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

Convergência do Método Subgradiente para Funções Quase-Convexas

João Carlos de Oliveira Souza

Teresina - 2010

João Carlos de Oliveira Souza

Dissertação de Mestrado:

Convergência do Método Subgradiente para Funções Quase-Convexas

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

Orientador:
Prof. Dr. João Xavier da Cruz Neto

Teresina - 2010

FICHA CATALOGRÁFICA

Serviço de Processamento Técnico da Universidade
Federal do Piauí

Biblioteca Comunitária Jornalista Carlos Castelo Branco

S729c Souza, João Carlos de Oliveira

Convergência do método subgradiente para funções quase-convexas/ João Carlos de Oliveira Souza.- 2010.

58 f. il.

Dissertação (Mestrado) - Pós-Graduação em Matemática da Universidade Federal do Paiuí, 2010.

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

1. Otimização matemática I. Título.

CDD: 519.3

Dedico este trabalho aos meus pais, Carlos e Domingas, a minha esposa Karyana e a Francisco Martins, José Alcione e Mara Beatriz. (In memoriam).

Agradecimentos

Agradeço, primeiramente a Deus, pela vida, pela saúde e disposição para os estudos.

Agradeço a meus pais, Carlos Henrique e Domingas, pelo amor, pela educação, pelo exemplo e principalmente pelo sacrifício que fizeram para que eu pudesse conseguir tudo o que tenho. Certamente, se não fosse por eles, eu não teria conquistado nada do que tenho. Ao meu pai, especialmente pelo exemplo de caráter e a minha mãe pelo exemplo de superação. Agradeço também a meus irmãos, José Rodolpho e Maria luiza pela convivência. Agradeço também a todos os outros familiares, que em algum momento, de alguma forma, me ajudaram.

Agradeço a minha esposa, Karyana, que sempre esteve ao meu lado ajudando, transmitindo muito amor, carinho e força. Com certeza você é responsável por grande parte de minhas conquistas e as dividirei para sempre com você, juntamete com as suas. Muito obrigado pela compreensão, paciência e dedicação. Te amo.

Agradeço ao meu orientador, Prof. Dr. João Xavier da Cruz Neto, por todas as opotunidades e confiança. Pelos ensinamentos matemáticos e da vida. Agradeço todos os seus conselhos e principalmente pela ajuda na preparação deste trabalho. Como grande matemático que é, contribuiu substancialmente na minha formação e como grande amigo, também contribuiu na formação do meu caráter. O senhor sempre terá meu respeito e gratidão.

Agradeço ao Prof. Dr. Paulo Roberto Oliveira por ter aceitado participar da banca, por ler este trabalho e por todas as sugestões feitas.

Agradeço a todos os professores da Pós-Graduação em Matemática, em especial Jurandir, Paulo Alexandre, Paulo Sérgio, Sissy, Barnabé, Newton, Juscelino, Humberto, Mancondes e Roger (com quem tive uma iniciação científica que mudou o meu olhar para matemática), que contribuiram bastante na miha formação matemática e ao Prof. Paulo Roberto por ter aceito participar da banca examinadora. Agradeço também a todos os professores da graduação, em especial Gilvan, João Benicio, Mário e Vicente. Agradeço também os funcionários do Departamento de Matemática.

Agradeço aos meus amigos de mestrado, com quem convivi por muitas horas nos últimos cinco anos, tanto estudando quanto nos poucos momentos de folga, Pedro Jorge, Arimatéa, João Santos (amigos desde os tempos de graduação), Gilberto, Daniel, Domingos, Venâncio, Cleyton e demais alunos. Aos amigos da graduação, em particular Maurício e Diego.

Agradeço ao CNPq pelo apoio financeiro.

Diante de Deus, somos todos igualmente sábios
e igualmente tolos.

Albert Einstein

Resumo

Neste trabalho, consideramos o problema de minimizar uma função quase-convexa e $Hölder$, não necessariamente diferenciável. Para tal, utilizamos as direções normalizadas do cone normal do conjunto de nível da função e empregamos a escolha do passo baseado no conhecimento *a priori* do valor ótimo da função custo. Mostramos que, tomado dessa forma, esse método possui informações adicinais sobre o problema considerado e que pode ser usado para melhorar taxas de convergência, baseado nas propriedades do subgradiente usual. Também apresentamos alguns resultados de convergência para versões inexatas do método e exemplos de implementações computacionais.

Abstract

In this work, we consider the problem of minimizing a quasiconvex and $H\ddot{o}lder$ function , not necessarily differentiable. To this end, we use the normalized directions of the normal cone of the set level of function and employ the stepsize rule based on *a priori* knowledge of the optimal value of the cost function. We show that this method possess additional information about the problem under consideration and can be used for improving convergence rates based on the usual subgradient properties. We also present several convergence results for inexact versions of the method and examples of computational implementations.

Sumário

Introdução

Seja uma função real $f : K \to \mathbb{R}$, onde $K \subset \mathbb{R}^n$ é um subconjunto não vazio. Considere o problema de minimizar a função f, em K. Este problema de minimização restrito pode ser denotado da seguinte maneira:

$$\min_{x \in K} f(x).$$

Quando, no problema acima, temos $K = \mathbb{R}^n$, dizemos que temos um problema irrestrito. Quando o conjunto K é convexo e a função f é convexa, dizemos que temos um problema de minimização convexa. Existem, na literatura, muitos métodos que resolvem estes problemas. Um método bastante utilizado é o método do subgradiente, no qual, dado um ponto inicial $x^0 \in \mathbb{R}^n$, contruímos uma sequência $\{x^k\}$, dada por

$$x^{k+1} = x^k - \lambda_k g^k,$$

com $\{\lambda_k\}$ uma sequência de números reais não negativos e $\{g^k\}$ uma sequência na qual $g^k \in \partial f(x^k),\ \forall k \in \mathbb{N}$, onde $\partial f(x^k)$ denota o subdiferencial de f no ponto x^k, para cada $k \in \mathbb{N}$. Com algumas hipóteses sob λ_k, prova-se que a sequência $\{x^k\}$ converge para a solução do problema. Em [2], [3] e [5-9] podemos encontrar algumas escolhas para $\{\lambda_k\}$ e observar várias propriededas de convergência deste e de outros métodos. Além da convergência, um outro fator importante é a velocidade que a sequência converge para a solução do problema. Para isso, veremos a definição e alguns tipos de taxa de convergência, que também podem ser encontradas em [10] e [11].

O objetivo central deste trabalho é utilizar uma variação do método do subgradiente para resolver problemas de minimização, onde a função custo é quase-convexa (em [15] e [18] obtemos várias propriedades de funções quase-convexas) e apresentar uma condição em que o método converge geometricamente para a solução do problema de minimização. Na demonstração da convergência do método, apresentamos um prova diferente da encontrada em [12]. A motivação para se trabalhar com funções quase-convexas é que, além de serem uma classe mais ampla que as funções convexas, existem várias ciências e campos da engenharia que utilizam a programação quase-convexa, tais como economia, teoria das aproximações, teoria do controle, etc. (para maiores detalhes, veja [19]).

Para isso, abordaremos no capítulo 1 noções de análise no $\mathbb{R}^n$, conceitos e resultados de otimização, tais como, ponto estacionário, minimizadores locais e globais, e conceitos e resultados de análise convexa, como função convexa, quase-convexa, pseudo-convexa e teorema de projeção e de separação (para um aprofundamento maior deste conteúdo veja [4] e [9]). No capítulo 2, veremos a definição e algumas propriedades do subdiferencial de Fénchel-Moreau, de Plastria (proposto em [13] e [14]) e definiremos sequência Féjer e quase-Féjer convergente (como em [16] e [17]). No capítulo 3, provaremos convergências do método do subgradiente para o caso convexo, como em [3]. Finalmente, no capítulo 4 estudaremos alguns tipos de taxa de convergência, provaremos a convergência do método do subgradiente para o caso quase-convexo, veremos algumas versões inexatas do método, que podem ser encontradas em [12] e exemplos de implementações computacionais.

Capítulo 1

Noções Preliminares

Neste capítulo, trataremos de conceitos básicos que serão utilizados posteriormente. Para isso, o dividiremos em seções que abordaremos conceitos e resultados de análise no $\mathbb{R}^n$, análise convexa e otimização.

1.1 Conceitos e resultados de Análise no $\mathbb{R}^n$

Inicialmente, veremos a definição de produto interno no $\mathbb{R}^n$.

Definição 1. *Um produto interno no espaço euclidiano $\mathbb{R}^n$ é uma regra que faz corresponder a cada par de vetores $x, y \in \mathbb{R}^n$ um número real, denotado por $\langle x, y \rangle$, de modo que, para quaisquer $x, y, z \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$, tenhamos:*

P1 $\langle x, y \rangle = \langle y, x \rangle$;

P2 $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$;

P3 $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle = \langle x, \lambda y \rangle$;

P4 $x \neq 0$ *implica* $\langle x, x \rangle > 0$.

Isto significa que um produto interno é uma função real simétrica, bilinear, positiva definida, $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. Um exemplo bastante importante é o produto interno canônico do espaço euclidiano $\mathbb{R}^n$, o qual é dado por

$$\langle x, y \rangle = x_1 y_1 + \cdots + x_n y_n,$$

onde $x = (x_1, \ldots, x_n)$ *e* $y = (y_1, \ldots, y_n)$*. Este será o produto interno que utilizaremos neste trabalho.*

Dado $x \in \mathbb{R}^n$*, escrevemos* $||x|| = \sqrt{x_1^2 + \ldots + x_n^2}$*. O número* $||x||$ *chama-se a norma euclidiana ou o comprimento do vetor* $x \in \mathbb{R}^n$*.*

Teorema 1. (Desigualdade de Cauchy-Schwarz) *Para quaisquer* $x, y \in \mathbb{R}^n$*, tem-se* $|\langle x, y\rangle| \leq ||x||||y||$*. Vale a igualdade se, e somente se, um dos vetores* x*,* y *é um múltiplo escalar do outro.*

Demonstração. Veja página 5 de [1]. □

A norma euclidiana $||x|| = \sqrt{\langle x, x\rangle}$ *goza das seguintes propriedades, onde* $x, y \in \mathbb{R}^n$*,* $\alpha \in \mathbb{R}$ *e* $|\alpha|$ *significa o valor absoluto do número real* α*:*

N1 $||x + y|| \leq ||x|| + ||y||$;

N2 $||\alpha x|| = |\alpha|||x||;$*;*

N3 $x \neq 0$ *implica* $||x|| > 0.$*.*

De um modo geral, uma norma num espaço vetorial E *é qualquer função real* $||\ || : E \to \mathbb{R}$ *que cumpra as condições N1, N2 e N3 acima.*

Uma norma arbitrária $||\ ||$ *num espaço vetorial* E *pode não provir de um produto interno, isto é, nem sempre existe um produto interno* $\langle\ ,\ \rangle$ *em* E *tal que* $||x||^2 = \langle x, x\rangle$ *para todo* $x \in E$*. Se a norma provém de um produto interno, então vale a identidade do paralelogramo*

$$||x + y||^2 + ||x - y||^2 = 2(||x||^2 + ||y||^2).$$

Uma norma em $\mathbb{R}^n$ *permite que sejam definidos alguns conceitos. Uma bola de centro* $a \in \mathbb{R}^n$ *e raio* $r > 0$ *é o conjunto*

$$B(a, r) = \{x \in \mathbb{R}^n\ ;\ ||x - a|| < r\}.$$

Analogamente definimos a bola fechada $B[a, r]$ *e a esfera* $S[a, r]$ *da seguinte maneira*

$$B[a, r] = \{x \in \mathbb{R}^n\ ;\ ||x - a|| \leq r\} \text{ e } S[a, r] = \{x \in \mathbb{R}^n\ ;\ ||x - a|| = r\}.$$

A seguir definiremos sequências no espaço euclidiano n-dimensional $\mathbb{R}^n$*.*

Definição 2. *Uma sequência em $\mathbb{R}^n$ é uma aplicação $x : \mathbb{N} \to \mathbb{R}^n$ definida nos naturais $\mathbb{N}$. O valor que essa aplicação assume no número k é indicado por x^k e denominado k-ésimo termo da sequência. Usaremos a notação $\{x^k\}$ para indicar a sequência cujo k-ésimo termo é $x^k \in \mathbb{R}^n$.*

Uma subsequência de $\{x^k\}$ é a restrição da sequência a um subconjunto infinito $\mathbb{N}' = \{k_1 < k_2 < \cdots < k_j < \cdots\}$. Uma subsequência será denotada por $\{x^{k_j}\}$.

Diz-se que uma sequência $\{x^k\}$ é limitada, quando existe $c > 0$ tal que $||x^k|| \leq c$, para todo $k \in \mathbb{N}$. Um ponto $a \in \mathbb{R}^n$ é limite de uma sequência $\{x^k\}$ quando, para todo $\epsilon > 0$ dado, é possível obter $k_0 \in \mathbb{N}$ tal que $k > k_0$ implica que $||x^k - a|| < \epsilon$. Também dizemos que $\{x^k\}$ converge para a e denotamos por $\lim\limits_{k\to\infty} x^k = a$. Caso contrário, diz-se que $\{x^k\}$ é divergente.

Diz-se que um ponto $a \in \mathbb{R}$ é valor de aderência de uma sequência $\{x^k\}$ quando existe uma subsequência de $\{x^k\}$ convergente para a.

Definição 3. *Uma aplicação $f : K \to \mathbb{R}$ definida no conjunto $K \subseteq \mathbb{R}^n$ é uma função contínua no ponto $a \in K$ quando,*

$$\forall \epsilon > 0 \;\; \exists \delta > 0; \;\; x \in K, \;\; ||x - a|| < \delta \Rightarrow |f(x) - f(a)| < \epsilon.$$

Diz-se que $f : K \to \mathbb{R}$ satisfaz a condição de $H\ddot{o}lder$ com constante $L > 0$ e grau β no ponto $x \in K \subseteq \mathbb{R}^n$ se existe um número $L \in \mathbb{R}$ tal que

$$|f(y) - f(x)| \leq L||y - x||^\beta \;\; \forall y \in K.$$

Em particular, quando $\beta = 1$ dizemos que f é Lipschitz. Obviamente, se f satisfaz a condição de $H\ddot{o}lder$, então f é contínua.

As funções contínuas tem a seguinte propriedade:

Proposição 1. *Uma aplicação $f : K \to \mathbb{R}$, definida no subconjunto $K \subseteq \mathbb{R}^n$, é contínua no ponto $a \in K$ se, e somente se, para toda sequência de pontos $\{x^k\} \subset K$, com $\lim\limits_{k\to\infty} x^k = a$, tem-se $\lim\limits_{k\to\infty} f(x^k) = f(a)$.*

Demonstração. Veja página 27 de [1]

1.2 Conceitos e resultados de Otimização

Por se tratar de um assunto extenso, trataremos apenas dos conceitos e resultados que serão utilizados no decorrer deste trabalho. Definiremos alguns conceitos introdutórios de otimização, como pontos estacionários, minimizadores locais e globais e alguns resultados básicos.

Consideremos um conjunto $K \subset \mathbb{R}^n$ e uma função $f : K \to \mathbb{R}$. Nosso objetivo será encontrar o minimizador de f em K. Este problema normalmente é exposto da seguinte maneira:

$$\min f(x), \quad x \in K, \tag{1.1}$$

onde K é chamado conjunto viável e f função objetivo, ou função custo, do problema (1.1).

Definição 4. *Seja a função $f : K \to \mathbb{R}$. Dizemos que $\overline{x} \in K$ é minimizador local do problema de (1.1), se existir $\delta > 0$, tal que*

$$f(\overline{x}) \leq f(x), \qquad \forall x \in B(\overline{x}, \delta). \tag{1.2}$$

E dizemos que $\overline{x}$ é minimizador global do problema de (1.1) se

$$f(\overline{x}) \leq f(x), \quad \forall x \in K. \tag{1.3}$$

Se as desigualdades (1.2) e (1.3) forem estritas, então $\overline{x}$ será chamado, respectivamente, de minimizador estrito local e minimizador estrito global.

Definição 5. *Dizemos que $\overline{v} \in [-\infty, +\infty)$ definido por*

$$\overline{v} = \inf_{x \in K} f(x)$$

é o valor ótimo do problema (1.1). O valor ótimo também é denotado por f^*.

Uma função pode admitir vários minimizadores globais, mas o valor ótimo do problema é sempre o mesmo.

