormente, isto é, se $\{y^k\}$ é uma sequência tal que $\lim_{k \to +\infty} y^k = y$ e $\xi^k \in \partial_c f(y^k)$, então todo ponto de acumulação ξ de $\{\xi^k\}$*

pertence a $\partial_c f(y)$.

Prova. Seja $\{\xi^{k_j}\}$ uma subsequência tal que $\xi^{k_j} \in \partial_c f(y^{k_j})$ com $\lim_{j\to\infty} \xi^{k_j} = \xi$. Então

$$\left\langle \xi^{k_j}, d \right\rangle \leq f^o(y^{k_j}, d), \qquad \forall d \in \mathbb{R}^n.$$

Assim, pelo Teorema 3.3, temos que

$$\limsup_{j\to\infty} \left\langle \xi^{k_j}, d \right\rangle = \langle \xi, d \rangle \leq \limsup_{j\to\infty} f^o(y^{k_j}, d) \leq f^o(y, d), \qquad \forall d \in \mathbb{R}^n.$$

Portanto $\xi \in \partial f(y)$. □

Definição 3.7 *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *localmente Lipschitz em* $x \in \operatorname{dom} f$. *Diremos que* x *é um ponto estacionário de* f *quando* $0 \in \partial_c f(x)$.

Proposição 3.8 *Sejam* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *um função localmente Lipschitz em* $x \in \operatorname{dom} f$ *e* $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *uma função continuamente diferenciável neste mesmo ponto. Então*

$$\partial_c (f+g)(x) \subseteq \partial_c f(x) + \{\nabla g(x)\}.$$

Prova. Primeiro observamos que

$$(f+g)^o(x,d) = \limsup_{\substack{y\to x \\ t\downarrow 0}} \left[\frac{f(y+td)-f(y)}{t} + \frac{g(y+td)-g(y)}{t} \right].$$

Usando propriedades de limite superior, obtemos

$$(f+g)^o(x;d) \leq f^o(x;d) + \langle \nabla g(x), d \rangle,$$

assim dado $\xi \in \partial(f+g)(x)$ temos, pela definição de subdiferencial e pelo Teorema 1.43, que

$$\langle \xi, d \rangle \leq (f+g)^o(x;d) \leq f^0(x;d) + \langle \nabla g(x), d \rangle, \qquad \forall d \in \mathbb{R}^n.$$

Logo,

$$\langle \xi - \nabla g(x), d \rangle \leq f^0(x;d), \qquad \forall d \in \mathbb{R}^n,$$

o que implica que $\xi - \nabla g(x) \in \partial f(x)$. Portanto $\xi \in \partial f(x) + \{\nabla g(x)\}$. □

Agora mostraremos que se f é convexa a derivada direcional e a derivada direcional generalizada coincidem em qualquer ponto no interior do domínio de f. Consequentemente o subdiferencial e o subdiferencial generalizado também irão coincidir.

Teorema 3.9 *Sejam* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *convexa e* $x \in \operatorname{int}(\operatorname{dom} f)$. *Então*

$$f^o(x,d) = f'(x,d), \qquad \forall\, d \in \mathbb{R}^n.$$

Prova. Seja $x \in \operatorname{int}(\operatorname{dom} f)$. Inicialmente observemos que pelo Teorema 1.30 f é localmente Lipschitz em x, assim a derivada direcional generalizada existe. Mostremos a igualdade. Usando propriedades do $\limsup$ é imediato concluir que

$$f^o(x,d) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y+td) - f(y)}{t} \geq \limsup_{t \downarrow 0} \frac{f(x+td) - f(x)}{t} = f'(x,d). \tag{3-3}$$

Agora, dado $\delta > 0$, pela definição de $\limsup$,

$$f^o(x,d) = \lim_{\varepsilon \downarrow 0} \sup_{\|y-x\| < \varepsilon\delta} \sup_{0<t<\varepsilon} \frac{f(y+td) - f(y)}{t}.$$

Como f é convexa, pelo Teorema 1.37, temos

$$f^o(x,d) = \lim_{\varepsilon \downarrow 0} \sup_{\|y-x\| < \varepsilon\delta} \frac{f(y+\varepsilon d) - f(y)}{\varepsilon}.$$

Por outro lado, f é localmente Lipschitz em x, então para ε suficientemente pequeno

$$\left| \frac{f(y+\varepsilon d) - f(y)}{\varepsilon} - \frac{f(x+\varepsilon d) - f(x)}{\varepsilon} \right| \leq \frac{2K_x}{\varepsilon} \|y-x\| \leq 2K_x\delta, \qquad \|y-x\| < \varepsilon\delta.$$

Assim, combinado as duas últimas expressões é fácil concluir que

$$f^o(x,d) \leq \lim_{\varepsilon \downarrow 0} \frac{f(x+\varepsilon d) - f(x)}{\varepsilon} + 2K_x\delta = f'(x,d) + 2K_x\delta, \qquad \forall\, \delta > 0.$$

Logo $f^o(x,d) \leq f'(x,d)$, o que junto com (3-3) implica o resultado desejado. $\square$

CAPÍTULO 4

Método de Ponto Proximal Local

Neste capítulo aplicaremos o Método de Ponto Proximal para minimizar funções não-convexas, que serão divididas em dois casos. Enfatizamos que os resultados obtidos são locais, pois dependem fortemente da escolha adequada de um certo conjunto de nível da função objetivo.

4.1 A Função Supremo de Funções Diferenciáveis

Neste seção vamos estudar uma classe especial de funções dadas pelo supremo de funções diferenciáveis a qual será nosso objeto de estudo neste cápitulo.