Existem alguns critérios que garantem a existência de solução global para o problema (1.1). Um critério bastante conhecido é o seguinte teorema.

Teorema 2. (Teorema de Weierstrass) *Sejam $K \subset \mathbb{R}^n$ um conjunto compacto e não vazio e $f : K \to \mathbb{R}$ uma função contínua. Então, o problema (1.1) tem solução global.*

Demonstração. Veja página 7 de [2]. □

Quando, no problema (1.1), consideramos o conjunto K como sendo o próprio $\mathbb{R}^n$, então dizemos que temos um problema irrestrito e apresentamos este problema da seguinte forma:

$$\min f(x), \quad x \in \mathbb{R}^n. \tag{1.4}$$

Definição 6. *Dizemos que um ponto* $\overline{x} \in \mathbb{R}^n$ *é um ponto estacionário ou crítico para problema* (1.4)*, quando*

$$\nabla f(\overline{x}) = 0. \tag{1.5}$$

A definição de ponto estacionário é importante, pois estes pontos são candidatos naturais a minimizadores do problema (1.4).

Teorema 3. (Condição necessária de primeira ordem) *Suponhamos que* $f : \mathbb{R}^n \to \mathbb{R}$ *seja diferenciável no ponto* $\overline{x} \in \mathbb{R}^n$*. Se* $\overline{x}$ *é um minimizador local do problema* (1.4)*, então*

$$\nabla f(\overline{x}) = 0.$$

Demonstração. Veja página 15 de [2]. □

É importante ressaltar que o resultado anterior também é válido quando tratamos de problemas restritos a um conjunto viável K*, desde que* $\overline{x}$ *esteja contido no interior do conjunto* K.

Portanto, se f *é diferenciável, as soluções locais do problema* (1.4) *devem ser pontos estacionários. A seguir, daremos algumas condições de segunda ordem.*

Teorema 4. (Condição necessária de segunda ordem) *Suponhamos que* $f : \mathbb{R}^n \to \mathbb{R}$ *seja duas vezes diferenciável no ponto* $\overline{x} \in \mathbb{R}^n$*. Se* $\overline{x}$ *é um minimizador local do problema* (1.4)*, então vale* (1.5) *e a matriz Hessiana de* f *no ponto* $\overline{x}$ *é semidefinida positiva, isto é,*

$$\langle \nabla^2 f(\overline{x})d, d\rangle \geq 0, \quad \forall d \in \mathbb{R}^n.$$

Demonstração. Veja página 16 de [2]. □

As condições no teorema acima não são suficientes para que um ponto $\overline{x}$ *seja minimizador, conforme exemplo abaixo.*

Exemplo 1. *Seja* $f : \mathbb{R} \to \mathbb{R}$*,* $f(x) = x^3$*. Essa função satisfaz a desigualdade do teorema acima no ponto* $\overline{x} = 0$*, mas este ponto não é minimizador local do problema* (1.4) *(veja figura 1.1).*

Teorema 5. (Condição suficiente de segunda ordem) *Suponhamos que* $f : \mathbb{R}^n \to \mathbb{R}$ *seja duas vezes diferenciável no ponto* $\overline{x} \in \mathbb{R}^n$*. Se* $\overline{x}$ *é um ponto estacionário e se a matriz Hessiana de* f *em* $\overline{x}$ *é definida positiva, isto é, se*

$$\langle \nabla^2 f(\overline{x})d, d\rangle > 0, \quad \forall d \in \mathbb{R}^n - \{0\},$$

então $\overline{x}$ *é minimizador local estrito do problema* (1.4).

Demonstração. Veja página 17 de [2]. □

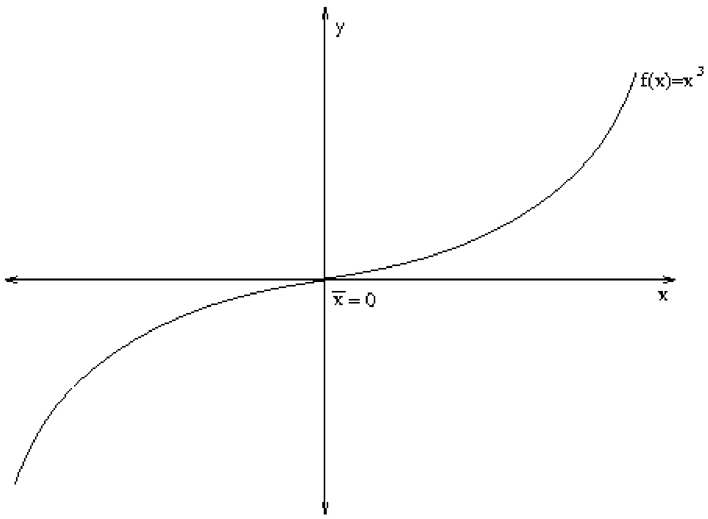

Figura 1.1: O ponto $\overline{x} = 0$ não é minimizador local de f.

No teorema acima, as condições não são necessárias para que um ponto $\overline{x}$ seja minimizador, conforme exemplo que segue.

Exemplo 2. *Seja $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^4$. Esta função não satisfaz a segunda condição do teorema em $\overline{x} = 0$, mas este ponto é minimizador do problema (1.4).*

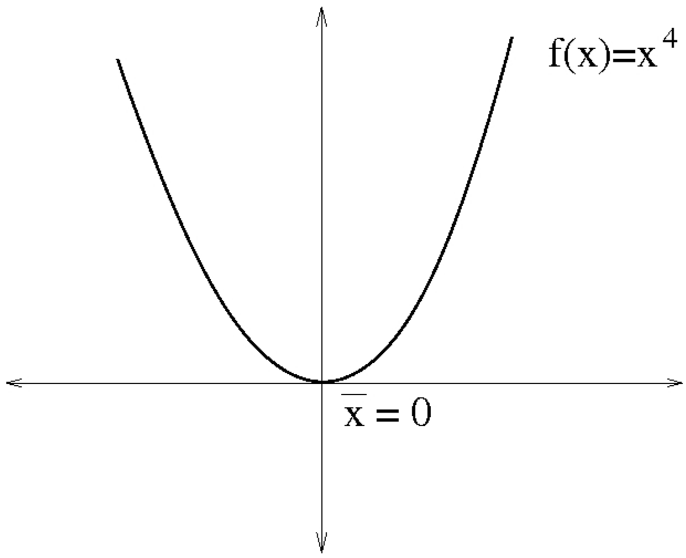

Figura 1.2: O ponto $\overline{x} = 0$ é minimizador global de f.

1.3 Conceitos e resultados de Análise Convexa

Nesta seção, veremos alguns conceitos e resultados sobre conjuntos convexos, funções convexas (e algumas de suas caracterizações), funções quase-convexas e pseudo-convexas, (com propriedades) projeção de um ponto a um conjunto e os teoremas de separação, que utilizaremos no capítulo seguinte.

1.3.1 Funções Convexa e Conjuntos Convexos

Iniciaremos esta subseção com o conceito de conjunto convexo, que é caracterizado por conter todos os segmento de retas cujos extremos pertecem ao conjunto.

Definição 7. *Um conjunto $K \subset \mathbb{R}^n$ é convexo se, para quaisquer $x, y \in K$ e $\lambda \in [0,1]$, tem-se que $\lambda x + (1-\lambda)y \in K$.*

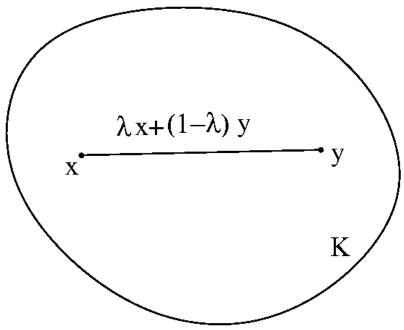

Figura 1.3: Conjunto convexo.

Exemplo 3. *Um hiperplano do $\mathbb{R}^n$ é um conjunto da forma $H(a,c) = \{x \in \mathbb{R}^n; \langle a, x\rangle = c\}$, onde $a \in \mathbb{R}^n$ e $c \in \mathbb{R}$. Dados $x, y \in H(a,c)$ e $\lambda \in [0,1]$, temos*

$$\langle a, \lambda x + (1-\lambda)y\rangle = \lambda\langle a, x\rangle + (1-\lambda)\langle a, y\rangle = \lambda c + (1-\lambda)c = c$$

Ou seja, $\lambda x + (1-\lambda)y \in H(a,c)$. Logo,$H(a,c)$ é um conjunto convexo.

Definição 8. *Dados $x^j \in \mathbb{R}^n$, $\lambda_j \geq 0$, $j = 1, ..., k$, com $\sum_{j=1}^{k} \lambda_j = 1$. O ponto $\sum_{i=j}^{k} \lambda_j x^j$ chama-se combinação convexa dos pontos x^j, com parâmetros λ_j.*

Proposição 2. *Um conjunto $K \subset \mathbb{R}^n$ é convexo se, e somente se, para quaisquer $p \in \mathbb{N}$, $x^j \in K$ e $\lambda_j \in [0,1]$, $j = 1, ..., k$, tais que $\sum_{j=1}^{k} \lambda_j = 1$, a combinação convexa $\sum_{j=1}^{k} \lambda_j x^j$ pertença a K.*

Demonstração. Veja página 75 de [2]. □

Definição 9. *Se $K \subset \mathbb{R}^n$ é um conjunto convexo, dizemos que a função $f : K \to \mathbb{R}$ é convexa em K quando, para quaisquer $x, y \in K$ e $\lambda \in [0,1]$, tem-se*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y).$$

Diz-se que f é estritamente convexa quando a desigualdade acima é estrita para todo $x \neq y$ e $\lambda \in (0,1)$.

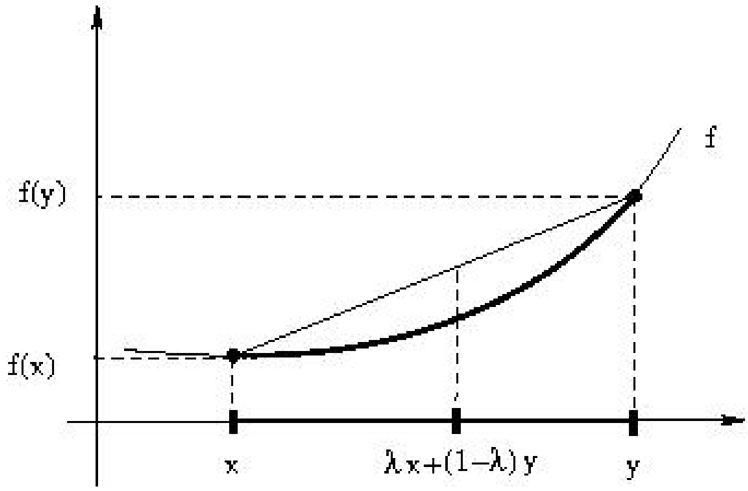

Figura 1.4: Função convexa

Definição 10. *O epígrafo de uma função $f : K \to \mathbb{R}$ é o conjunto*

$$E_f = \{(x,c) \in S \times \mathbb{R}\ ;\ f(x) \leq c\}.$$

A proposição abaixo estabelece uma relação entre a convexidade de uma função e seu epígrafo.

Proposição 3. *Seja $K \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : K \to \mathbb{R}$ é convexa em K se, e somente se, o epígrafo de f é um conjunto convexo em $\mathbb{R}^n \times \mathbb{R}$.*

Demonstração. Veja página 67 de [2]. □

Em otimização é comum a utilização de funções que assumem o "valor"∞, onde $\mathbb{R} \cup \{\infty\}$ é denominado reta estendida. A seguir veremos um exemplo de função deste tipo.

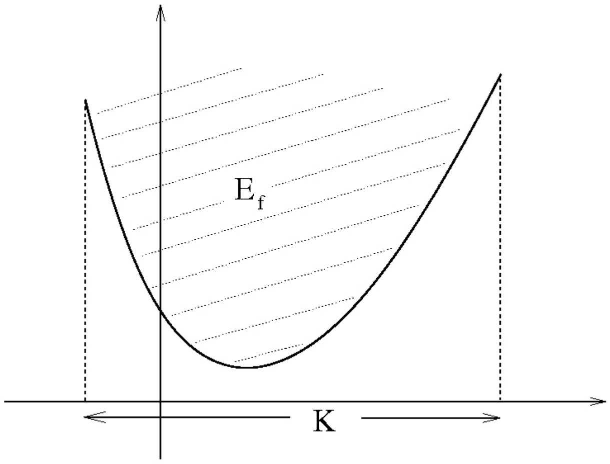

Figura 1.5: Função convexa.

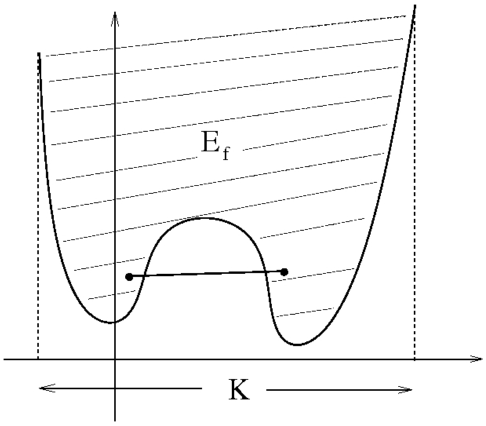

Figura 1.6: A função não é convexa.

Exemplo 4. *Seja $K \subset \mathbb{R}^n$ um conjunto convexo. A função indicadora de K, $\delta_K : K \to \mathbb{R} \cup \{\infty\}$ definida por*

$$\delta_K(x) = \begin{cases} 0, & \textit{se} \quad x \in K \\ \infty, & \textit{se} \quad x \notin K \end{cases}$$

é convexa porque $E_{\delta_K} = K \times [0, \infty)$.

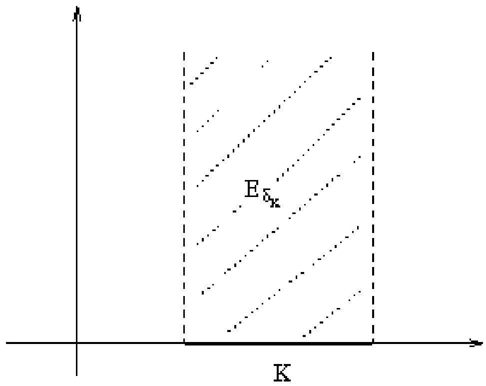

Figura 1.7: Epígrafo de δ_K

Definiremos agora, conjunto de nível de uma função e daremos uma condição necessária para que uma função real seja convexa em um conjunto convexo $K \subset \mathbb{R}^n$.

Definição 11. *O conjunto de nível de uma função $f : K \to \mathbb{R}$ associado a um ponto $c \in \mathbb{R}$ é dado por*

$$L_{f,K}(c) = \{x \in K \ ; \ f(x) \leq c\}.$$

Proposição 4. *Seja f uma função real definida no convexo $K \subset \mathbb{R}^n$. Então $L_{f,K}(c)$ é convexo, para todo $c \in \mathbb{R}$.*

Demonstração. Veja página 133 de [2]. □

A convexidade de todos os conjuntos de níveis de uma uma função não garante a sua convexidade, conforme exemplo abaixo.

Exemplo 5. *Seja $f : \mathbb{R} \to \mathbb{R}$, com $f(x) = x^3$. A função f tem todos os seus conjuntos de níveis convexos, mas f não é convexa (basta tomar $x = -1$, $y = 0$ e $\lambda = 1/2$ e verificar que $f(\lambda x + (1-\lambda)y > \lambda f(x) + (1-\lambda)f(y))$.*

Na subseção seguinte, veremos uma classe de funções onde vale esta propriedade. Essas funções serão chamadas de quase-convexas.

Dizemos que (1.1) é um problema de minimização convexa, quando $K \subset \mathbb{R}^n$ é um conjunto convexo e $f : K \to \mathbb{R}$ é uma função convexa em K. A importância da convexidade já pode ser vista no seguinte resultado.

Teorema 6. (Teorema de minimização convexa) *Sejam $K \subset \mathbb{R}^n$ um conjunto convexo e $f : K \to \mathbb{R}$ é uma função convexa em K. Então, todo minimizador local do problema (1.1) é global. Além disso, o conjunto de minimizadores é convexo. Se f é estritamente convexa, não pode haver mais de um minimizador.*

Demonstração. Suponhamos que $\overline{x} \in S$ seja um minimizador local e não global. Então, existe $\delta > 0$, tal que $f(\overline{x}) \leq f(x)$ para todo $x \in B(\overline{x}, \delta)$, e $\hat{x} \in S$, tal que $f(\hat{x}) < f(\overline{x})$.
Defina $x_\lambda = \lambda\hat{x} + (1-\lambda)\overline{x}$. Pela convexidade de S, $x_\lambda \in S$ para todo $\lambda \in [0,1]$. Agora, pela convexidade da f, para todo $\lambda \in (0,1]$, temos

$$f(x_\lambda) \leq \lambda f(\hat{x}) + (1-\lambda)f(\overline{x}) = f(\overline{x}) + \lambda(f(\hat{x}) - f(\overline{x})) < f(\overline{x}).$$

Dessa forma, podemos tomar λ suficientemente pequeno, de modo que $x_\lambda \in B(\overline{x}, \delta)$, e ainda temos que $f(x_\lambda) < f(\overline{x})$. Isto contradiz o fato de que $\overline{x}$ é minimizador local do problema (1.1). Portanto, qualquer solução local deve ser global.

Seja $S^* \subset S$ o conjunto dos minimizadores (globais) e $\overline{v} \in \mathbb{R}$ o valor ótimo do problema. Para quaisquer $x, y \in S^*$ e $\lambda \in [0,1]$, pela convexidade da f, temos

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

isto é,

$$f(\lambda x + (1-\lambda)y) \leq \lambda\overline{v} + (1-\lambda)\overline{v} = \overline{v}.$$

Como $\overline{x}$ é minimizador global, temos

$$f(\lambda x + (1-\lambda)y) = \overline{v}.$$

Então, $\lambda x + (1-\lambda)y \in S^*$. Logo S^* é convexo.

Suponhamos agora que f seja estritamente convexa e que existam $x, \overline{x} \in S^*$, com $x \neq \overline{x}$. Como x e $\overline{x}$ são minimizadores globais e $\lambda x + (1-\lambda)\overline{x} \in S$, para $\lambda \in (0,1)$, pela convexidade de S, segue que

$$f(\lambda x + (1-\lambda)\overline{x}) \geq f(x) = f(\overline{x}) = \overline{v}.$$

Por outro lado, pela convexidade estrita da f,

$$f(\lambda x + (1-\lambda)\overline{x}) < \lambda f(x) + (1-\lambda)f(\overline{x}) = \lambda\overline{v} + (1-\lambda)\overline{v} = \overline{v},$$

que resulta numa contradição. Portanto, o minimizador é único. □

A seguir, daremos uma condição necessária e suficiente para que um problema de minimização convexa tenha solução.

Teorema 7. (Condição necessária e suficiente para problema de minimização convexa) *Sejam $K \subset \mathbb{R}^n$ um conjunto convexo e $f : J \to \mathbb{R}$ uma função convexa e diferenciável no aberto J que contém K. Então $\overline{x}$ é um minimizador de f em K se, e somente se,*

$$\langle \nabla f(\overline{x}), x - \overline{x} \rangle \geq 0, \quad \forall x \in K.$$

Demonstração. Veja página 151 de [2]. □

Quando uma função é difereciável, existem caracterizações que são úteis para determinar se uma função é convexa ou não, conforme veremos a seguir.

Teorema 8. (Caracterização de Funções Convexas) *Sejam $K \subset \mathbb{R}^n$ um conjunto convexo e aberto e $f : K \to \mathbb{R}$ uma função diferenciável em K. Então as afirmações abaixo são equivalentes:*

1) *A função f é convexa em K.*

2) *Para todo $x, y \in K$,*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle.$$

3) *Para todo $x, y \in K$,*

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

Quando f é duas vezes diferenciável em K, as propriedades acima também são equivalentes a:

4) *A matriz Hessiana de f é semidefinida positiva em todo ponto de K:*

$$\langle \nabla^2 f(x)d, d \rangle \geq 0 \ \ \forall x \in K, \ \ \forall d \in \mathbb{R}^n.$$

Demonstração. $1) \Rightarrow 2)$ Suponha que f seja convexa, então para $x, y \in K$ e $\lambda \in (0, 1]$, tem-se que

$$f(\lambda y + (1-\lambda)x) \leq \lambda f(y) + (1-\lambda) f(x).$$

Considerando $d = y - x$, temos $x + \lambda d = x + \lambda(y - x) = (1-\lambda)x + \lambda y$. Pela convexidade da f, segue que $f(x + \lambda d) \leq \lambda f(y) + (1-\lambda) f(x)$. Então,

$$\lambda[f(y) - f(x)] \geq f(x + \lambda d) - f(x).$$

Isto implica que

$$f(y) - f(x) \geq \frac{f(x + \lambda d) - f(x)}{\lambda},$$

pois $\lambda > 0$.

Daí, passando o limite quando $\lambda \to 0^+$, obtemos

$$f(y) - f(x) \geq \lim_{\lambda \to 0^+} \frac{f(x + \lambda d) - f(x)}{\lambda} = \frac{\partial f}{\partial d}(x) = \langle \nabla f(x), d \rangle = \langle \nabla f(x), y - x \rangle.$$

Logo, para todo $x, y \in K$

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle. \tag{1.6}$$

$2) \Rightarrow 3)$ Trocando x por y em (1.6), obtemos

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \tag{1.7}$$

Somando (1.6) e (1.7), tem-se

$$\begin{aligned}
\langle \nabla f(x), y - x \rangle + \langle \nabla f(y), x - y \rangle &\leq 0 \\
\langle \nabla f(x), y - x \rangle - \langle \nabla f(y), y - x \rangle &\leq 0 \\
\langle \nabla f(x) - \nabla f(y), y - x \rangle &\leq 0 \\
\langle \nabla f(y) - \nabla f(x), y - x \rangle &\geq 0, \ \forall x, y \in K.
\end{aligned}$$

$3) \Rightarrow 2)$ Sejam $x, y \in K$. Pelo teorema do valor médio, existe $\lambda \in (0,1)$ tal que

$$f(y) - f(x) = \langle \nabla f(x + \lambda d), d \rangle \tag{1.8}$$

onde $d = y - x$.

Aplicando 3) para $x + \lambda d$ e x, obtemos

$$0 \leq \langle \nabla f(x + \lambda d) - \nabla f(x), x + \lambda d - x \rangle = \langle \nabla f(x + \lambda d) - \nabla f(x), \lambda d \rangle.$$

Então

$$\langle \nabla f(x + \lambda d), \lambda d \rangle \geq \langle \nabla f(x), \lambda d \rangle, \text{ que implica em } \langle \nabla f(x + \lambda d), d \rangle \geq \langle \nabla f(x), d \rangle. \tag{1.9}$$

Combinando (1.8) e (1.9) e usando $d = y - x$, obtemos

$$f(y) - f(x) \geq \langle \nabla f(x), d \rangle \Rightarrow f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle.$$

Logo, para todo $x, y \in K$

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle.$$

$2) \Rightarrow 1)$ Aplicando 2) para x e $x + \lambda d$, com $d = y - x$, segue que

$$f(x) \geq f(x + \lambda d) + \langle \nabla f(x + \lambda d), x - (x + \lambda d) \rangle = f(x + \lambda d) + \langle \nabla f(x + \lambda d), -\lambda d \rangle.$$

Portanto,

$$f(x) \geq f(x + \lambda d) - \lambda \langle \nabla f(x + \lambda d), d \rangle.$$

Daí,

$$(1 - \lambda) f(x) \geq (1 - \lambda) f(x + \lambda d) - (1 - \lambda)\lambda \langle \nabla f(x + \lambda d), d) \rangle. \tag{1.10}$$

Analogamente, aplicando 2) para y e $x + \lambda d$, obtemos

$$\lambda f(y) \geq \lambda f(x + \lambda d) + \lambda(1 - \lambda) \langle \nabla f(x + \lambda d), d) \rangle. \tag{1.11}$$

Somando (1.10) e (1.11), tem-se

$$(1 - \lambda) f(x) + \lambda f(y) \geq (1 - \lambda) f(x + \lambda d) + \lambda f(x + \lambda d) = f(x + \lambda d). \tag{1.12}$$

Subtituindo $d = y - x$ em (1.12), temos

$$(1 - \lambda) f(x) + \lambda f(y) \geq f(x + \lambda(y - x)) = f(x + \lambda y - \lambda x) = f((1 - \lambda)x + \lambda y).$$

Logo f é convexa em K.

Suponhamos agora que f é duas vezes diferenciável em K. É sufuciente mostrar que 2) $\Leftrightarrow$ 4). Fixemos $x \in K$ e $d \in \mathbb{R}^n$ quaisquer. Como K é aberto, $x + td \in K$ para $t > 0$ sufucientemente pequeno. De 2),

$$f(x + td) - f(x) \geq t\langle \nabla f(x), d\rangle.$$

Usando a diferenciabilidade da f,

$$0 \leq f(x + td) - f(x) - t\langle \nabla f(x), d\rangle = \frac{t^2}{2}\langle \nabla^2 f(x)d, d\rangle + o(t^2).$$

Dividindo por $t^2 > 0$ e tomando o limite quando $t \to 0^+$, obtemos 4).

Agora, sejam $x, y \in K$ quaisquer. Pelo Teorema do Valor Médio, existe $\theta \in (0, 1)$ tal que

$$f(y) - f(x) - \langle \nabla f(x), y - x\rangle = \frac{1}{2}\langle \nabla^2 f(x + \theta(y - x))(y - x), y - x\rangle \geq 0,$$

onde a desigualdade em 2) segue de 4). $\square$

Uma outra importante propriedade das funções convexas será provada no seguinte teorema.

Teorema 9. *Sejam $K \subset \mathbb{R}^n$ um conjunto convexo, aberto e $f : K \longrightarrow \mathbb{R}$ uma função convexa. Então f é localmente Lipschitz.*

Demonstração. Seja $\overline{x} \in K$. Como K é aberto, existe $\delta > 0$ (que depende de $\overline{x}$) tal que $U \subset K$, onde $U = \{x \in \mathbb{R}^n; -\delta \leq x_i - \overline{x}_i \leq \delta, \ \ i = 1, \ldots, n\}$. Chamaremos de V o conjunto de vértices de U. Assim V está composto por 2^n pontos e $U = conv(V)$.

Seja $\beta = \max_{x \in V} f(x)$. Pela convexidade de f, segue que $L_{f,K}(\beta)$ é convexo. Como $V \subset L_{f,K}(\beta)$, segue que

$$U = conv(V) \subset conv(L_{f,K}(\beta)) = L_{f,K}(\beta)$$

que implica que $f(x) \leq \beta$, $\forall x \in U$.

Seja $x \in U$, tal que $0 < ||x - \overline{x}|| < \delta$. Defina

$$d := \frac{\delta(x - \overline{x})}{||x - \overline{x}||} = \frac{x - \overline{x}}{\alpha}, \quad \alpha = \frac{||x - \overline{x}||}{\delta} \in (0, 1).$$

Daí, $|d_i| < \delta$, $\forall i = 1, \ldots, n$, donde segue que $-\delta \leq \overline{x}_i + d_i - \overline{x}_i \leq \delta, \ \ \forall i = 1, \ldots, n$.

Assim, $\overline{x} + d \in U$ e $\overline{x} - d \in U$. Além disso,

$$x = \overline{x} + \frac{\alpha(x - \overline{x})}{\alpha} = \overline{x} + \alpha d = \alpha\overline{x} + \overline{x} + \alpha d - \alpha\overline{x} = \alpha(\overline{x} + d) + (1 - \alpha\overline{x}),$$

logo

$$f(x) = f(\alpha(\overline{x}+d) + (1-\alpha)\overline{x}) \leq \alpha f(\overline{x}+d) + (1-\alpha)f(\overline{x}) \leq \alpha\beta + (1-\alpha)f(\overline{x}) \quad (1.13)$$

Analogamente, obtemos

$$f(\overline{x}) \leq \frac{f(x)+\alpha\beta}{1+\alpha} \quad (1.14)$$

De (1.13) segue que $f(x) \leq \alpha\beta + (1-\alpha)f(\overline{x})$, então

$$f(x) - f(\overline{x}) \leq \alpha(\beta - f(\overline{x})) \quad (1.15)$$

De (1.14) segue que $f(\overline{x}) + \alpha f(\overline{x}) \leq f(x) + \alpha\beta$, que implica que $-\alpha\beta + \alpha f(\overline{x}) \leq f(x) - f(\overline{x})$. Logo

$$-\alpha(\beta - f(\overline{x})) \leq f(x) - f(\overline{x}) \quad (1.16)$$

De (1.15) e (1.16) segue que

$$|f(x) - f(\overline{x})| \leq \alpha(\beta - f(\overline{x})) = \frac{||x-\overline{x}||}{\delta}(\beta - f(\overline{x})),$$

portanto

$$|f(x) - f(\overline{x})| \leq \delta^{-1}(\beta - f(\overline{x}))||x-\overline{x}|| = M||x-\overline{x}||$$

que prova o teorema. □

Conforme vimos no primeiro capítulo, se f for Lipschitz, então f é contínua. Em particular, no teorema acima, f é contínua.

No teorema acima, caso o domínio da função não seja aberto, garantimos que f é contínua no interior de seu domínio.

Exemplo 6. *Sejam $K = \{x \in \mathbb{R} \; ; \; x \geq -1\}$ e $f : K \to \mathbb{R}$,*

$$f(x) = \begin{cases} 2, & \text{se} \quad x = -1 \\ x^2, & \text{se} \quad x > -1. \end{cases}$$

É fácil ver que f é convexa em K (seu epígrafo é convexo), mas f não é contínua em $x = -1$ (na fronteira de seu domínio).

Definição 12. *Seja $K \subset \mathbb{R}^n$ um conjunto convexo. Dizemos que $f : K \to \mathbb{R}$ é fortemente convexa, com módulo $\alpha > 0$ em K, quando para quaisquer $x, y \in K$ e $\lambda \in [0,1]$, tem-se*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \alpha\lambda(1-\lambda)||x-y||^2.$$

Obviamente, uma função fortemente convexa é estritamente convexa, e uma função estritamente convexa é convexa.

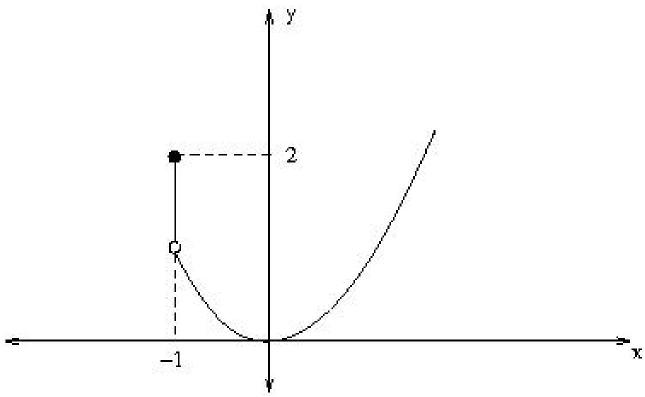

Figura 1.8: f é convexa e descontínua em $x = -1$.

Exemplo 7. *Seja* $f : \mathbb{R} \to \mathbb{R}$, *com* $f(x) = x^2$, *é fortemente convexa.*

Exemplo 8. *Seja* $f : \mathbb{R} \to \mathbb{R}$, *com* $f(x) = e^x$, *é estritamente (mas não fortemente) convexa.*

Exemplo 9. *Seja* $f : \mathbb{R} \to \mathbb{R}$, *com* $f(x) = x$, *é convexa (mas não estritamente convexa).*

O seguinte resultado tem grande importância do ponto de vista computacional, porque, com frequência, métodos numéricos são baseados na resolução de uma sequência de subproblemas cuja função objetivo é fortemente convexa. Como consequência do próximo teorema, veremos que o problema de minimização de uma função fortemente convexa, num conjunto fechado e não vazio, tem sempre solução única.

Teorema 10. (Compacidade de conjunto de nível de funções fortemente convexas) *Suponhamos que a função* $f : \mathbb{R}^n \to \mathbb{R}$ *seja fortemente convexa em* $\mathbb{R}^n$. *Então, o conjunto de nível*

$$L_{f,\mathbb{R}^n}(c) = \{x \in \mathbb{R}^n \ ; \ f(x) \leq c\}$$

é compacto para todo $c \in \mathbb{R}$.

Demonstração. Veja página 139 de [2]. □

Corolário 1. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função fortemente convexa e* $K \subset \mathbb{R}^n$ *um subconjunto não vazio qualquer. Então* f *tem um único minimizador em* K.