Sejam $T \subset \mathbb{R}^m$, $\Omega \subset \mathbb{R}^n$ um conjunto aberto e convexo e $\varphi : \mathbb{R}^n \times T \to \mathbb{R} \cup \{+\infty\}$. Suponhamos que a função

$$\begin{array}{rcl} \varphi_\tau : \mathbb{R}^n & \to & \mathbb{R} \cup \{+\infty\} \\ x & \mapsto & \varphi_\tau(x) := \varphi(x, \tau), \end{array}$$

seja continuamente diferenciável em Ω para todo $\tau \in T$. Defina $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ como

$$f(x) = \sup_{\tau \in T} \varphi_\tau(x). \tag{4-1}$$

Além disso, suponhamos que as seguintes condições são satisfeitas:

H0) f é semicontínua inferiormente;

H1) $\sup_{\tau \in T} \|\nabla \varphi_\tau(\bar{x})\| := \sup_{\tau \in T} \|\nabla_x \varphi(\bar{x}, \tau)\| < +\infty$, para algum $\bar{x} \in \Omega \cap \operatorname{dom} f$;

H2) Para cada $\tau \in T$, o gradiente $\nabla \varphi_\tau := \nabla_x \varphi(\cdot, \tau)$ é Lipschitz em Ω com constantes $L_\tau > 0$.

H3) $L = sup_{\tau \in T} L_\tau < +\infty$.

De agora em diante f denotará a função definida em (4-1). Além disso, vamos assumir que as hipóteses **H0**, **H1**, **H2** e **H3** são sempre válidas.

Para simplificar a notação vamos definir a seguinte função: Dados $\bar{x} \in \mathbb{R}^n$ e $\lambda \geq 0$ a função $f_{(\lambda,\bar{x})} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ é definida por

$$f_{(\lambda,\bar{x})}(x) := f(x) + \frac{\lambda}{2}\|x - \bar{x}\|^2. \tag{4-2}$$

Proposição 4.1 *Sejam $\bar{x} \in \mathbb{R}^n$, $\lambda \geq 0$ e $\tilde{f}_{(\lambda,\bar{x})} = f_{(\lambda,\bar{x})}{}_{|_\Omega}$ a restrição de $f_{(\lambda,\bar{x})}$, definida em (4-2), ao conjunto Ω. Se $\lambda > L$ então a função $\tilde{f}_{(\lambda,\bar{x})}$ é fortemente convexa com módulo $\lambda - L$. Além disso, $\Omega \subset \mathrm{dom}(f)$ e como consequência, $\mathrm{dom}(\tilde{f}_{(\lambda,\bar{x})}) = \Omega$.*

Prova. Sejam $\tau \in T$ e $\tilde{\varphi}_\tau = \varphi_\tau{}_{|_\Omega}$, a restrição de φ_τ ao conjunto Ω como na Definição 1.7. Primeiramente note que $\tilde{\varphi}_\tau$ é continuamente diferenciável em Ω e, além disso, as hipóteses **H1**, **H2** e **H3** são satisfeitas para $\tilde{\varphi}_\tau$. Agora tome $\tau \in T$ e $x, y \in \Omega$. De **H2**, **H3** e da hipótese $\lambda > L$, temos

$$\langle \nabla\tilde{\varphi}_\tau(x) + \lambda x - \nabla\tilde{\varphi}_\tau(y) - \lambda y,\ x - y\rangle \geq -L\|x-y\|^2 + \lambda\|x-y\|^2 \geq 0.$$

Portanto, pela Proposição 1.44, a função $\mathbb{R}^n \ni x \mapsto \tilde{\varphi}_\tau(x) + (\lambda/2)\|x - \bar{x}\|^2$ é fortemente convexa com módulo $\lambda - L > 0$ para todo $\tau \in T$. Note ainda que

$$\sup_{\tau \in T}\left\{\tilde{\varphi}_\tau(x) + \frac{\lambda}{2}\|x - \bar{x}\|^2\right\} = \tilde{f}_{(\lambda,\bar{x})}(x).$$

Como a função $\mathbb{R}^n \ni x \mapsto \tilde{\varphi}_\tau(x) + (\lambda/2)\|x - \bar{x}\|^2$ é fortemente convexa com módulo $\lambda - L > 0$, pela Proposição 1.31, $\tilde{f}_{(\lambda,\bar{x})}$ também é fortemente convexa com módulo $\lambda - L$.

Por outro lado, pelo Teorema Fundamental do Cálculo temos

$$\tilde{\varphi}_\tau(x) = \tilde{\varphi}_\tau(y) + \int_0^1 \langle \nabla\tilde{\varphi}_\tau(y + t(x-y)) - \nabla\tilde{\varphi}_\tau(y), x - y\rangle\, dt + \langle \nabla\tilde{\varphi}_\tau(y), x - y\rangle,$$

onde $\tau \in T$ e $x, y \in \Omega$. Agora usando o hipótese **H2** e algumas manipulações algébricas simples, obtemos da última expressão que

$$\tilde{\varphi}_\tau(x) \leq \tilde{\varphi}_\tau(y) + \frac{L_\tau}{2}\|y - x\|^2 + \langle \nabla\tilde{\varphi}_\tau(y), x - y\rangle.$$

Fazendo $y = \bar{x}$ na última expressão, de **H1** obtemos que $\sup_{\tau \in T}\|\nabla\tilde{\varphi}_\tau(\bar{x})\| = c < +\infty$, o que implica, usando ainda desigualdade de Schwarz, que

$$f(x) \leq f(\bar{x}) + \frac{L}{2}\|x - \bar{x}\|^2 + c\|x - \bar{x}\| < +\infty, \qquad \forall x \in \Omega,$$

pois, $\bar{x} \in \mathrm{dom}\, f$ e assim $\Omega \subset \mathrm{dom}(f)$. Assim concluímos a prova do teorema. $\square$

Corolário 4.2 *A função f é localmente Lipschitz em Ω.*

Prova. Primeiramente tome $\lambda > L$ e $x, y \in \Omega$ e $\tilde{f}_{(\lambda,\bar{x})} = f_{(\lambda,\bar{x})}|_{\Omega}$ a restrição de $f_{(\lambda,\bar{x})}$, definida em (4-2), ao conjunto Ω, $\bar{x} \in \mathbb{R}^n$. Usando a desigualdade triangular obtemos, após simples manipulações algébricas o seguinte resultado

$$|f(x) - f(y)| \leq \left|\tilde{f}_{(\lambda,\bar{x})}(x) - \tilde{f}_{(\lambda,\bar{x})}(y)\right| + \left|\frac{\lambda}{2}\|x-\bar{x}\|^2 - \frac{\lambda}{2}\|y-\bar{x}\|^2\right|.$$