Demonstração. Veja página 140 de [2]. □

Para funções diferenciáveis fortemente convexas, temos critérios que são análogos aos de funções convexas.

Teorema 11. (Caracterização de funções fortemente convexas) *Sejam $K \subset \mathbb{R}^n$ um conjunto convexo, aberto e $f : K \to \mathbb{R}$ uma função diferenciável em K. Então as afirmações abaixo são equivalentes:*

1) *A função f é fortemente convexa em K com módulo $\alpha > 0$.*

2) *Para todo $x, y \in K$,*

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

3) *Para todo $x, y \in K$,*

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 2\alpha ||y - x||^2.$$

Quando f é duas vezes diferenciável em K, as propriedades acima também são equivalentes a:

4) *A matriz Hessiana de f é definida positiva uniformemente em todo ponto de K, isto é,*

$$\langle \nabla^2 f(x) d, d \rangle \geq 2\alpha ||y - x||^2, \quad \forall x \in K, \quad \forall d \in \mathbb{R}^n.$$

Demonstração. Veja página 156 de [2]. □

1.3.2 Funções Quase-convexas e Pseudo-convexas

Definição 13. *Seja $f : K \longrightarrow \mathbb{R}$, onde $K \subset \mathbb{R}^n$ é não vazio e convexo. A função f é dita quase-convexa se, para todo $x, y \in K$*

$$f(\lambda x + (1 - \lambda) y) \leq \max\{f(x), f(y)\}, \quad \lambda \in (0, 1).$$

Sabemos que funções convexas são caracterizadas pela convexidade dos seus epígrafos. Veremos agora que as funções quase-convexas são caracterizadas pela convexidade dos seus conjuntos de níveis.

Teorema 12. *Seja $f : K \longrightarrow \mathbb{R}$, onde $K \subset \mathbb{R}^n$ é não vazio e convexo. A função f é quase-convexa se, e somente se, $L_{f,K}(c) = \{x \in S; f(x) \leq c\}$ é convexo para todo $c \in \mathbb{R}$.*

Demonstração. Suponha que f é quase-convexa e tome $x, y \in L_{f,K}(c)$. Assim, $x, y \in K$ e $\max\{f(x), f(y)\} \leq c$. Considere $z = \lambda x + (1-\lambda)y$, com $\lambda \in (0,1)$. Pela convexidade de K, $z \in K$. Pela quase-convesidade de f temos $f(z) \leq \max\{f(x), f(y)\} \leq c$. Logo, $z \in L_{f,K}(c)$, que implica em $L_{f,K}(c)$ convexo.

Agora, suponha que $L_{f,K}(c)$ é convexo para cada $c \in \mathbb{R}$ e tome $x, y \in K$. Considere $z = \lambda x + (1-\lambda)y$, com $\lambda \in (0,1)$. Note que $x, y \in L_{f,K}(\bar{c})$, para $\bar{c} = \max\{f(x), f(y)\}$. Por hipótese, $L_{f,K}(\bar{c})$ é convexo, então $z \in L_{f,K}(\bar{c})$. Dessa forma, $f(z) \leq \bar{c} = \max\{f(x), f(y)\}$. Portanto f é quase-convexa. $\square$

O seguinte teorema nos dá uma condição necessária e sufuciente para que funções diferenciáveis sejam quase-convexas.

Teorema 13. *Sejam K um conjunto aberto, convexo e não vazio em $\mathbb{R}^n$ e $f : K \to \mathbb{R}$ diferenciável em K. Então, f é quase-convexa se, e somente se, $x, y \in K$ e $f(x) \leq f(y)$, implicar que $\langle \nabla f(y), x - y \rangle \leq 0$.*

Demonstração. Sejam f quase-convexa e $x, y \in K$ tal que $f(x) \leq f(y)$. Sendo K convexo, $\lambda x + (1-\lambda)y \in K$, com $\lambda \in (0,1)$. Pela diferenciabilidade da f, temos

$$f(\lambda x+(1-\lambda)y)-f(y) = f(y+\lambda(x-y))-f(y) = \lambda\langle \nabla f(y), x-y\rangle + \lambda||x-y||\rho(\lambda(x-y)),$$

onde $\rho(\lambda(x-y)) \to 0$ quando $\lambda \to 0$. Sendo f quase-convexa, temos $f(\lambda x+(1-\lambda)y) \leq f(y)$, isto é, $f(\lambda x + (1-\lambda)y) - f(y) \leq 0$. Então,

$$\lambda\langle \nabla f(y), x-y\rangle + \lambda||x-y||\rho(\lambda(x-y)) \leq 0.$$

Logo,

$$\langle \nabla f(y), x-y\rangle + ||x-y||\rho(\lambda(x-y)) \leq 0.$$

Fazendo , $\lambda \to 0$ na desigualdade acima, obtemos o resultado.

Reciprocamente, suponhamos $x, y \in K$ tal que $f(x) \leq f(y)$. Mostraremos que $f(\lambda x + (1-\lambda)y) \leq f(y)$ para cada $\lambda \in (0,1)$. Faremos isso mostrando que

$$L = \{z \in K \; ; \; z = \lambda x + (1-\lambda)y, \; \lambda \in (0,1) \text{ e } f(z) > f(y)\}$$

é vazio. Por contradição, suponha que existe $x' \in L$. Assim $x' = \lambda x + (1-\lambda)y$ para algum $\lambda \in (0,1)$ e $f(x') > f(y)$. Sendo f diferenciável, f é contínua, ou seja, existe $\delta \in (0,1)$ tal que

$$f(\mu x' + (1-\mu)y) > f(y) \;\text{ para todo }\; \mu \in [\delta, 1] \tag{1.17}$$

e, sendo $\delta > 0$, $f(x') > f(\delta x' + (1-\delta)y)$, ou seja, $f(x') - f(\delta x' + (1-\delta)y) > 0$. Aplicando o teorema do valor médio a função $\phi : [\delta, 1] \longrightarrow \mathbb{R}$ tal que $\phi(\mu) = f(\mu x' + (1-\mu)y)$, temos

$$0 < f(x') - f(\delta x' + (1-\delta)y) = (1-\delta)\langle \nabla f(\hat{x}), x' - y \rangle, \tag{1.18}$$

onde $\hat{x} = \hat{\mu}x' + (1-\hat{\mu})y$ para algum $\hat{\mu} \in (\delta, 1)$. Por (1.17) temos que $f(\hat{x}) > f(y)$. Dividindo (1.18) por $1 - \delta > 0$, segue que $\langle \nabla f(\hat{x}), x' - y \rangle > 0$ que implica em

$$\langle \nabla f(\hat{x}), x - y \rangle > 0. \tag{1.19}$$

Por outro lado, $f(\hat{x}) > f(y) \geq f(x)$, e $\hat{x}$ é a combinação convexa de x e y, isto é, $\hat{x} = \hat{\lambda}x + (1-\hat{\lambda})y$, onde $\hat{\lambda} \in (0,1)$. Pela hipótese do teorema, $\langle \nabla f(\hat{x}), x - \hat{x} \rangle \leq 0$. Logo,

$$0 \geq \langle \nabla f(\hat{x}), x - \hat{x} \rangle = (1-\hat{\lambda})\langle \nabla f(\hat{x}), x - y \rangle$$

que contradiz a inequação (1.19). Portanto L é vazio, que prova o teorema. □

Exemplo 10. *Seja $f(x) = x^3$ e suponha $f(a) \leq f(b)$, isto é, $a^3 \leq b^3$. Isso é verdade apenas quando $a \leq b$. Agora, considere $\langle \nabla f(b), a - b \rangle = 3(a-b)b^2$. Como $a \leq b$, temos $3(a-b)b^2 \leq 0$. Isso mostra, pelo teorema anterior que f é quase-convexa.*

Sabemos que toda função convexa é quase-convexa. As funções convexas, definidas num convexo, tem a propriedade de que todo minimizador local é minimizador global da função. Esta propriedade não se verifica para funções quase-convexas (veja figura 1.1).

Afim de mantermos esta propriedade, definiremos as funções estritamente quase-convexas.

Definição 14. *Seja $f : K \to \mathbb{R}$, onde $K \subset \mathbb{R}^n$ é não vazio e convexo. A função f é dita estritamente quase-convexa se, para cada $x, y \in K$, com $f(x) \neq f(y)$, tivermos*

$$f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\} \ \ \forall \lambda \in (0,1).$$

Teorema 14. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ estritamente quase-convexa. Considere o problema de minimizar $f(x)$, com $x \in K$, onde K é um subconjunto não vazio e convexo do $\mathbb{R}^n$. Se $\overline{x}$ é solução local do problema, então $\overline{x}$ é solução global deste problema.*

Demonstração. Suponha, por absurdo, que existe $\hat{x} \in K$ tal que $f(\hat{x}) < f(\overline{x})$. Pela convexidade de K, $\lambda\hat{x} + (1-\lambda)\overline{x} \in K$ para cada $\lambda \in (0,1)$. Como $\overline{x}$ é mínimo local, então $f(\overline{x}) \leq \lambda\hat{x} + (1-\lambda)\overline{x}$ para todo $\lambda \in (0,\delta)$, para algum $\delta \in (0,1)$. Sendo f estritamente quase-convexa e $f(\hat{x}) < f(\overline{x})$, temos $f(\lambda\hat{x} + (1-\lambda)\overline{x}) < f(\overline{x})$ para todo $\lambda \in (0,1)$. Isto contradiz o fato de $\overline{x}$ ser mínimo local. Portanto, $\overline{x}$ é mínimo global. □

Sabemos que as funções estritamente convexas são convexas. Este resultado não vale para funções quase-convexa, conforme exemplo abaixo.

Exemplo 11. *Seja*

$$f(x) = \begin{cases} 1, & \text{se} \quad x = 0 \\ 0, & \text{se} \quad x \neq 0 \end{cases}$$

Pela definição, f é estritamente quase-convexa. Porém, f não é quase-convexa. Tome $a = -1$ e $b = 1$. Assim, $f(a) = f(b) = 0$, mas $f(\frac{1}{2}a + \frac{1}{2}b) = f(0) = 1 > f(b)$.

Contudo, se f é semicontínua inferiormente e estritamente quase-convexa, garantimos que f é quase-convexa.

Teorema 15. *Se K é um subconjunto não vazio e convexo do $\mathbb{R}^n$ e $f : K \to \mathbb{R}$ é estritamente quase-convexa e semicontínua inferiormente, então f é quase-convexa.*

Demonstração. Sejam $x, y \in K$. Se $f(x) \neq f(y)$, então sendo f estritamente quase-convexa, temos $f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\}$ para cada $\lambda \in (0,1)$. Agora suponha que $f(x) = f(y)$. Para mostrar que f é quase-convexa, precisamos mostrar que $f(\lambda x + (1-\lambda)y) \leq f(x)$ para cada $\lambda \in (0,1)$. Por contradição, suponha que $f(\mu x + (1-\mu)y) > f(x)$ para algum $\mu \in (0,1)$. Denote por $z = \mu x + (1-\mu)y$. Como f é semicontínua inferiormente, existe $\lambda \in (0,1)$ tal que

$$f(z) > f(\lambda x + (1-\lambda)z) > f(x) = f(y) \tag{1.20}$$

Sendo f estritamente quase-convexa, $f(z) = f(\mu x + (1-\mu)y) < f(y) < f(\lambda x + (1-\lambda)z)$, contradizendo (1.20). Portanto, f é quase-convexa. □

Se f é estritamente convexa, um minimizador local é único minimizador global. Na direção deste resultado, definiremos funções fortemente quase-convexa.

Definição 15. *Seja $f : K \to \mathbb{R}$, onde $K \subset \mathbb{R}^n$ é não vazio e convexo. A função f é dita fortemente quase-convexa se, para cada $x, y \in K$ com $x \neq y$, temos*

$$f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\}$$

para cada $\lambda \in (0,1)$.

Conforme definição, toda função estritamente convexa é fortemente quase-convexa e toda função fortemente quase-convexa é estritamente quase-convexa.

Teorema 16. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ fortemente quase-convexa. Considere o problema de minimizar $f(x)$, com $x \in K$, onde K é um subconjunto não vazio e convexo do $\mathbb{R}^n$. Se $\overline{x}$ é solução local do problema, então $\overline{x}$ é única solução global deste problema.*

Demonstração. Sendo $\overline{x}$ minimizador local, existe uma $\epsilon-$vizinhança $V_\epsilon(\overline{x})$ tal que $f(\overline{x}) \leq f(x)$ para todo $x \in S \cap V_\epsilon(\overline{x})$. Suponha, por contradição, que existe $\hat{x} \neq \overline{x}$ e $f(\hat{x}) \leq f(\overline{x})$. Sendo f fortemente quase-convexa, segue que

$$f(\lambda\hat{x} + (1-\lambda)\overline{x}) < \max\{f(\hat{x}), f(\overline{x})\} = f(\overline{x})$$

para todo $\lambda \in (0,1)$. Para λ sufucientemente pequeno, $\lambda\hat{x} + (1-\lambda)\overline{x} \in S \cap V_\epsilon(\overline{x})$ e contradiz o fato de $\overline{x}$ ser mínimo local. Portanto, $\overline{x}$ é única solução global. □

A seguir, daremos uma outra condição necessária e suficiente para que uma função seja quase-convexa.

Teorema 17. *Sejam K um subconjunto não vazio, aberto e convexo do $\mathbb{R}^n$ e $f : K \to \mathbb{R}$ uma função duas vezes diferenciável em K. Se f é quase-convexa, então, para todo $x \in K$ e $y \in \mathbb{R}^n$ temos*

$$\langle y, \nabla f(x)\rangle = 0 \ \Rightarrow \langle y, \nabla^2 f(x) y\rangle \geq 0.$$

Demonstração. Considerando a função $h : [0,1] \longrightarrow \mathbb{R}$ dada por $h(t) = f(x + (y-x)t)$ a condição

$$\langle y, \nabla f(x)\rangle = 0 \ \Rightarrow \ \langle y, \nabla^2 f(x) y\rangle \geq 0.$$

equivale a

$$h'(0) = 0 \ \Rightarrow \ h''(0) \geq 0.$$

Além disso,

$$h(t) \leq \max\{h(0), h(1)\}$$

equivale a

$$f(ty + (1-t)x) \leq \max\{f(x), f(y)\}.$$

Ou seja, h é quase-convexa se, e somente se, f é quase-convexa. Logo é suficiente mostrar o caso onde h é uma função real.

Suponha que h é quase-convexa em $(0,1)$ e $h'(c) = 0$, com $c \in (0,1)$. Devemos mostrar que $h''(c) \geq 0$. Se $h'(c) = 0$ e $h''(c) < 0$, então para $\epsilon > 0$ suficientemente pequeno temos

$h(c-\epsilon) < h(c)$ e $h(c+\epsilon) < h(c)$. Assim, o conjunto de nível $N = \{x \in \mathbb{R}; h(x) \leq h(c) - \epsilon\}$ não é convexo, que contradiz o fato de h ser quase-convexa.

$\square$

1.3.3 Projeções e Teoremas de Separação

Seja K um conjunto não vazio, convexo e fechado em $\mathbb{R}^n$. Para algum $x \in \mathbb{R}^n$ fixado, considere o seguinte problema:

$$\inf\{||y-x||^2\ ;\ y \in K\}. \tag{1.21}$$

Dado $x^0 \in K$, considere o conjunto $S := \{y \in \mathbb{R}^n\ ;\ ||y-x|| \leq ||x^0 - x||\}$.

Então, (1.21) é equivalente a

$$\inf\{||y-x||\ ;\ y \in K \cap S\},$$

o qual possui solução, pois a função $y \mapsto ||y-x||$ é contínua e $K \cap S$ é compacto, pois S é compacto. Portanto, deduzimos a existência de um ponto em K que minimiza a distância de x. Veremos a seguir, que existe um único ponto que minimiza a distância de K ao ponto x.

Teorema 18. (Teorema da Projeção) *Seja $K \subset \mathbb{R}^n$ um conjunto convexo e fechado. Então, para todo $x \in \mathbb{R}^n$, a projeção de x sobre K, denotada por $P_K(x)$, existe e é único. Além disso, $\overline{x} = P_K(x)$ se, e somente se,*

$$\overline{x} \in K,\ \ \langle x - \overline{x}, y - \overline{x}\rangle \leq 0\ \forall y \in K.$$

Demonstração. Veja página 94 de [2]. $\square$

Sejam K_1 e K_2 conjuntos não-vazios em $\mathbb{R}^n$ e

$$H(a,c) = \{x \in \mathbb{R}^n; \langle a, x\rangle = c\}$$

um hiperplano, onde $a \in \mathbb{R}^n/\{0\}$, $c \in \mathbb{R}$.