Por outro lado, a Proposição 4.1 implica que $\tilde{f}_{(\lambda,\bar{x})}$ é convexa. Como $x \mapsto \|x-\bar{x}\|^2$ também é convexa, segue do Teorema 1.30 que ambas são localmente Lipschitz em Ω. Então, pela desigualdade acima, para x e y suficientemente próximos existem $K_1, K_2 > 0$ tais que

$$|f(x) - f(y)| \leq K_1\|x-y\| + \frac{\lambda}{2}K_2\|x-y\| = (K_1 + \frac{\lambda}{2}K_2)\|x-y\| = K\|x-y\|,$$

e assim concluimos que f é localmente Lipschitz em Ω. □

Proposição 4.3 *A seguinte inclusão é válida:*

$$\operatorname{conv}\{\nabla\varphi_{\bar{\tau}}(x) : \bar{\tau} \in T(x)\} \subset \partial_c f(x), \qquad \forall x \in \Omega,$$

onde $T(x) = \{\tau \in T : f(x) = \varphi(x,\tau)\}$.

Prova. Tome $x \in \Omega$. Como, pelo Corolário 4.2, f é localmente lipschitz em Ω temos

$$f^0(x;d) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y+td) - f(y)}{t} \geq \limsup_{t \downarrow 0} \frac{f(x+td) - f(x)}{t}.$$

Agora, usando a definição de f e do conjunto $T(x)$ obtemos que

$$f^0(x;d) \geq \limsup_{t \downarrow 0} \frac{\varphi_{\bar{\tau}}(x+td) - \varphi_{\bar{\tau}}(x)}{t}.$$

Como φ_τ é diferenciável para todo $\tau \in T$ segue imediatamente da última expressão que

$$f^0(x;d) \geq \varphi'_{\bar{\tau}}(x,d) = \langle \nabla\varphi_{\bar{\tau}}(x), d\rangle, \qquad \forall\, \bar{\tau} \in T(x), \quad d \in \mathbb{R}^n,$$

isto é, $\nabla\varphi_{\bar{\tau}}(x) \in \partial_c f(x)$ qualquer que seja $\bar{\tau} \in T(x)$. Assim pela convexidade de $\partial_c f(x)$

$$conv\{\nabla\varphi_{\bar{\tau}}(x) | \bar{\tau} \in T(x)\} \subset \partial_c f(x).$$

□

Exemplo 4.4 *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *continuamente diferenciável em algum conjunto convexo e aberto* $\Omega \subset \mathbb{R}^n$*. Definamos* $\varphi : \mathbb{R}^n \times [0,1] \to \mathbb{R} \cup \{+\infty\}$ *por*

$$\varphi(x,t) = \varphi_t(x) := -t^2 + f(x).$$

Claramente $f(x) = \sup_{t\in[0,1]} \varphi_t(x)$ *qualquer que seja* $x \in \mathbb{R}^n$*. Afirmamos que* **H1** *é automaticamente satisfeita e ainda as hipóteses* **H0**, **H2** *e* **H3** *são satisfeitas se* f *possuir as duas propriedades*

H0*) f *é semicontínua inferiormente;*

H2*) $\nabla f \in Lip_L(\Omega)$*;*

De fato, pois

$$\nabla \varphi_t(x) = \nabla f(x).$$

4.2 Método de Ponto Proximal para a Função Supremo

Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ como definida na seção 4.1 e satisfazendo as condições **H0**, **H1**, **H2** e **H3**. O método de *ponto proximal* para resolver o problema de otimização

$$\min_{x\in\mathbb{R}^n} f(x), \tag{4-3}$$

é definido formalmente da sequinte forma: Dados a sequência $\{\lambda_k\}$ de números reais positivos e $x_0 \in \mathrm{dom}(f)$ defina

$$x^{k+1} = \arg\min_{x\in\mathbb{R}^n} \left\{ f(x) + \frac{\lambda_k}{2} \|x - x^k\|^2 \right\}, \qquad k = 0, 1, \ldots. \tag{4-4}$$

Vamos supor ainda que f seja limitada inferiormente, isto é, $\inf_{x\in\mathbb{R}^n} f(x) > -\infty$.

4.2.1 Primeiro Caso

Nesta seção aplicaremos o método de ponto proximal para minimizar funções dadas pelo máximo de funções diferenciáveis, sobre um conjunto de índices finito. A prova de convergência será fortemente baseada nos Teoremas 1.40 e 4.3.

Nosso objetivo é provar o seguinte teorema de convergência para o método de ponto proximal.

Teorema 4.5 *Sejam* $y \in \mathbb{R}^n$ *e* $T = \{1, ..., m\}$ *um conjunto de índices. Suponhamos que* $\{x : f(x) \le f(y)\} \subset \Omega$*. Escolhendo* $\bar{\lambda} > 0$ *e* $\{\lambda_k\}$ *tais que*

$$0 < L = \max_{\tau\in T} L_\tau < \lambda_k \le \bar{\lambda}, \qquad k = 0, 1, \ldots. \tag{4-5}$$

Então a seqüência $\{x^k\}$, gerada pelo de método de ponto proximal com ponto inicial x^0, tal que $f(x^0) \leq f(y)$, está bem definida e satisfaz uma das seguinte afirmações:

i) a seqüência $\{x^k\}$ é finita, isto é, $x^k = x^{\bar{k}}$ para todo $k \geq \bar{k}$. Neste caso, $x^{\bar{k}}$ é um ponto estacionário de f;

ii) a seqüência $\{x^k\}$ é infinita e qualquer um de seus pontos de acumulação é ponto estacionário de f.