Definição 16. *Dizemos que o hiperplano $H(a.c)$ separa os conjuntos K_1 e K_2 se*

$$\langle a, x\rangle \leq c \leq \langle a, y\rangle\ \ \forall x \in K_1,\ \ \forall y \in K_2.$$

Dizemos que $H(a,c)$ separa estritamente K_1 e K_2 quando as desigualdades acima são estritas.

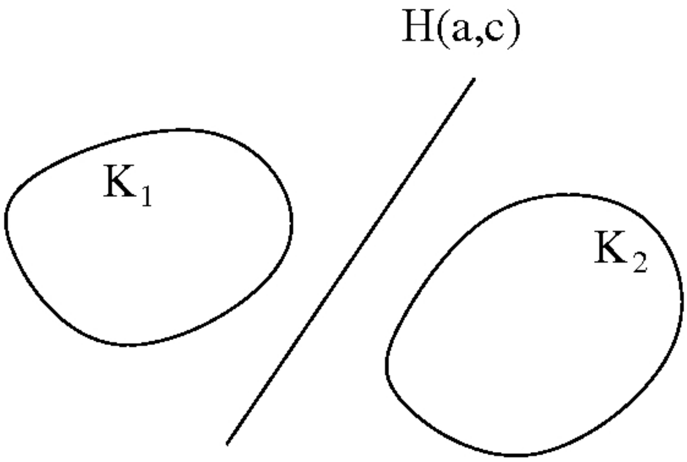

Figura 1.9: Separação deconjuntos.

A noção de separação de conjuntos é muito importante. No sentido geométrico, separação significa que um dos conjuntos fica de um lado do hiperplano $H(a,c)$, enquanto o outro conjunto fica do outro lado, veja figura abaixo.

Como pode ser visto geometricamente, a possibilidade de separar dois conjuntos está ligado ao fato da interseção deles ser vazia ou não, e à convexidade deles.

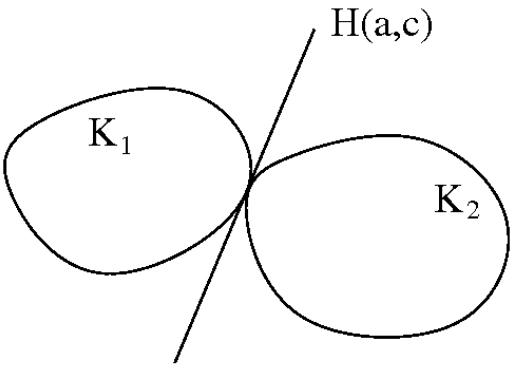

Figura 1.10: Interseção não vazia.

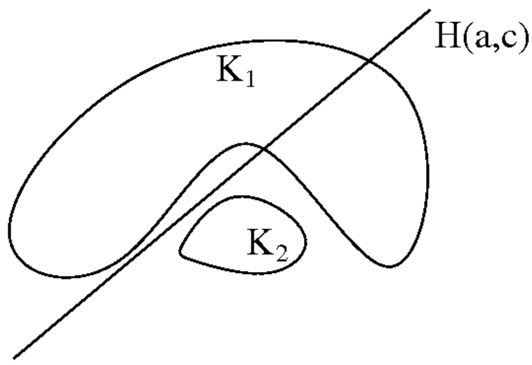

Figura 1.11: Conjunto não convexo.

Até o final desta seção provaremos os teoremas de separação que utilizaremos em algumas demonstrações de importantes teoremas deste trabalho.

Lema 1. (Lema de Minkowski) *Seja $K \subset \mathbb{R}^n$ um conjunto convexo e não-vazio. Se $x \notin \overline{K}$, então existem $a \in \mathbb{R}^n - \{0\}$ e $c \in \mathbb{R}$ tais que*

$$\langle a, x \rangle = c, \quad \langle a, y \rangle \ > c \ \ \forall y \in K.$$

Demonstração. O conjunto $\overline{K}$ é fechado e convexo (porque K é convexo). Seja $\overline{x}$ a projeção de x sobre $\overline{K}$, cuja existência e unicidade são asseguradas pelo Teorema da Projeção.

Definimos $a = \overline{x} - x$ ($a \neq 0$ porque $x \notin \overline{K}$) e $c = \langle \overline{x} - x, x \rangle$. Pela definição, $\langle a, x \rangle = c$. Seja $y \in K$. Temos que

$$\langle a, y \rangle = \langle \overline{x} - x, y \rangle \geq \langle \overline{x} - x, \overline{x} \rangle,$$

onde a desigualdade vale para todo $y \in \overline{K}$ (e portanto para todo $y \in K$), pelo Teorema da Projeção. Obtemos que

$$\begin{aligned} \langle \overline{x} - x, \overline{x} \rangle &= ||x - \overline{x}||^2 + \langle \overline{x} - x, x \rangle \\ &= ||x - \overline{x}||^2 + c > c, \end{aligned}$$

onde usamos que $||x - \overline{x}|| > 0$ (pela hipótese de que $x \notin \overline{K}$). Temos então o resultado desejado. □

Exemplo 12. *Consideremos o conjunto*

$$K = \{(x, y) \in \mathbb{R}^2 / \{(0,0)\}\ ;\ x \geq 0 \text{ e } y \geq 0\} \cup \{(x, y) \in \mathbb{R}^2 ; x < 0 < y\}.$$

Conforme figura 1.12 abaixo, K é convexo, $(0,0) \notin K$ *e* $(0,0) \in \overline{K}$*. Observe que* $(0,0)$ *não pode ser separado estritamente, mas pode ser separado (não estrito).*

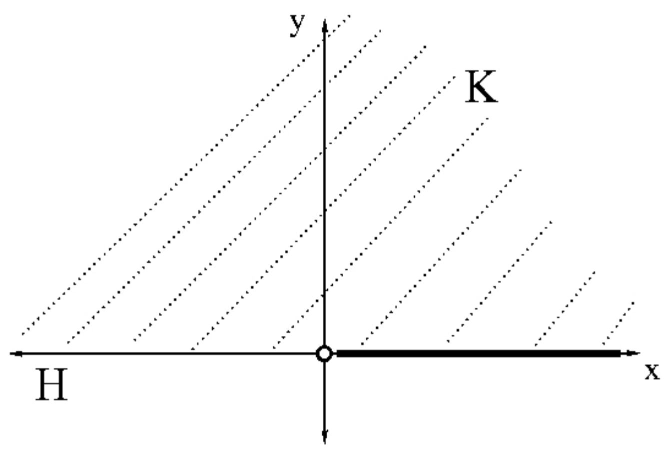

Figura 1.12: Separação deconjuntos.

Agora mostraremos que um ponto na fronteira de um conjunto convexo pode ser separado dele.

Lema 2. *Seja* $K \subset \mathbb{R}^n$ *um conjunto convexo não vazio. Se* $x \in \partial K$*, então existem* $a \in \mathbb{R}^n - \{0\}$ *e* $c \in \mathbb{R}$ *tais que*

$$\langle a, x \rangle = c \quad \langle a, y \rangle \geq c \ \ \forall y \in K.$$

Demonstração. Como $x \in \partial K$, podemos escolher uma sequência $\{x^k\} \to x$ tal que $x^k \notin \overline{K}$ $\forall k \in \mathbb{N}$.

Pelo Lema de Minkowski, todo ponto x^k pode ser separado de K estritamente, isto é, para todo $k \in \mathbb{N}$ existem $a^k \in \mathbb{R}^n - \{0\}$ e $c_k \in \mathbb{R}$ tais que

$$\langle a^k, x^k \rangle = c_k, \quad \langle a^k, y \rangle > c_k \quad \forall y \in K.$$

Podemos admitir (tomando uma subsequência, caso necessário) que $\left\{ \frac{a^k}{||a^k||} \right\} \to a \neq 0$. Obtemos que

$$\lim_{k\to\infty} \frac{c_k}{||a^k||} = \lim_{k\to\infty} \frac{\langle a^k, x^k \rangle}{||a^k||} = \langle a, x \rangle.$$

Definindo $c = \langle a, x \rangle$, dividindo por $||a^k|| > 0$ na desigualdade $\langle a^k, y \rangle > c_k$ e passando o limite, obtemos que para todo $y \in K$ tem-se

$$\langle a, y \rangle \geq c,$$

o que conclui a prova. □

Teorema 19. (Teorema da Separação) *Sejam $K_1, K_2 \subset \mathbb{R}^n$ conjuntos convexos e não vazios tais que $K_1 \cap K_2 = \emptyset$. Então existem $a \in \mathbb{R}^n - \{0\}$ e $c \in \mathbb{R}$ tais que*

$$\langle a, x \rangle \leq c \leq \langle a, y \rangle \;\; \forall x \in K_1, \;\; \forall y \in K_2.$$

Demonstração. O conjunto $D = K_1 - K_2$ é convexo e $0 \notin D$ porque $K_1 \cap K_2 = \emptyset$. Nesta situação há duas possibilidades: ou $0 \notin \overline{D}$, ou $0 \in \partial D$. Em ambos os casos , podemos separar 0 do conjunto D. Concluímos que existe $a \in \mathbb{R}^n - \{0\}$ tal que

$$\langle a, z \rangle \geq 0 \;\; \forall z \in D,$$

isto é, $\langle a, y - x \rangle \geq 0$ para todo $y \in K_2$ e $x \in K_1$. Ou ainda,

$$\langle a, y \rangle \geq \langle a, x \rangle \;\; \forall y \in K_2, \;\; \forall x \in K_1.$$

Em particular, a função $\langle a, . \rangle$ é limitada inferiormente em K_2 e superiormente em K_1.

Dessa forma, obtemos que

$$\gamma_2 = \inf_{y \in K_2} \langle a, y \rangle \geq \sup_{x \in K_1} \langle a, x \rangle = \gamma_1.$$

Definindo $c = (\gamma_1 + \gamma_2)/2$, temos que $\gamma_1 \geq c \geq \gamma_2$ e

$$\forall x \in K_1, \;\; \langle a, x \rangle \leq \sup_{x \in K_1} \langle a, x \rangle = \gamma_1 \leq c,$$

$$\forall y \in K_2, \;\; \langle a, y \rangle \geq \sup_{y \in K_2} \langle a, y \rangle = \gamma_2 \geq c,$$

o que completa a prova do resultado. □

Para garantir a separação estrita, são necessárias hipóteses adicionais.

Teorema 20. (Teorema da Separação Estrita) *Sejam $K_1, K_2 \subset \mathbb{R}^n$ conjuntos convexos, fechados e não vazios. Suponhamos que um deles seja também limitado (logo, compacto). Então $K_1 \cap K_2 = \emptyset$ se, e somente se, existem $a \in \mathbb{R}^n - \{0\}$ e $c \in \mathbb{R}$ tais que*

$$\langle a, x\rangle < c < \langle a, y\rangle \ \ \forall x \in K_1, \ \ \forall y \in K_2.$$

Demonstração. Se temos separação estrita mas existe algum $x \in K_1 \cap K_2$, logo chegamos à contradição: $\langle a, x\rangle < c < \langle a, x\rangle$.

Suponhamos agora que K_2 seja limitado e $K_1 \cap K_2 = \emptyset$. Como o conjunto K_1 é convexo e fechado, o operador de projeção $P_{K_1} : \mathbb{R}^n \to K_1$ é bem definido e contínuo.

Logo a função

$$\Psi : \mathbb{R}^n \to \mathbb{R}_+, \ \Psi(x) = dist(x, K_1) = ||x - P_{K_1}(x)||$$

é contínua. Portanto, como K_2 é compacto, o problema

$$\min_{x \in K_2} \psi(x)$$

tem solução global (Teorema de Weierstrass), que denotamos por a^2. Denotamos $a^1 = P_{K_1}(a^2)$. Pelo Teorema da Projeção,

$$\langle a^2 - a^1, x\rangle \leq \langle a^2 - a^1, a^1\rangle = c_1 \ \ \forall x \in K_1. \tag{1.22}$$

Para todo $x \in K_2$, tem-se que

$$||x - a^1|| \geq dist(x, K_1) \geq \min_{x \in K_2}\{dist(x, K_1)\} = \Psi(a^2)$$

$$= dist(a^2, K_1) = ||a^2 - P_{K-1}(a^2)|| = ||a^2 - a^1||,$$

o que significa que $a^2 = P_{K_2}(a^1)$. Usando novamente o Teorema da Projeção, obtemos que

$$\langle a^2 - a^1, x\rangle \geq \langle a^2 - a^1, a^2\rangle = c_2 \ \ \forall x \in K_2. \tag{1.23}$$

Definimos $a = a^2 - a^1 \neq 0$ ($K_1 \cap K_2 = \emptyset$ implica que $a^2 \neq a^1$) e $c = (c_1 + c_2)/2$. Observamos que $c_2 > c > c_1$ porque

$$c_2 - c_1 = ||a^1 - a^2|| > 0.$$

Agora, o resultado segue de (1.22) e (1.23). □

Exemplo 13. *Consideremos os conjuntos*

$$K_1 = \{(x, y) \in \mathbb{R}^2 \ ; \ y \leq 0\}, \quad K_2 = \{(x, y) \in \mathbb{R}^2 \ ; \ y \geq 1/x, \ x > 0\}.$$

Os dois conjuntos são fechados mas nenhum é limitado. O eixo $Ox = \{(x, y) \in \mathbb{R}^2; y = 0\}$ *é o único hiperplano que separa os conjuntos, mas não separa estritamente.*

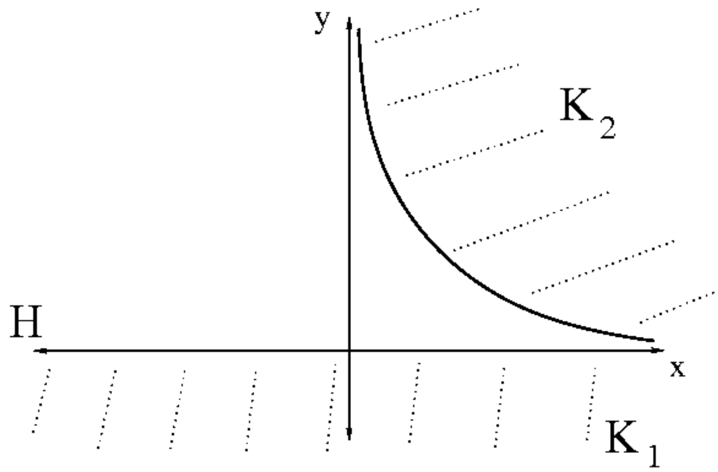

Figura 1.13: Separação de conjuntos.

Capítulo 2

Subdiferencial

Neste capítulo, consideramos dois dos principais subdiferenciais utilizados em otimização. Na primeira seção, definiremos o subdiferencial de Fénchel-Moreau, o subdiferencial usual, e provaremos algumas de suas principais propriedades. Na segunda seção, abordaremos a definição e propriedades do subdiferencial de Plastria, um subdiferencial que contém o de Fénchel-Moreau. Na seção seguinte, definiremos sequência Fejér e quase-Fejér convergente e algumas consequências dessas definições.

2.1 Subdiferencial de Fénchel-Moreau

Antes de definirmor subdiferencial de Fénchel-Moreau, vejamos o seguinte teorema.