Se além disso, tivermos as seguintes condições satisfeitas:

a) $U^* = \{z \in \Omega : f(z) = \inf_{x\in\mathbb{R}^n} f(x)\} \neq \emptyset$;

b) *Dados $\lambda \in (L, \bar{\lambda}]$ e $f_\lambda(x) := f(x) + (\lambda/2)\|x\|^2$, existem $c_0 \in [f^*, f(y))$ e $d > 0$ tais que*

$$\|s(x)\| > d > 0, \qquad \forall\, x \in L_f(f(y)) \setminus L_f(c_0), \quad \forall\, s(x) \in \partial f_\lambda(x) - \lambda x.$$

c) *$L_f(c_0)$ é convexo, e $\tilde{f} = f_{|_{L_f(c_0)}}$ é convexo.*

Então a sequência $\{x^k\}$, gerada pelo de método de ponto proximal, converge para $x^* \in U^*$.

Observação 4.6 *Se f é convexa, não constante em Ω e com conjunto solução não vazio e limitado, então as condições* **b** *e* **c** *são automaticamente satisfeitas. Com efeito, como f é convexa qualquer conjunto de nível será convexo, Proposição 1.33, a condição* **c** *é satisfeita.*

Agora veremos que f satisfaz a condição **b**. *Seja $y \in \Omega$ tal que $c = f(y) > f^*$. Então tomando $c_0 \in (f^*, c)$ e $z \in U^*$, a convexidade de f nos dá:*

$$f(z) \geq f(x) + \langle s(x), z - x\rangle \qquad \forall\, x \in L_f(c) \setminus L_f(c_0), \qquad s(x) \in \partial f(x) = \partial f_\lambda(x) - \lambda x.$$

Fazendo $\delta = c_0 - f^* > 0$, como $c \geq f(x) > c_0 > f^*$, temos

$$\langle s(x), x - z\rangle \geq f(x) - f(z) > \delta > 0. \tag{4-6}$$

Por outro lado, como f é convexa e o conjunto solução é limitado segue da Proposição 1.34 que $L_f(c_0)$ é limitado. Logo existe $M > 0$ tal que

$$M = \sup\left\{\|x - z\| : x \in L_f(c) \setminus L_f(c_0)\right\}.$$

Seja $d = \delta/M$, então usando (4-6) e a desigualdade de Cauchy-Schwarz, obtemos

$$0 < \delta < \|s(x)\|\|x - z\| \leq M\|s(x)\|, \qquad \forall\, x \in L_f(c) \setminus L_f(c_0), \qquad s(x) \in \partial f_\lambda(x) - \lambda x,$$

que é a condição **b**.

Apresentamos abaixo um exemplo interessante mostrando que, em geral, a condição **b** não vale para uma função convexa com conjunto solução ilimitado.

Exemplo 4.7 *Seja a função* $f : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ *definida por*

$$f(x,y) = \frac{x^2}{1+y}.$$

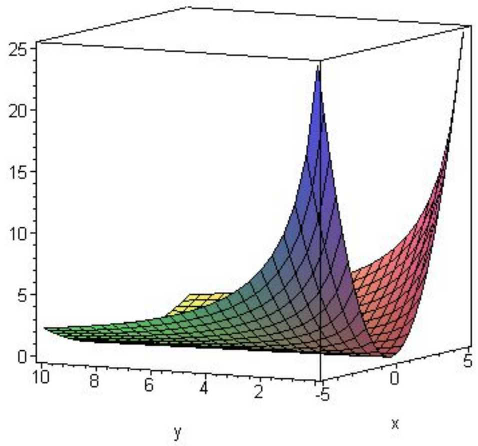

Figura 4.1: *Função convexa que não satisfaz a condição* **b**.

Afirmamos que f *é convexa e não satisfaz a condição* **b**. *Com efeito, o gradiente e a hessiana de* f *são, respectivamente, dados por*

$$\nabla f(x,y) = \left(\frac{2x}{1+y}, \frac{-x^2}{(1+y)^2} \right), \qquad \nabla^2 f(x,y) = \begin{pmatrix} \dfrac{2}{1+y} & \dfrac{-2x}{(1+y)^2} \\ \\ \dfrac{-2x}{(1+y)^2} & \dfrac{2x^2}{(1+y)^3} \end{pmatrix}.$$

A hessiana de f *possui determinante nulo e o termo* $2/(1+y) > 0$ *para todo ponto* $(x,y) \in \mathbb{R} \times \mathbb{R}_+$*, o que implica que ela é semi-definida positiva. Logo* f *é convexa. Tome* $c > \bar{c} > c_0 \geq 0 = f^*$ *e note que*

$$\left\{ (x, \frac{x^2}{\bar{c}} - 1) : x \in \mathbb{R} \right\} \subset L_f(c) \setminus L_f(c_0),$$

e assim concluímos que

$$\lim_{x\to+\infty} \nabla f(x, \frac{x^2}{\bar{c}} - 1) = \lim_{x\to+\infty} \left(\frac{2c}{x}, \frac{-c^2}{x^2} \right) = (0,0),$$

que contraria a condição **b**.

Antes de provarmos o Teorema 4.5 mostraremos alguns lemas importantes, de maneira que sua prova será consequência imediata destes lemas. Assumimos todas as hipóteses do Teorema 4.5, exceto **a**, **b** e **c** a menos que elas sejam mencionadas explicitamente.

Lema 4.8 *A sequência $\{x^k\}$, gerada pelo método de ponto proximal (4-4) aplicado a f com ponto inicial $x^0 \in \{x : f(x) \le f(y)\}$ e $\{\lambda_k\}$ satisfazendo (4-5), está bem definida e $\{x^k\} \subset \{x : f(x) \le f(y)\}$.*

Prova. Inicialmente observe que, tomando $\bar{x} \in L_f(f(y))$ e $\lambda \in (L, \bar{\lambda}]$, $f_{(\lambda,\bar{x})}$, definida em (4-2), é semicontínua inferiormente e

$$L_{f_{(\lambda,\bar{x})}}(f(\bar{x})) \subset L_f(f(y)) \subset \Omega. \tag{4-7}$$

Assim, pela Proposição 1.33, $L_{f_{(\lambda,\bar{x})}}(f(\bar{x}))$ é fechado. Por outro lado, pela Proposição 4.1,

$$\tilde{f}_{(\lambda,\bar{x})} := f_{(\lambda,\bar{x})}{}_{|_\Omega},$$

é fortemente convexa com módulo $(\lambda - L)$. Assim, pelo Lema 1.46 juntamente com (4-7), o conjunto $L_{f_{(\lambda,\bar{x})}}(f(\bar{x}))$ é limitado, logo compacto. Note ainda que, como $f_{(\lambda,\bar{x})}$ é semicontínua inferiormente, $f_{(\lambda,\bar{x})}$ possui um mínimo em $L_{f_{(\lambda,\bar{x})}}(f(\bar{x}))$ que, obviamente, é um mínimo global de $f_{(\lambda,\bar{x})}$ e ainda, de $\tilde{f}_{(\lambda,\bar{x})}$ ser fortemente convexa juntamente com (4-7), este é o único mínimo. Portanto, como $x^0 \in L_f(f(y))$ e a sequência $\{\lambda_k\}$ satisfaz

$$0 < L = \max_{i\in I} L_i < \lambda_k \le \bar{\lambda}, \qquad k = 0, 1, \ldots,$$

o resultado segue por argumento trivial de indução. □

Seja $\{x^k\}$ a sequência gerada pelo método de ponto proximal (4-4) aplicado a f com ponto inicial $x^0 \in L_f(f(y))$ e $\{\lambda_k\}$ satisfazendo (4-5). Defina $f_k : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ como

$$f_k(x) := f_{(\lambda_k, x^k)}(x) = f(x) + \frac{\lambda_k}{2}\|x - x^k\|^2. \tag{4-8}$$

Lema 4.9 *Seja $\{x^k\}$ a sequência gerada pelo método de ponto proximal (4-4) aplicado a f com ponto inicial $x^0 \in L_f(f(y))$ e $\{\lambda_k\}$ satisfazendo (4-5). Então* *i)* $f_k(x^{k+1}) \leq f(x^k), \quad k = 0, 1, \ldots$, *isto é,* $\{f(x^k)\}$ *é monótona não-crescente;*

ii) $\{x^k\} \subset L_f(f(x_0))$*;*

iii) $0 \in \partial_c f_k(x^{k+1}), \quad k = 0, 1, \ldots$*;*

iv) *a sequência* $\{f(x^k)\}$ *converge;*

v) $\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$.

Prova. Os itens *i* e *iii* seguem de x^{k+1} minimizar a função f_k, necessariamente a restrição de f_k a Ω que é fortemente convexa, Proposição 4.1, e do Teorema 1.47. Do item *i*, segue imediatamente que $f(x^k) \leq f(x^{k-1}) \leq \cdots \leq f(x^0)$, isto é, $\{f(x^k)\}$ é monótona não-crescente. Logo o item *ii* segue-se. Agora, como $\{f(x^k)\}$ é monótona não-crescente e f é limitada inferiormente, $\{f(x^k)\}$ converge e o item *iv* fica provado. E por fim o item *v* segue combinando os itens *i* e *iv*. □

Lema 4.10 *Seja* $\{x^k\}$ *a sequência gerada pelo método de ponto proximal* (4-4) *aplicado a* f *com ponto inicial* $x^0 \in L_f(f(y))$ *e* $\{\lambda_k\}$ *satisfazendo* (4-5)*. Então* $\{x^k\}$ *satisfaz uma das seguinte afirmações:*

i) *a seqüência* $\{x^k\}$ *é finita, isto é,* $x^k = x^{\bar{k}}$ *para todo* $k \geq \bar{k}$*. Neste caso,* $x^{\bar{k}}$ *é um ponto estacionário de* f*;*

ii) *a seqüência* $\{x^k\}$ *é infinita e qualquer um de seus pontos de acumulação é ponto estacionário de* f*.*

Prova. Para cada k, segue do Teorema 1.48 combinando com a Proposição 4.1 que

$$\partial f_k(x^{k+1}) = \operatorname{conv}\left\{\nabla\varphi_j(x^{k+1}) + \lambda_k(x^{k+1} - x^k) : j \in T(x^{k+1})\right\}, \tag{4-9}$$

onde $T(x^{k+1}) = \left\{j \in T : \varphi_j(x^{k+1}) = f(x^{k+1})\right\}$. Por outo lado, o Lema 4.9 item *iii* implica que

$$0 \in \partial f_k(x^{k+1}).$$

Assim, usando a última expressão junto com (4-9) concluímos que existem $\alpha_j^{k+1} \geq 0$, com $j \in T(x^{k+1})$, tais que

$$0 = \sum_{j\in T(x^{k+1})} \alpha_j^{k+1}\nabla\varphi_j(x^{k+1}) + \lambda_k(x^{k+1} - x^k), \qquad \sum_{j\in T(x^{k+1})} \alpha_j^{k+1} = 1. \tag{4-10}$$

Deste modo temos duas possibilidades: a sequência $\{x^k\}$ é finita, isto é, $x^k = x^{\bar{k}}$ para $k > \bar{k}$ ou a sequência $\{x^k\}$ é infinita

No primeiro caso, de (4-10) obtemos

$$0 = \sum_{j \in T(x^{\bar{k}})} \alpha_j^{\bar{k}+1} \nabla \varphi_j(x^{\bar{k}}), \qquad \sum_{j \in T(x^{\bar{k}})} \alpha_j^{\bar{k}+1} = 1.$$

Portanto, pela Proposição 4.3, $0 \in \partial_c f(x^{\bar{k}})$, isto é, $x^{\bar{k}}$ é um ponto estacionário de f. E i fica provado.