Teorema 21. *Sejam $K \subset \mathbb{R}^n$ um conjunto não vazio e convexo e $f : K \to \mathbb{R}$ uma função convexa. Então, para todo $x \in$ int K, existe $s = s(x) \in \mathbb{R}^n$ tal que*

$$\langle s, y - x\rangle + f(x) \leq f(y), \quad \forall y \in K.$$

Demonstração. Pela Proposição 3, E_f é convexo em $\mathbb{R}^n \times \mathbb{R}$. Dados $x \in \mathbb{R}^n$ temos que $(x, f(x)) \in \partial E_f$, donde podemos tomar, pelo Teorema da Separação, um hiperplano suporte a E_f em $(x, f(x))$, ou seja, existe $s_0 \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$, não nulos, tal que

$$\langle s_0, y - x\rangle + \alpha(r - f(x)) \leq 0, \quad \forall (y, r) \in E_f. \tag{2.1}$$

Observe que em (2.1) o número α não pode ser positivo, pois, caso fosse, tomando em (2.1) $y = x + s_0$ e $r > \max\{f(x), f(y)\}$, teríamos:

$f(x) \leq \max\{f(x), f(y)\} < r$, ou seja, $(x, y) \in E_f$ e

$$\langle s_0, x + s_0 - x\rangle + \alpha(r - f(x)) \leq 0,$$

que implicaria em

$$||s_0||^2 \leq \alpha(f(x) - r) < 0$$

que é um absurdo. Dessa forma, temos $\alpha \leq 0$. Novamente em (2.1), se $\alpha = 0$, temos

$$\langle s_0, y - x\rangle \leq 0, \ \ \forall y \in K. \tag{2.2}$$

Como $x \in \text{int } K$, existe $\lambda > 0$, suficientemente pequeno, tal que $x + \lambda s_0 \in K$. Como vale (2.2), para todo $y \in K$, em particular, para $y = x + \lambda s_0$ temos

$$\langle s_0, x + \lambda s_0 - x\rangle \leq 0,$$

ou seja, $\lambda||s_0||^2 \leq 0$, que implica em $s_0 = 0$, que é um absurdo. Logo, em (2.1) temos $\alpha < 0$. Dividindo a desigualdade (2.1) por $|\alpha| > 0$ temos,

$$\langle s, y - x\rangle + f(x) \leq r \ \ \forall (y, r) \in E_f,$$

onde $s = \dfrac{s_0}{|\alpha|}$. Em particular,

$$\langle s, y - x\rangle + f(x) \leq f(y) \ \ \forall y \in K.$$

□

Na verdade, no teorema acima, prova-se que $\alpha \leq -1$.

Definição 17. *Seja $f : K \to \mathbb{R}$ uma função. Dizemos que $s \in \mathbb{R}^n$ é um subgradiente de f no ponto $x \in K$ se*

$$\langle s, y - x\rangle + f(x) \leq f(y), \ \ \forall y \in K.$$

O conjunto de todos os subgradientes de f em x chama-se subdiferencial de f em x (ou subdiferencial de Fénchel-Moreau) e é denotado por $\partial f(x)$. Veja figura 2.1.

A seguir veremos alguns exemplos de subdiferencial.

Exemplo 14. *Considere a função $f : \mathbb{R} \to \mathbb{R}$, dada por $f(x) = |x|$. Dado $x > 0$, da definição de subdiferencial, temos*

$$s(y - x) + x \leq y \ \ \forall y \in \mathbb{R}. \tag{2.3}$$

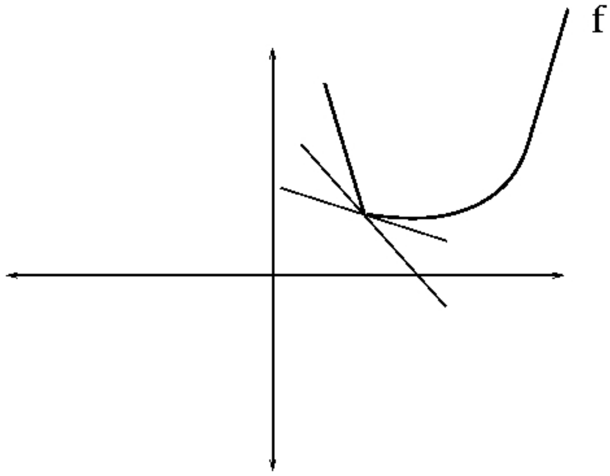

Figura 2.1: Subgradientes.

Se $x < y$, então $s(y-x) \leq y-x$, implica que $s \leq 1$. Caso $x > y$, temos $s \geq 1$. E se $x = y$, a desigualdade (2.3) vale para todo $s \in \mathbb{R}$. Tomando a interseção dos casos, temos $s = 1$.

Analogamente, se tomarmos $x < 0$ temos $s = -1$. Agora, se $x = 0$, temos

$$sy \leq y.$$

Caso $y = 0$, temos que vale (2.3) para todo $s \in \mathbb{R}$. Se $y > 0$, então $s \leq 1$. E se $y < 0$, temos $s \geq -1$. Novamente tomando a interseção dos casos, temos que qualquer $s \in [-1, 1]$, satisfaz (2.3). Portanto,

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

Vemos que nos pontos onde f é diferenciável (em $x \neq 0$) o subdiferencial de f é só um ponto. No caso $x = 0$, o subdiferencial de f é um subconjunto de $\mathbb{R}$.

A seguir, veremos um teorema que garante boas propriedades ao subdiferencial de Fénchel-Moreau.

Teorema 22. *Seja $f : K \to \mathbb{R}$ uma função convexa. Então o conjunto $\partial f(x)$ é não vazio, convexo e compacto.*

Demonstração. Dado $x \in K$, pelo Teorema 21, $\partial f(x) \neq \emptyset$. Provemos agora, que $\partial f(x)$ é convexo.

Sejam $s, w \in \partial f(x)$ e $t \in [0, 1]$. Então

$$f(x) + \langle s, y-x \rangle \leq f(y) \text{ e } f(x) + \langle w, y-x \rangle \leq f(y), \quad \forall y \in \mathbb{R}^n.$$

Então,

$$f(x) + \langle tw + (1-t)s, y - x\rangle = f(x) + (1-t)\langle s, y-x\rangle + t\langle w, y-x\rangle =$$

$$= f(x) + tf(x) - tf(x) + (1-t)\langle s, y-x\rangle + t\langle w, y-x\rangle =$$

$$= (1-t)f(x) + (1-t)\langle s, y-x\rangle + t(f(x) + \langle w, y-x\rangle) =$$

$$= (1-t)(f(x) + \langle s, y-x\rangle) + t(f(x) + \langle w, y-x\rangle) \leq$$

$$\leq (1-t)f(y) + tf(y) = f(y), \quad \forall y \in \mathbb{R}^n,$$

ou seja, $(1-t)s + tw \in \partial f(x)$. Logo, $\partial f(x)$ é convexo.

Vamos mostrar agora que $\partial f(x)$ é fechado e limitado.

Seja $s \in \overline{\partial f(x)}$, logo existe $\{s^k\} \subset \partial f(x)$ tal que $\lim\limits_{k\to\infty} s^k = s$. Como $s^k \in \partial f(x)$, então $\langle s^k, y-x\rangle + f(x) \leq f(y)$, $\forall k \in \mathbb{N}$ e $\forall y \in K$. Dessa forma,

$$\lim_{k\to\infty} \langle s^k, y-x\rangle + f(x) \leq f(y),$$

ou seja,

$$\langle s, y-x\rangle + f(x) \leq f(y).$$

Donde $s \in \partial f(x)$. Logo, $\partial f(x)$ é fechado.

Agora, mostremos que $\partial f(x)$ é limitado. Supoha, por absurdo, que $\partial f(x)$ não seja limitado, isto é, existe $\{s^k\} \subset \partial f(x)$ tal que

$$\lim_{k\to\infty} ||s^k|| = \infty.$$

Sendo $\{\frac{s^k}{||s^k||}\}$ limitada, existe

$$\lim_{k\to\infty} \frac{s^k}{||s^k||} = s, \;\; \text{com} \;\; ||s|| = 1,$$

a menos de subsequência.

Como $s^k \in \partial f(x)$, temos que

$$\langle s^k, y-x\rangle + f(x) \leq f(y), \;\; \forall\, y \in \mathbb{R}^n.$$

Em particular, para $y = x + s$, temos

$$\langle \frac{s^k}{||s^k||}, y-x\rangle \leq \frac{f(y) - f(x)}{||s^k||},$$

que fazendo $k \to \infty$, segue que

$$\langle s, s \rangle \leq 0,$$

pois $||s^k|| \to \infty$. Logo, $1 = ||s|| \leq 0$, que é um absurdo.

Portanto, $\partial f(x)$ é limitado, o que demonstra o teorema. $\square$

Conforme veremos nos exemplos a seguir, a hipótese de f ser convexa não pode ser retirada para que o teorema seja válido.

Exemplo 15. *Seja*

$$f(x) = \begin{cases} x+1, & \text{se} \quad x > 0 \\ x, & \text{se} \quad x \leq 0 \end{cases}.$$

Observe que f não é convexa. Um cálculo simples nos dá que $\partial f(0) = [1, +\infty)$.

Exemplo 16. *Seja $f(x) = x^3$, donde $\partial f(0) = \emptyset$. Neste caso a função f também não é convexa.*

Como consequência do teorema, temos o seguinte corolário.

Corolário 2. *Seja $f : K \to \mathbb{R}$ é uma função diferenciável e convexa no ponto $x \in int\ K$, $K \subset \mathbb{R}^n$ não vazio e convexo. Então o conjunto $\partial f(x)$ contém um só elemento. Neste caso $\partial f(x) = \{\nabla f(x)\}$.*

Demonstração. Pelo teorema 22, $\partial f(x)$ é não vazio, ou seja, existe $s \in \partial f(x)$. Então,

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

Em particular, a desigualdade acima vale para $y = x + \lambda d$, para $\lambda > 0$ suficientemente pequeno e para todo $d \in \mathbb{R}^n$, pois $x \in int\ K$. Logo,

$$\lambda \langle s, d \rangle + f(x) \leq f(x + \lambda d). \tag{2.4}$$

Pela diferenciabilidade da f no ponto x, temos

$$f(x + \lambda d) = f(x) + \lambda \langle \nabla f(x), d \rangle + \lambda ||d|| \alpha(x, \lambda d). \tag{2.5}$$

Subtraindo (2.5) de (2.4), temos

$$\lambda \langle s - \nabla f(x), d \rangle - \lambda ||d|| \alpha(x, \lambda d) \leq 0.$$

Dividindo por $\lambda > 0$ e fazendo $\lambda \to 0^+$, segue que

$$\langle s - \nabla f(x), d \rangle \leq 0.$$

Tomando $d = s - \nabla f(x)$, pela desigualdade anterior, temos $s = \nabla f(x)$. $\square$

Para mostrar que a convexidade da função f no teorema anterior é necessária, temos o seguinte exemplo.

Exemplo 17. *Seja $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = |x|^{1/2}$. Usando a definição de subdiferencial obtemos que $\partial f(0) = \{0\}$, mas f não é diferenciável em $x = 0$.*

A seguir, definiremos cone normal.

Definição 18. *Sejam $K \subset \mathbb{R}^n$ um conjunto convexo e $\overline{x} \in K$. O cone normal no ponto $\overline{x}$ em relação ao conjunto K é dado por*

$$N_K(\overline{x}) = \{y \in \mathbb{R}^n \ ; \ \langle y, x - \overline{x}\rangle \leq 0, \ \ \forall x \in K\}.$$

No próximo teorema, temos uma caracterização do subdiferencial utilizando a definição de cone normal.

Teorema 23. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Um vetor $w \in \partial f(x)$ se, e somente se, o vetor $(w, -1) \in N_{E_f}(x, f(x))$, onde E_f é o epígrafo de f.*

Demonstração. Seja $w \in \partial f(x)$, isto é, $\langle w, y - x\rangle + f(x) \leq f(y)$, para todo $y \in \mathbb{R}^n$. Dessa forma

$$\langle w, y - x\rangle + f(x) - r \leq 0,$$

para todo $y \in \mathbb{R}^n$ tal que $f(y) \leq r$. Então,

$$\langle w, y - x\rangle + (-1)(r - f(x)) \leq 0,$$

para todo $(y, r) \in E_f$. Logo,

$$(w, -1) \in N_{E_f}(x, f(x)).$$

Reciprocamente, dado $(w, -1) \in N_{E_f}(x, f(x))$, temos

$$\langle (w, -1), (z - x, r - f(x))\rangle \leq 0,$$

para todo $(z, r) \in E_f$, ou seja, $\langle w, z - x\rangle + f(x) \leq r$, para todo $z \in \mathbb{R}^n$ tal que $f(z) \leq r$. Em particular, tomando $z = y$ e $r = f(y)$, temos, $\langle w, y - x\rangle + f(x) \leq f(y)$. Logo, $w \in \partial f(x)$. $\square$

No exemplo 23, em $x = 0$ (nos demais pontos a função é diferenciável) podemos verificar este resultado através do cone normal, conforme figura 2.2.

Tendo em vista o Teorema da Separação, podemos garantir que o conjunto $N_{E_f}(x, f(x))$ é não vazio, convexo e fechado. Dessa forma, usando o teorema 23, temos uma outra maneira de demonstrar o teorema 22.

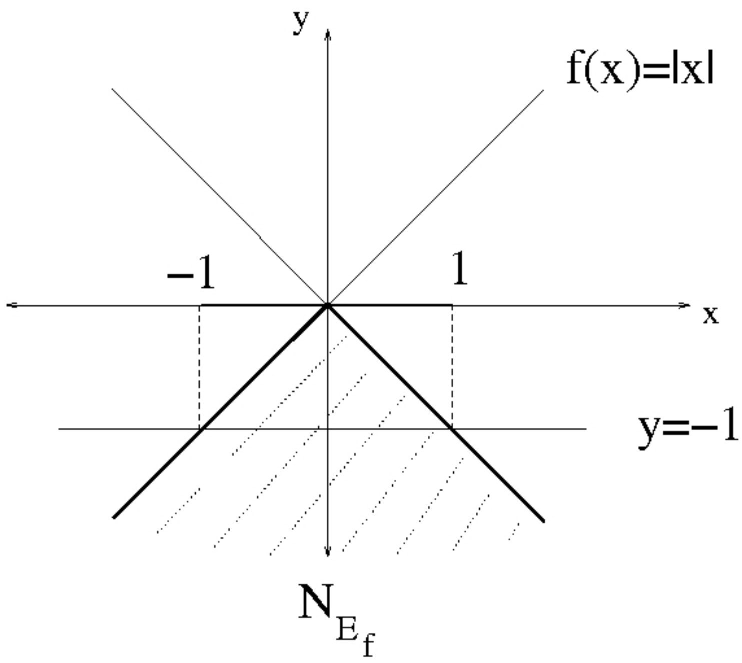

Figura 2.2: Cone normal em relação ao epígrafo de f.

Observação 1. *Sabemos que* $f(x) = x^3$ *é quase-convexa e pelo exemplo 16 temos que* $\partial f(0) = \emptyset$*. Logo, o teorema 22 não vale para funções quase-convexas.*

2.2 Subdiferencial de Plastria

Com o objetivo de obtermos um resultado semelhante ao do teorema 22 para funções quase-convexas, iremos definir nesta subseção o subdiferencial de Plastria e abordar algumas de suas principais propriedades.

Definição 19. *Seja* $f : K \to \mathbb{R}$ *definida em* $K \subset \mathbb{R}^n$*. Dizemos que* $s \in \mathbb{R}^n$ *é um subgradiente de Plastria da função* f *no ponto* $x \in K$ *se*

$$\langle s, y - x\rangle + f(x) \leq f(y), \tag{2.6}$$

para todo $y \in K$*, tal que* $f(y) < f(x)$*. O conjunto de todos os subgradientes de Plastria é chamado de subdiferencial de Plastria e será denotado por* $\partial^P f(x)$*.*

Se $s \in \partial f(x)$*, então vale* (2.6)*,* $\forall y \in K$*. Em particular, vale* (2.6) *para todo* $y \in K$ *tal que* $f(y) < f(x)$*. Logo* $\partial f(x) \subset \partial^P f(x)$*. A recíproca deste fato não é verdadeira. Com efeito, dado* $s \in \partial^P f(x)$*, temos que*

$$\langle s, y - x\rangle + f(x) \leq f(y),$$

para todo $y \in K$ tal que $f(y) < f(x)$. Dessa forma, para $\lambda \geq 1$, temos

$$\langle \lambda s, y - x \rangle \leq \lambda[f(y) - f(x)] \leq f(y) - f(x).$$

Logo, $\{\lambda s \; ; \; \lambda \geq 1\} \subset \partial^P f(x)$. Isto mostra que $\partial^P f(x)$ não é limitado.

De um modo geral, $\partial f(x)$ e $\partial^P f(x)$ são diferentes para f convexa. Na verdade $\partial f(x) = \partial^P f(x)$, somente quando $\partial f(x) = \partial^P f(x) = \{0\}$.

Para o caso de funções convexas, temos que $\nabla f(x) \in \partial f(x)$. No caso do subdiferencial de Plastria, isto geralmente não ocorre, conforme exemplo abaixo.

Exemplo 18. *Seja $f : \mathbb{R} \to \mathbb{R}$ dada por*

$$f(x) = \begin{cases} -27, & \text{se} \quad x \leq -3 \\ x^3, & \text{se} \quad -3 < x < 3 \\ 27, & \text{se} \quad x \geq 3 \end{cases} .$$

É fácil ver que $\nabla f(-2) = 12 \notin \partial^P f(-2) = [19, \infty)$.

A seguir, veremos uma importante propriedade do subdiferencial de Plastria.

Teorema 24. *Se $f : \mathbb{R}^n \to \mathbb{R}$ é quase-convexa e Lipschitz, com constante L, então:*

1) *$\partial^P f(x) \neq \emptyset$ para todo $x \in \mathbb{R}^n$. Mais precisamente, existe $s \in \partial^P f(x)$, com $||s|| \leq L$;*

2) *$\partial^P f(x)$ é um conjunto convexo e fechado, para todo $x \in \mathbb{R}^n$.*

Demonstração. Para provar 1, considere $a \in \mathbb{R}^n$ e o conjunto de nível estrito $L_f^<(f(a)) = \{x \in \mathbb{R}^n \; ; \; f(x) < f(a)\}$, que é convexo, pela quase-convexidade da f. Sendo f Lipschitz, f é contínua e $L_f^<(f(a))$ é um conjunto aberto que não contém $f(a)$. Então existe um hiperplano separando estritamente o ponto a e $L_f^<(f(a))$, isto é, existe $s_0 \in \mathbb{R}^n$ (que pode ser tomado $||s_0|| = 1$) tal que, para todo $x \in L_f^<(f(a))$,

$$\langle x - a, s_0 \rangle < 0. \tag{2.7}$$

Tomando $s = Ls_0$, temos que $||s|| = L$, restando provar que $s \in \partial^P f(a)$. Para todo $x \in L_f^<(f(a))$, denote por x' a projeção ortogonal de x ao hiperplano

$$\langle z - a, s \rangle = 0 \tag{2.8}$$

Por (2.7) e (2.8), $x' \notin L_f^<(f(a))$, ou seja, $f(x') \geq f(a)$. Além disso,

$$\langle a - x, s \rangle = ||x - x'|| \; ||s|| = L||x - x'||.$$

Logo,

$$f(a) - f(x) \leq f(x') - f(x) \leq L||x - x'|| = \langle a - x, s \rangle,$$

para todo $x \in L_f^{<}(f(a))$. Portanto, $s \in \partial^P f(a)$.

A demonstração de 2) é inteiramente análoga ao teorema 22. □

As funções Lipschitz possuem boas propriedades e o teorema abaixo nos dá condições para que uma função seja Lipschitz.

Teorema 25. *Seja $f : K \to \mathbb{R}$ definida no subconjunto $K \subset \mathbb{R}^n$. Suponha que, existe $L > 0$ tal que , para cada $x \in K$, exite $x^0 \in \partial^P f(x)$, com $||x^0|| \leq L$. Então f é Lipschitz em K.*

Demonstração. É sufuciente considerar $x, y \in K$ tal que $f(x) \neq f(y)$. Suponha, sem perda de generalidade, que $f(x) < f(y)$. Aplicando a hipótese do teorema ao ponto $y \in K$, exite $y^0 \in \partial^P f(y)$, com $||y^0|| \leq L$, tal que

$$0 < f(y) - f(x) \leq \langle y^0, y - x \rangle \leq L||y - x||,$$

que prova o teorema. □

A seguir, apresentaremos uma outra importante propriedade do subdiferencial de Plastria.

Teorema 26. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função quase-convexa e Lipschitz, com constante de Lipschitz L e suponha que $z \in \mathbb{R}^n$. Então, para todo $v \in S(0,1) \cap N_{L_f^{<}(f(z))}(z)$, o vetor*

$$\overline{v} = Lv$$

pertence a $\partial^P f(z)$. Adicionalmente, para todo $u \in \partial^P f(z)/\{0\}$, o vetor

$$\overline{u} = L\frac{u}{||u||}$$

pertence a $\partial^P f(z)$.

Demonstração. Seja $z \in \mathbb{R}^n$. Como f é Lipschitz, segue que f é contínua e $L_f^{<}(f(z))$ é aberto e convexo, que não contém $z \in \mathbb{R}^n$. Considere $v \in S(0,1) \cap N_{L_f^{<}(f(z))}(z)$ e tome $\overline{v} = Lv$. Para todo $y \in L_f^{<}(f(z))$, denotaremos por $P(y)$ a projeção ortogonal de y ao hiperplano

$$\{x \in \mathbb{R}^n \ ; \ \langle x - z, \overline{v} \rangle = 0\}.$$

Então, $P(y) \notin L_f^{<}(f(z))$, donde $f(P(y)) \geq f(z)$. Além disso, temos

$$\langle z - y, \overline{v} \rangle = ||P(y) - y||||\overline{v}|| = L||P(y) - y||.$$

Sendo f Lipschitz, temos

$$f(z) - f(y) \leq f(P(y)) - f(y) \leq L||P(y) - y|| = \langle z - y, \overline{v}\rangle,$$

mostrando que $\overline{v} \in \partial^P f(z)$.

Agora, seja $u \in \partial^P f(x)/\{0\}$. Dessa forma,

$$\langle \frac{u}{||u||}, y - x\rangle = \frac{1}{||u||}\langle u, y - x\rangle \leq \frac{f(y) - f(x)}{||u||} < 0.$$

Logo,

$$\frac{u}{||u||} \in S(0,1) \cap N_{L_f^{<}(f(z))}(z).$$

Portanto, pela primeira parte do teorema, $\overline{u} \in \partial^P f(z)$. □

2.3 Sequência Féjer e quase-Féjer convergente

Uma importante propriedade em otimização é a de uma sequência Fejér convergente, que definiremos a seguir.

Definição 20. *Uma sequência $\{x^k\}$ em $\mathbb{R}^n$ é Fejér convergente a um conjunto $U \subset \mathbb{R}^n$ se*

$$||x^{k+1} - x|| \leq ||x^k - x||, \quad \forall k \in \mathbb{N}, \quad \forall x \in U. \tag{2.9}$$

Como consequência temos a seguinte proposição.

Proposição 5. *Se $\{x^k\}$ é Fejér convergente ao conjunto não vazio $U \subset \mathbb{R}^n$, então $\{x^k\}$ é limitada. Se $x \in U$ é ponto de acumulação de $\{x^k\}$, então $\lim\limits_{k\to\infty} x^k = x$.*

Demonstração. Segue de (2.9) que $||x^k - x|| \leq ||x^0 - x||$, para todo $x \in U$. Então, $\{x^k\}$ está contida na bola de centro $x \in U$ e raio $||x^0 - x||$, donde $\{x^k\}$ é limitada. Agora, considere $\{x^{k_j}\}$ uma subsequência de $\{x^k\}$, tal que $\lim\limits_{k_j\to\infty} x^{k_j} = x$. Como $x \in U$, de (2.9), temos que a sequência de números reais não negativos, $\{||x^k - x||\}$ é decrescente e possui uma subsequência, $\{||x^{k_j} - x||\}$, que converge a zero. Logo, $\lim\limits_{k\to\infty} ||x^k - x|| = 0$. Portanto, $\lim\limits_{k\to\infty} x^k = x$. □

A seguir definiremos a propriedade de quase-Fejér convergente.

Definição 21. *Uma sequência $\{x^k\}$ é quase-Fejér convergente para o subconjunto $U \subset \mathbb{R}^n$ se, para todo $u \in U$, existe uma sequência $\{\epsilon_k\} \subset \mathbb{R}$, tal que $\epsilon_k \geq 0$, $\sum\limits_{k=0}^{\infty} \epsilon_k < \infty$ e $||x^{k+1} - u||^2 \leq ||x^k - u||^2 + \epsilon_k$, para todo $k \in \mathbb{N}$.*

Proposição 6. *Se* $\{x^k\}$ *é quase-Fejér convergente ao conjunto não vazio* $U \subset \mathbb{R}^n$*, então* $\{x^k\}$ *é limitada. Além disso, se* $x \in U$ *é ponto de acumulação de* $\{x^k\}$*, então* $\lim\limits_{k\to\infty} x^k = x$*.*

Demonstração. Dado $u \in U$ aplicando a definição de quase-Fejér convergente, temos

$$||x^k - u||^2 \leq ||x^0 - u||^2 + \sum_{j=0}^{k-1} \epsilon_j \leq ||x^0 - u||^2 + \sum_{j=0}^{\infty} \epsilon_j.$$

Como $\sum\limits_{k=0}^{\infty} \epsilon_k < \infty$, temos que $\{x^k\}$ é limitada.

Agora, sejam $x \in U$ ponto de acumulação de $\{x^k\}$ e $\delta > 0$. Pela hipótese de quase-Fejér convergente, garantimos que existe $k_0 \in \mathbb{N}$ tal que

$$\sum_{j=k_0}^{\infty} \epsilon_j < \delta/2.$$

Denotando por $\{x^{k_i}\}$ a subsequência que converge para x, existe $k_1 \in \mathbb{N}$ tal que

$$||x^{k_i} - x||^2 < \delta/2, \quad \forall k_i > k_1.$$

Logo, para todo $k \geq \max\{k_0, k_1\}$, temos

$$||x^k - x||^2 \leq ||x^{k_i} - x||^2 + \sum_{j=k_i}^{k-1} \epsilon_j \leq ||x^{k_i} - x||^2 + \sum_{j=k_i}^{\infty} \epsilon_j < \delta/2 + \delta/2 = \delta.$$

Portanto, $\lim\limits_{k\to\infty} x^k = x$. □

Capítulo 3

Método de Subgradiente: Caso convexo

A seguir introduziremos o método de subgradiente para funções convexas, onde utilizaremos subgradientes aproximados. Lembre-se que para $\epsilon \geq 0$*, o* ϵ*-subdiferencial de* f *no ponto* $x \in \mathbb{R}^n$ *é o conjunto dos* ϵ*-subgradientes, dado por*

$$\partial_\epsilon f(x) = \{y \in \mathbb{R}^n \ ; \ \langle y, z - x \rangle + f(x) - \epsilon \leq f(z), \ \ \forall z \in \mathbb{R}^n\}. \tag{3.1}$$

O subdiferencial $\partial f(x)$ *é obtido tomando* $\epsilon = 0$*.*

Algoritmo 1. (O método de subgradiente) *Dado* $x^0 \in \mathbb{R}^n$ *definiremos a sequência* $\{x^k\}$ *como segue*

$$x^{k+1} = x^k - \lambda_k g^k, \tag{3.2}$$

onde $\{\epsilon_k\}$ *e* $\{\lambda_k\}$ *são sequências de números reais não negativos e* $g^k \in \partial_{\epsilon_k} f(x^k)$*.*

Estudaremos algumas propriedades de convergência do Algoritmo 1.

Teorema 27. (Convergência do método de subgradiente I) *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função convexa no* $\mathbb{R}^n$*. Suponhamos que no Algoritmo 1 as sequências* $\{\epsilon_k\}$*,* $\{\lambda_k\}$ *e* $\{g^k\}$ *satisfaçam as condições*

$$\sum_{k=0}^{\infty} \lambda_k = \infty, \tag{3.3}$$

$$\lim_{k \to \infty} \epsilon_k = 0, \tag{3.4}$$

$$\lim_{k \to \infty} \lambda_k ||g^k||^2 = 0. \tag{3.5}$$

Então, para a sequência $\{x^k\}$ *gerada pelo Algoritmo 1, tem-se que*

$$\liminf_{k \to \infty} f(x^k) = f^*,$$

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

Demonstração. Dado $x \in \mathbb{R}^n$ qualquer, de (3.1) e (3.2), temos que

$$
\begin{aligned}
||x^{k+1}-x||^2 &= ||x^{k+1}-x^k||^2 + ||x^k-x||^2 + 2\langle x^{k+1}-x^k, x^k-x\rangle \\
&= ||x^k-x||^2 + 2\lambda_k\langle g^k, x-x^k\rangle + {\lambda_k}^2||g^k||^2 \\
&\leq ||x^k-x||^2 + 2\lambda_k(f(x)-f(x^k)+\epsilon_k) + {\lambda_k}^2||g^k||^2. \quad (3.6)
\end{aligned}
$$

Suponha que a afirmação não seja verdade, isto é, que $\liminf_{k\to\infty} f(x^k) > f^*$. Então, existem $x \in \mathbb{R}^n$, $\delta > 0$ e $k_1 \in \mathbb{N}$, tal que

$$f(x) < f(x^k) - \delta, \quad \forall k \geq k_1. \quad (3.7)$$

De (3.4) e (3.5), segue que podemos tomar $k_2 \in \mathbb{N}$ sufucientemente pequeno, de modo que

$$\lambda_k||g^k||^2 + 2\epsilon_k \leq \delta, \quad \forall k \geq k_2.$$

Combinando esta última relação com (3.6) e (3.7), obtemos que

$$
\begin{aligned}
||x^{k+1}-x||^2 &= ||x^k-x||^2 + \lambda_k(2\epsilon_k + \lambda_k||g^k||^2 - 2\delta) \\
&\leq ||x^k-x||^2 - \delta\lambda_k, \quad \forall k \geq k_0,
\end{aligned}
$$

onde $k_0 = \max\{k_1, k_2\}$.

Portanto,

$$
\begin{aligned}
\delta\sum_{j=k_0}^{k}\lambda_j &\leq \sum_{j=k_0}^{k}(||x^j-x||^2 - ||x^{j+1}-x||^2) \\
&= ||x^{k_0}-x||^2 - ||x^{k+1}-x||^2 \\
&\leq ||x^{k_0}-x||^2,
\end{aligned}
$$

que fazendo $k \to \infty$, obtemos uma contradição com (3.3). □

Se fizermos normalização das direções empregadas no Algoritmo 1, as afirmações sobre convergência do método podem ser fortalecidas, para escolha de comprimentos de passo mais específicos. Para simplificar, consideraremos o caso de subgradientes exatos. Obtemos então o seguinte esquema

$$x^{k+1} = x^k - \beta_k\frac{g^k}{||g^k||}, \quad g^k \in \partial f(x^k), \quad k = 0, 1, \ldots, \quad (3.8)$$

onde $\beta_k > 0$ é o parâmetro de comprimento de passo.

Teorema 28. (Convergência do método de subgradiente II) *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa em $\mathbb{R}^n$. Suponhamos que o problema (1.4) tenha solução. Suponha ainda que no Algoritmo 1,*

$$\epsilon_k = 0, \quad \lambda_k = \frac{\beta_k}{||g^k||} \quad \forall k \in \mathbb{N}, \tag{3.9}$$

o que corresponde a (3.8), e que a sequência $\{\beta_k\}$ satisfaça as condições

$$\sum_{k=0}^{\infty} \beta_k = \infty, \quad \sum_{k=0}^{\infty} {\beta_k}^2 < \infty. \tag{3.10}$$

Então a sequência $\{x^k\}$ gerada pelo Algoritmo 1 converge para uma solução do problema (1.4).

Demonstração. Seja $x^* \in \mathbb{R}^n$ uma solução do problema (1.4). Obviamente, $f(x^*) \leq f(x^k)$, para todo $k \in \mathbb{N}$. Tomando em (3.6) $x = x^*$, e usando (3.9), obtemos

$$||x^{k+1} - x^*||^2 \leq ||x^k - x^*||^2 + {\beta_k}^2.$$

Seja k_0 arbitrário, porém fixo. Utilizando a última desigualdade, para qualquer $k \geq k_0 + 1$, obtemos

$$||x^k - x^*||^2 \leq ||x^{k_0} - x^*||^2 + \sum_{i=k_0}^{k-1} {\beta_i}^2 \tag{3.11}$$

$$\leq ||x^{k_0} - x^*||^2 + \sum_{i=0}^{\infty} {\beta_i}^2 < \infty.$$

Dessa forma, obtemos que a sequência $\{x^k\}$ é limitada. Por isso, de (3.9) e (3.10) obtemos as relações (3.3)-(3.5). Logo pelo Teorema 27, temos que $\liminf_{k\to\infty} f(x^k) = f^* = f(x^*)$. Usando a continuidade da função f em $\mathbb{R}^n$ e o fato que $\{x^k\}$ é limitada, concluímos que $\{x^k\}$ tem ponto de acumulação $\hat{x} \in \mathbb{R}^n$ tal que $f(\hat{x}) = f(x^*)$, isto é, $\hat{x}$ é solução do problema (1.4).