No segundo caso, suponha ainda que $\{x^k\}$ possui um ponto de acumulação $\bar{x}$. Assim, pelo Lema 4.9 itens i e ii concluímos que $\bar{x} \in L_f(f(y))$. Agora considere a seguinte sequência $\{\alpha^{k+1}\} \subset \mathbb{R}^m$ definida por

$$\alpha^{k+1} = (\alpha_1^{k+1}, \ldots, \alpha_m^{k+1}), \qquad \alpha_j^{k+1} = 0, \quad j \in T \backslash T(x^{k+1}).$$

Como $\sum_{j \in T(x^{k+1})} \alpha_j^{k+1} = 1$ nos temos $\|\alpha_j^{k+1}\|_1 = 1$ para todo k, onde $\| \ \|_1$ denota a norma da soma em $\mathbb{R}^m$. Sendo assim é possível tomar $\{x^{k_i+1}\}$ e $\{\alpha_j^{k_i+1}\}$ subsequências de $\{x^{k+1}\}$ e $\{\alpha_j^{k+1}\}$, respectivamente, tais que

$$\lim_{i \to +\infty} x^{k_i+1} = \bar{x}, \qquad \lim_{i \to +\infty} \alpha_j^{k_i+1} = \bar{\alpha}_j.$$

Como T finito podemos, passando a outra subsequência se necessário, assumir que

$$T(x^{k_1+1}) = T(x^{k_2+1}) = \cdots = \bar{T}.$$

Por outro lado, pelo Lema 4.9, $\lim_{k \to \infty} \|x^{k+1} - x^k\| = 0$. Portanto tomando $k = k_i$ em (4-10) e fazendo i ir a infinito, da continuidade de $\nabla \varphi_j$, concluímos

$$0 = \sum_{j \in \bar{T}} \bar{\alpha}_j \nabla \varphi_j(\bar{x}), \qquad \sum_{j \in \bar{T}} \bar{\alpha}_j = 1.$$

Como, claramente, $T(\bar{x}) \supset \bar{T}$, a expressão anterior junto com a Proposição 4.3 implica que $0 \in \partial_c f(\bar{x})$, isto é, $\bar{x}$ é um ponto estacionário de f, o que prova o item ii. $\square$

Assumindo também, daqui por diante, as condições **a**, **b** e **c** demonstraremos que a sequência gerada pelo do método, com $\max_{\tau \in T} L_\tau < \lambda_i < \bar{\lambda}$ e $x^0 \in L_f(f(y))$, converge para um ponto solução $x^* \in U^*$. Antes para isto precisaremos de dois resultados preliminares.

Lema 4.11 *Seja $\{x^k\}$ a sequência gerada pelo método de ponto proximal* (4-4) *aplicado a f com ponto inicial $x^0 \in L_f(f(y))$ com λ_k satisfazendo* (4-5)*. Se* **a**-**c** *valem então existe k_0 tal que*

$$x^k \in L_f(c_0), \qquad \forall\, k > k_0.$$

Prova. Tome $x^0 \in L_f(f(y)) \setminus L_f(c_0)$, $\lambda \in (L, \bar{\lambda}]$ e f_λ, como na condição **b** do Teorema 4.5.

Considere a seguinte igualdade

$$f(x)+\frac{\lambda_k}{2}\|x-x^k\|^2 = f_\lambda(x)+\frac{\lambda_k-\lambda}{2}\|x\|^2-\lambda_k\langle x,x_k\rangle+\frac{\lambda_k}{2}\|x^k\|^2. \tag{4-11}$$

Suponha, por absurdo, que $x^k \in L_f(f(y))\setminus L_f(c_0)$ para todo k. Assim, como $f_{k|_\Omega}$ e $f_{\lambda|_\Omega}$ são convexas, Proposição 4.1, de (4-11) juntamente com a Proposição 4.9 obtemos

$$0\in\partial f_k(x^{k+1}) = \left(\partial f_\lambda(x^{k+1})-\lambda x^{k+1}+\lambda_k(x^{k+1}-x^k)\right),$$

implicando assim que $\lambda_k(x^k-x^{k+1})\in\left(\partial f_\lambda(x^{k+1})-\lambda x^{k+1}\right)$. Logo, pela condição **b** do Teorema 4.5

$$\lambda_k\|x^{k+1}-x^k\|>d \quad\Rightarrow\quad |x^{k+1}-x^k\|>\frac{d}{\bar{\lambda}},$$

o que pelo item *v* da Lema 4.9 é um absurdo. Portanto, a sequência $\{x^k\}$ entra no conjunto $L_f(c_0)$, isto é, $x^k\in L_f(c_0)$ para todo k maior que algum k_0. □

Lema 4.12 *Seja $\{x^k\}$ a sequência gerada pelo método de ponto proximal* (4-4) *aplicado a f com ponto inicial $x^0\in L_f(f(y))$ e λ_k satisfazendo* (4-5). *Então existe k_0 tal que*

$$\|x-x^{k+1}\|^2\le\|x-x^k\|^2+\frac{1}{\lambda_k}(f(x)-f(x^{k+1})), \qquad x\in L_f(c_0), \quad k>k_0. \tag{4-12}$$

Prova. Seja k_0 como no Lema 4.11. Então $x^k\in L_f(c_0)$ para $k>k_0$. Pelo o Lema 4.9 item *iii* temos que

$$0\in\partial f_k(x^{k+1}), \qquad k>k_0.$$

Assim, como $f_{|_{L_f(c_0)}}$ é convexa, do Teorema 1.41 juntamente com (4-8) temos

$$2\lambda_k(x^k-x^{k+1})\in\partial f(x^{k+1}), \qquad k>k_0,$$

e pela Proposição 1.42

$$\frac{1}{\lambda_k}(f(x)-f(x^{k+1}))\ge 2\langle x^k-x^{k+1},x-x^{k+1}\rangle \qquad x\in L_f(c_0), \quad k>k_0.$$

Por outro lado,

$$\|y-x^k\|^2=\|y-x^{k+1}\|^2-2\langle x^k-x^{k+1},y-x^{k+1}\rangle+\|x^{k+1}-x^k\|^2.$$

Então, combinando as duas últimas expressão obtemos:

$$\|x-x^{k+1}\|^2\le\|x-x^k\|^2-\|x^{k+1}-x^k\|^2+\frac{1}{\lambda_k}\left(f(x)-f(x^{k+1})\right), \qquad x\in L_f(c_0), \quad \forall k>k_0.$$

Como $\|x^{k+1}-x^k\|^2$ é maior ou igual a zero o resultado segue. $\square$

Lema 4.13 *Seja $\{x^k\}$ a sequência gerada pelo método de ponto proximal (4-4) aplicado a f com ponto inicial $x^0 \in L_f(f(y))$ e $\{\lambda_k\}$ satisfazendo (4-5). Suponha que* **a**-**c** *valem. Então $\{x^k\}$ converge para $x^* \in U^* \subset \Omega$.*

Prova. Seja k_0 como no Lema 4.11. Trocando em (4-12) x por $z \in U^* \subset L_f(c_0)$ teremos que

$$\|z-x^{k+1}\|^2 \leq \|z-x^k\|^2, \qquad k > k_0, \tag{4-13}$$

pois $f(z)-f(x^{k+1}) \leq 0$. O que implica que $\{x^k\}$ é limitada, ou seja,

$$\|x^{k+1}\| \leq \|z-x^{k_0}\| + \|z\|, \qquad k > k_0.$$

Portanto $\{x^k\} \subset L_f(f(y)) \subset \Omega$ possui um ponto de acumulação $\bar{x} \in \mathbb{R}^n$. Como f é semicontínua inferiormente e $x^k \in L_f(c_0)$ para $k \geq k_0$ então $\bar{x} \in L_f(c_0)$. Assim, pelo Lema 4.10, $\bar{x}$ é estacionário. Como $f_{|_{L_f(c_0)}}$ é convexa o Teorema 1.47 nos dá que $\bar{x} \in U^*$. Por outro lado (4-13) mostra que $\{x^k\}$ converge para $\bar{x}$. O que conclui a prova. $\square$

Prova.**Teorema 4.5** O teorema sai imediatamente combinando os Lemas 4.10 e 4.13. $\square$

4.2.2 Segundo Caso

Nesta seção aplicaremos o método de ponto proximal para minimizar funções dadas pelo supremo de funções diferenciáveis, agora, sobre um conjunto não necessariamente finito. É importante observar que o caso anterior pode ser visto como um caso particular deste, mas a sua prova foi feita pois usa uma técnica diferente da que será usada agora. A técnica que será utilizada neste caso será fortemente baseada no Teorema 3.6, que se refere o subdiferencial de Clarke.

Assim nosso objetivo, aqui, é provar o seguinte teorema de convergência para o método de ponto proximal.

Teorema 4.14 *Seja $y \in \mathbb{R}^n$. Suponhamos que $\{x : f(x) \leq f(y)\} \subset \Omega$. Escolhendo $\bar{\lambda} > 0$ e $\{\lambda_k\}$ tais que*

$$0 < L = \sup_{\tau \in T} L_\tau < \lambda_k \leq \bar{\lambda}, \qquad k = 0,1,\ldots. \tag{4-14}$$

Então a seqüência $\{x^k\}$, gerada pelo de método de ponto proximal, com ponto inicial x^0 tal que $f(x^0) \leq f(y)$ está bem definida e satisfaz uma das seguinte afirmações:

i) a seqüência $\{x^k\}$ é finita, isto é, $x^k = x^{\bar{k}}$ para todo $k \geq \bar{k}$. Neste caso, $x^{\bar{k}}$ é um ponto estacionário de f;

ii) a seqüência $\{x^k\}$ é infinita e qualquer um de seus pontos de acumulação é ponto estacionário de f.

Se além disso, tivermos as seguintes condições:

a) *$U^* = \{z \in \Omega : f(z) = \inf_{x\in\mathbb{R}^n} f(x)\} \neq \emptyset$;*

b) *Dados $\lambda \in (L, \bar{\lambda}]$ e $f_\lambda(x) := f(x) + (\lambda/2)\|x\|^2$, existem $c_0 \in [f^*, f(y))$ e $d > 0$ tais que*

$$\|s(x)\| > d > 0, \qquad \forall\, x \in L_f(f(y)) \setminus L_f(c_0), \quad \forall\, s(x) \in \partial f_\lambda(x) - \lambda x;$$

c) *$L_f(c_0)$ é convexo e $\tilde{f} = f_{|_{L_f(c_0)}}$ é convexa.*

Então a sequência $\{x^k\}$, gerada pelo método de ponto proximal converge para $x^* \in U^*$.

Observação 4.15 *Podemos observar que o Lema 4.8 também é válido, pois não depende da finitudo do conjunto T. Assim isto nos dá que a sequência gerada pelo método de ponto proximal aplicado a f, com λ_k satisfazendo (4-14), está bem definida. Observamos ainda que o Lema 4.9 é válido neste caso.*

Agora para demonstração do Teorema 4.14 será de crucial importância os seguintes lemas.

Lema 4.16 *Para f satisfazendo as condições do Teorema 4.14 e $\{x^k\}$ a sequência gerada pelo método de ponto proximal (4-4) com ponto inicial $x^0 \in L_f(f(y))$ e $\{\lambda_k\}$ satisfazendo (4-14). A seguinte afirmação é verdadeira*

$$\partial f_k(x) \subset \partial_c f(x) + \lambda_k(x - x^k), \qquad x \in \Omega.$$

Prova. É consequência imediata da Proposição 3.8 e comentário anterior ao Teorema 3.9. □

Lema 4.17 *Seja $\{x^k\}$ a sequência gerada pelo método de ponto proximal (4-4) aplicado a f com ponto inicial $x^0 \in L_f(f(y))$ e $\{\lambda_k\}$ satisfazendo (4-14). Então $\{x^k\}$ satisfaz uma das seguinte afirmações:*

i) a seqüência $\{x^k\}$ é finita, isto é, $x^k = x^{\bar{k}}$ para todo $k \geq \bar{k}$. Neste caso, $x^{\bar{k}}$ é um ponto estacionário de f;

ii) a seqüência $\{x^k\}$ é infinita e qualquer um de seus pontos de acumulação é ponto estacionário de f.

Prova. Dos Lemas 4.9, 4.16 e de (4-8) temos que

$$0 \in \partial f_k(x^{k+1}) \subset \partial_c f(x^{k+1}) + \lambda_k(x^{k+1} - x^k),$$

e portanto $\lambda_k(x^k - x^{k+1}) \in \partial_c f(x^{k+1})$. Assim se $x^k = x^{\bar{k}}$ para $k > \bar{k}$ temos que $0 \in \partial_c f(x)$, ou seja, $x^{\bar{k}}$ é um ponto estacionário de f.

Agora suponhamos que $\{x^k\}$ seja uma seqüência infinita e $\bar{x} = \lim_{i\to\infty} x^{k_i}$, onde $\{x^{k_i}\}$ é uma subsequência de $\{x^k\}$. Note que, pelo Teorema 3.6, ∂f é semicontínua superiormente. Então, como $\lambda_k(x^k - x^{k+1}) \in \partial_c f(x^{k+1})$ e λ_k satisfaz 4-14, o Lema 4.9 item v junto com a semicontinuidade superior de ∂f implica que

$$0 = \lim_{k\to\infty} \lambda_k(x^{k_i+1} - x^{k_i}) \in \partial f(\bar{x}),$$

e assim $\bar{x}$ é um ponto estacionário de f. □

Observação 4.18 *Também é importante ressaltar que o Lema 4.13 também é verificado, mas agora, usando o Lema 4.17 em lugar do Lema 4.10.*

Prova.(**Teorema 4.14**) A demonstração segue imediatamente dos Lemas 4.17 e 4.13. □

Exemplo 4.19 *Conforme foi feito no Exemplo 4.4. Podemos tomar*

$$f(x) = \begin{cases} x^2 \ln(x), & x \in (0, +\infty), \\ 0, & x = 0, \\ +\infty, & x \in (-\infty, 0), \end{cases}$$

*e $\Omega = (\varepsilon, d)$, com $0 < \varepsilon < \sqrt{e^{-1}} < d$, assim teremos satisfeitas todas as condições do Teorema 4.14. (**H0***) f é, obviamente, semicontínua inferiormente, (**H2***) f'' é limitada em Ω e, além disso, Ω contém um conjunto de nível de f.*

Observações Finais

Nesta dissertação, que tem como referência principal o trabalho de Kaplan, A. e Tichatschke [8], fizemos um estudo do método de ponto proximal aplicado a uma classe de funções não convexas dadas formalmente por, dados $T \subset \mathbb{R}^m$, $\Omega \subset \mathbb{R}^n$ um conjunto aberto e convexo e $\varphi : \mathbb{R}^n \times T \to \mathbb{R} \cup \{+\infty\}$,

$$f(x) = \sup_{\tau \in T} \varphi_\tau(x).$$

onde $\varphi_\tau(x) = \varphi(x,\tau)$ para todo $x \in \mathbb{R}^n$ e $\tau \in T$. Ainda é assumido algumas propriedades para as funções pertencentes a esta classe, de modo que se obtenha boa definição e convergência local do método. Em particular, a convergência local é garantida quando se impõe, entre outras coisa, a convexidade da função objetivo sobre um determinado conjunto de nível. Uma continuação natural deste trabalho seria um estudo da convergência local do método para funções "$Lower - C^2$"como é apresentado em [3] e ainda nesta direção, mas em um contexto mais abrangente, a extensão dos resultado para operadores monótonos seria também muito interessante.

Referências Bibliográficas

[1] Brezis, hanm *Analisis funcional/ Functional Analysis (Spanish Edition)*. Alianza Editorial S.A, Madrid (1984).

[2] Clarke, F. H. *Optimization and Nonsmooth Analysis*. Classics in applied mathematics. 5, SIAM, New York (1983).

[3] Hare, W. e Sagastizábal, C. *Computing proximal points of nonconvex functions*. Math. Program., Ser. B no. 116, 221-258, (2009).

[4] Hiriart-Urruty, J.-B e Lemaréchal, C. *Convex analysis and minimization algorithms I and II*. Springer-Verlag, (1993).

[5] Iusem, A. N. *Métodos de ponto proximal em otimização* IMPA, Rio de Janeiro (1995).

[6] Iusem, A. N. . *Augmented Lagrangian Methods and Proximal Point Methods for Convex Optimization*. Investigación Operativa, v. 8, p. 11-49, (1999).

[7] Kolmogorov, A. N e Fomin, S. V. *Elementos da teoria das funções e de análise funcional*. Editora Mir, Moscou (1982).

[8] Kaplan, A., Tichatschke, R. *Proximal point methods and nonconvex optimization*. J. Global Optim. 13, no. 4, 389-406, (1998).

[9] Makela, Marko M. e Neittaanmaki, Pekka *Nonsmooth Optimization Analysis and Algorithms With Applications to Optimal Control*. Singapore ; World Scientific, New Jersey (1992).

[10] Martinet, B. *Régularisation, d'inéquations variationelles par approximations successives*. (French) Rev. Française Informat. Recherche Opérationnelle 4 , Ser. R-3, 154-158, (1970).

[11] Pennanen, Teemu *Local convergence of the Proximal Point Algorithm and Multiplier Methods without Monotonicity*. Mathematics of Operations Research. Vol 27,. No. 1. , 170-191, (2002).

[12] Pennanen, T. e Svaiter, B. F. *Solving monotone inclusions with linear multi-step methods*. SIAM J. Optim. Vol 13,. No. 4., 1080-1097, (2003).

[13] Pennanen, T., Iusem, A. N. e Svaiter, B. F. *Inexact variants or the Proximal Poinst Algorithm without Monotonicity*. SIAM J. Optim. Vol 13,. No. 4. , 1080-1097, (2003).

[14] Rockafellar, R. T. *Monotone operators and the proximal point algorithm*. SIAM J. Control. Optim. 14, 877-898, (1976).

[15] Solodov, M. V. ; Izmailov, A. *Otimização, Volume 1: Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade*. IMPA, Rio de Janeiro (2005).