Podemos então tomar $x^* = \hat{x}$ na análise acima, para concluir, de (3.11) que para todo $k \geq k_j + 1$

$$||x^k - \hat{x}||^2 \leq ||x^{k_j} - \hat{x}||^2 + \sum_{i=k_j}^{\infty} {\beta_i}^2, \tag{3.12}$$

onde, de (3.10),

$$\lim_{k_j \to \infty} \left(\sum_{i=k_j}^{\infty} {\beta_i}^2 \right) = 0.$$

Em particular, dado $\delta > 0$ arbitrariamente pequeno, existe $k_{j_0} \in \mathbb{N}$, tal que

$$\sum_{i=k_{j_0}}^{\infty} \beta_i{}^2 < \delta^2/8$$

e

$$||x^{k_j} - \hat{x}||^2 < \delta^2/8,$$

para todo $k_j \geq k_{j_0}$. De (3.12), concluímos que

$$||x^k - \hat{x}|| < \delta, \quad \forall k \geq k_j + 1.$$

Isto significa que $\{x^k\}$ converge para $\hat{x}$, que é a solução do problema. $\square$

Observação 2. *Na demonstração do teorema acima mostramos que a sequência $\{x^k\}$ é quase-Fejér convergente a solução do problema em questão. Logo, $\{x^k\}$ converge para a solução do problema, pela proposição 6.*

A propriedade que faz o método do subgradiente funcionar é a seguinte. Se x^* é minimizador de f, de (3.6) temos que

$$||x^{k+1} - x^*||^2 \leq ||x^k - x^*||^2 - 2\lambda_k(f(x^k) - f(x^*)) + \lambda_k||g^k||^2,$$

onde $g^k \in \partial f(x^k)$. Donde obtemos que

$$||x^{k+1} - x^*||^2 \leq ||x^k - x^*||^2$$

se

$$0 < \lambda_k < \frac{2(f(x^k) - f(x^*))}{||g^k||^2}.$$

Ou seja, para passos suficientemente curtos, a distância ao conjunto solução diminui. No entanto, a escolha do passo está condicionada ao conhecimento do valor ótimo do problema. É por isso que temos que empregar regras do tipo (3.10), para que valores dos comprimentos de passos sejam suficientemente curtos. Porém, métodos com comprimento de passos dados por uma regra do tipo (3.10) converge, de uma certa forma, lentamente.

Capítulo 4

Método de Subgradiente: Caso quase-convexo

Neste capítulo, abordaremos o conceito e alguns tipos de taxas de convergência. Além disso, apresentaremos algumas propriedades de convergência do método de subgradiente normalizado para funções quase-convexas, que como vimos é uma classe de funções mais ampla que as convexas. Com uma hipótese adicional, melhoraremos a taxa de convergência do método. Também apresentaremos versões inexatas deste modelo.

4.1 Taxas de convergência

Taxa de convergência nada mais é que a velocidade com que a sequência converge para a solução. Utilizaremos o conceito de convergência relativo a

$$\limsup_{k\to\infty} \frac{||x^{k+1} - x^*||}{||x^k - x^*||^p} = \beta < \infty.$$

Se $p = \beta = 1$, a convergência é dita sub-linear, o pior resultado.

Se $p = 1$ e $\beta < 1$, então dizemos que a sequência tem taxa de convergência linear. Neste caso, quando assintoticamente tivermos $||x^{k+1} - x^*|| \leq \beta||x^k - x^*||$, a convergência linear é chamada de convergência geométrica.

Se $p \geq 1$ e $\beta = 0$, dizemos que a sequência tem taxa de convegência super-linear.

Se $p = 2$ e $\beta < \infty$, dizemos que a sequência tem taxa de convergência quadrática.

4.2 Método exato

Seja $f : \mathbb{R}^n \to \mathbb{R}$ *uma função contínua, mas não necessariamente diferenciável definida no espaço Euclidiano n-dimensional* $\mathbb{R}^n$*. Assim, podemos considerar problema de minimização sobre o* $\mathbb{R}^n$ *como em* (1.4).

Denotaremos por D^* *e* f^* *o conjunto solução do problema* (1.4) *e o valor ótimo em* (1.4)*, respectivamente. Se a função custo* f *é convexa, obtemos a boa definição do problema de minimização convexa, que pode ser resolvida com a ajuda de vários métodos de otimização não-suave, onde um deles, o método do subgradiente, que foi visto no capítulo anterior, tinha a seguinte forma:*

$$x^{k+1} := x^k - \lambda_k g^k, \quad g^k \in \partial f(x^k), \quad \lambda_k > 0, \tag{4.1}$$

tal que

$$\sum_{k=0}^{\infty} \lambda_k = \infty, \quad \sum_{k=0}^{\infty} {\lambda_k}^2 < \infty. \tag{4.2}$$

Contudo, a convergência de tal método é muito devagar por causa da primeira condição em (4.2)*. Na verdade, esta condição evita o método de alcançar uma taxa linear de convergência. Por essa razão, outras regras podem ser sugeridas, mas elas também não permitirão obter resultados de alta convergência para classes amplas de problemas. Como resultado, muitos trabalhos em otimização convexa não suave foram desenvolvidos para outras classes de métodos de soluções. No entanto, o método do subgradiente usual pode ser usado na solução de problemas de dimensões altas e estaremos interessados em investigar e revelar versões efetivas desse método. Em alguns trabalhos é sugerido a escolha do passo em* (4.1) *como segue:*

$$\lambda_k = \frac{(f(x^k) - f^*)}{||g^k||^2}. \tag{4.3}$$

Essa versão claramente requer o conhecimento do valor ótimo em (1.4)*, mas possui melhores propriedades de convergência em comparação com a variante inicial* (4.1) *e* (4.2)*. Na verdade, se o problema* (1.4) *é soluvel,* (4.1) *e* (4.3) *geram a sequência* $\{x_k\}$ *que é Fejér convergente para a solução. Além disso, ela pode alcançar uma taxa de convergência linear. Vários algoritimos, que podem ser vistos como intermediários entre* (4.1), (4.2) *e* (4.1), (4.3)*, foram considerados em [20,21]. Contudo, todos esses algoritimos possuem a mesma taxa de convergência. O principal objetivo deste capítulo é apresentar uma modificação do algoritimo* (4.1), (4.3) *que alcança uma taxa de convergência linear para classes amplas de problemas de otimização. A idéia desta modificação decorre do fato de que o subgradiente usual repre-*

senta apenas uma aproximação afim da função custo em cada ponto, enquanto que o subgradiente normalizado utiliza a informação sobre algum majorante superior da função custo começando de um ponto ótimo. Essa aproximação pode ser estendida para o caso quase-convexo. Neste capítulo, consideraremos propriedades de convergência do método modificado com computações inexatas.

Dado um conjunto $X \subseteq \mathbb{R}^n$*, denotaremos por*

$$N(X,x) = \{q \in \mathbb{R}^n; \langle q, y-x\rangle \leq 0 \ \ \forall y \in X\}$$

e $G_\epsilon(x)$ *o correspondente conjunto de nível estrito da função f, isto é,*

$$G_\epsilon(x) = \{z \in \mathbb{R}^n; f(z) < f(x) - \epsilon\}.$$

Por simplicidade, o conjunto $G(x) = G_0(x)$.

Neste capítulo consideraremos o problema de otimização (1.4) *onde iremos supor que a função custo* f *é quase-convexa e que o problema* (1.4) *é solúvel. Então* D^* *é claramente não vazio, convexo e fechado. Neste caso, podemos escrever em analogia com o método do subgradiente* (4.1) *como segue:*

$$x^{k+1} := x^k - \lambda_k q^k, \quad q^k \in Q(x^k),\ \lambda_k > 0, \tag{4.4}$$

onde o conjunto

$$Q(x) = S(0,1) \bigcap N(G(x), x).$$

No que segue, supomos que todos os pontos iterados são não ótimos, isto é, o método gera uma sequência infinita. Então os conjuntos $Q(x^k)$ *são não vazios, para cada* $k \in \mathbb{N}$*, devido ao teorema da separação. No caso convexo, o cone* $N(G(x), x)$ *concide com o subdiferencial de Fénchel-Moreau, conforme vimos no capítulo 2, então* (4.4) *torna-se uma versão normalizada do método do subgradiente. A fim de escolher a regra do passo em* (4.4) *e obter a taxa de convergência necessitamos de propriedades dos vetores normais que são sililares aos do subgradiente. Essas propriedades são dadas aos elementos do conjunto*

$$Q_\epsilon(x) = S(0,1) \bigcap N(G_\epsilon(x), x)$$

com $\epsilon > 0$.

Proposição 7. *Suponha que a função f satisfaz a condição de $Hölder$ com grau $\beta > 0$ no ponto $x^* \in D^*$, no conjunto $\overline{G(x)}$ para todo ponto $x \in \mathbb{R}^n \backslash D^*$. Então, para todo $\epsilon \geq 0$ tal que $f(x) - f^* > \epsilon$, temos*

$$f(x) - f^* \leq L\langle q, x - x^*\rangle^{\beta} + \epsilon, \quad \forall q \in Q_{\epsilon}(x). \tag{4.5}$$

Demonstração. Sendo f contínua e quase-convexa, $G_{\epsilon}(x)$ é convexo e aberto. Com efeito, dados $a, b \in G_{\epsilon}(x)$, tomemos $c = \lambda a + (1-\lambda)b$, com $\lambda \in [0,1]$. Então,

$$f(c) = f(\lambda a + (1-\lambda)b) \leq max\{f(a), f(b)\} < f(x) - \epsilon$$

que implica em $c \in G_{\epsilon}(x)$. Logo $G_{\epsilon}(x)$ é convexo.

Provaremos que o complementar de $G_{\epsilon}(x)$ é fechado. Tome $z^k \in G_{\epsilon}(x)^c$, tal que $z^k \to \overline{z}$. Como f é contínua $f(z^k) \to f(\overline{z})$. Como $f(z^k) \geq f(x) - \epsilon$, temos $f(\overline{z}) \geq f(x) - \epsilon$. Logo $\overline{z} \in G_{\epsilon}(x)^c$, e isto implica que $G_{\epsilon}(x)^c$ é fechado, donde $G_{\epsilon}(x)$ é aberto.

Agora, tome $r = inf\{||z - x^*||;\ z \in \partial G_{\epsilon}(x)\}$, onde $\partial G_{\epsilon}(x)$ denota a fronteira do conjunto $G_{\epsilon}(x)$. Pela definição de ínfimo existe $\{z^k\} \subset \partial G_{\epsilon}(x)$ e $\{\delta_k\} \subset \mathbb{R}$, $\delta_k > 0$, com $\delta_k \to 0$, tal que $||z^k - x^*|| \leq r + \delta_k$, $\forall k \in \mathbb{N}$.

Então,

$$f(x) - \epsilon - f^* = f(z^k) - f^* \leq L||z^k - x^*||^{\beta} \leq L(r + \delta_k)^{\beta} \leq Lr^{\beta}.$$

Logo,

$$f(x) - f^* \leq Lr^{\beta} + \epsilon. \tag{4.6}$$

Como $x^* + rz \in \overline{G_{\epsilon}(x)}$, para todo $z \in S(0,1)$, temos $\langle q, x^* + rz - x\rangle \leq 0$, $\forall q \in Q_{\epsilon}(x)$. Tomando $z = q$, temos:

$$r\langle q, q\rangle + \langle q, x^* - x\rangle = \langle q, x^* + rq - x\rangle \leq 0$$

Com isso,

$$r \leq \langle q, x - x^*\rangle, \quad \forall q \in Q_{\epsilon}(x).$$

Aplicando este fato em (4.6), obtemos $f(x) - f^* \leq L\langle q, x - x^*\rangle^{\beta} + \epsilon, \quad \forall q \in Q_{\epsilon}(x)$ □

Considere o seguinte algoritmo:

Algoritmo 2. *Dado um ponto inicial $x^0 \in \mathbb{R}^n \backslash D^*$, construiremos a sequência $\{x^k\}$ da seguinte maneira*

$$x^{k+1} := x^k - \lambda_k g^k, \quad g^k \in Q(x^k), \quad \lambda_k > 0,$$

onde

$$\lambda_k = \gamma[(f(x^k) - f^*)/L]^{1/\beta}, \quad \gamma \in (0,2) \tag{4.7}$$

e L (respectivamente, β) é a constante (respectivamente, grau) na condição de $Hölder$ para a função f no ponto $x^* \in D^*$.

Teorema 29. *Suponha que f satisfaz a condição de $Hölder$ com constante L e grau β no ponto $x^* \in D^*$. Então*

$$||x^{k+1} - x^*||^2 \leq ||x^k - x^*||^2 - \gamma(2-\gamma)[(f(x^k) - f^*)/L]^{2/\beta}. \tag{4.8}$$

Demonstração. Pela definição do algoritmo, temos

$$||x^{k+1} - x^*||^2 = ||(x^k - \lambda_k q^k) - x^*||^2 \leq ||x^k - x^*||^2 - 2\lambda_k \langle q^k, x^k - x^* \rangle + {\lambda_k}^2.$$

Usando (4.7) e (4.5) temos

$$\begin{aligned} ||x^{k+1} - x^*||^2 &\leq ||x^k - x^*||^2 - 2\gamma \left(\frac{f(x^k) - f^*}{L} \right)^{2/\beta} + \gamma^2 \left(\frac{f(x^k) - f^*}{L} \right)^{2/\beta} \\ &\leq ||x^k - x^*||^2 - \gamma(2-\gamma) \left(\frac{f(x^k) - f^*}{L} \right)^{2/\beta}, \end{aligned}$$

que prova o teorema. □

Podemos substituir o espaço $\mathbb{R}^n$ na condição de $Hölder$ por um conjunto contendo a união $\cup_{k=0}^{\infty} G(x^k)$. Se f tem conjunto de nível limitado, esse conjunto será limitado também.

Agora estamos prontos para estabelecer um resultado de convergência.

Corolário 3. *Se todas as hipóteses do teorema 28 são satisfeitas, então*

i) $\lim_{k\to\infty} f(x^k) = f^*$;

ii) *A sequência $\{x^k\}$ tem ponto de acumulação e esses pontos de acumulação pertencem a D^*.*

Demonstração. Segue do teorma 28, na desigualdade (4.8), que a sequência $\{x^k\}$ é limitada, isto é, tem ponto de acumulação. Além disso, temos

$$\sum_{k=0}^{\infty} (f(x^k) - f^*)^{2/\beta} < \infty,$$

donde

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

Assim, se z é ponto de acumulação de $\{x^k\}$, então $f(z) = f^*$ pela continuidade de f. □

Observe que, sendo f quase-convexa, a continuidade segue do fato de f satisfazer a condição de $H\ddot{o}lder$. Note ainda que, a condição do teorema 28 afirma sobre a existência de uma majorante superior da função $f(x) - f^*$. Esta condição implica na convergência. Afim de obter estimativas de convergência, precisamos de funções majorante e minorante. Para este fim, introduziremos a seguinte condição.

Condição (G) *Dado um ponto $x^* \in D^*$, existem números $\alpha > 0$, $\beta > 0$, $l > 0$ e $L > 0$ tais que*

$$l||x - x^*||^{\alpha\beta} \leq f(x) - f^* \leq L||x - x^*||^{\beta} \;\; \forall x \in \mathbb{R}^n / D^*. \tag{4.9}$$

Teorema 30. *Suponha que a condição (G) se verifica. Então*

i)

$$||x^{k+1} - x^*||^2 \leq ||x^k - x^*||^2 - \gamma(2-\gamma)\left(\frac{l}{L}\right)^{2/\beta} ||x^k - x^*||^{2\alpha}; \tag{4.10}$$

ii) $\lim\limits_{k\to\infty} x^k = x^*$.

Demonstração. Sob as hipóteses deste teorema, temos que vale (4.8). Aplicando agora o lado esquerdo da inequação (4.9) temos

$$||x^{k+1} - x^*||^2 \leq ||x^k - x^*||^2 - \gamma(2-\gamma)\left(\frac{1}{L}\right)^{2/\beta} (l^{1/\beta}||x^k - x^*||^{\alpha})^2$$

$$\leq ||x^k - x^*||^2 - \gamma(2-\gamma)\left(\frac{l}{L}\right)^{2/\beta} ||x^k - x^*||^{2\alpha},$$

isto é, (4.10) se verifica. A afirmação (ii) segue obviamente de (4.10). □

Observação 3. *Na demonstração do teorema 30, que pode ser encontrada em [12], utlizamos a condição (G) para provarmos que $x^k \to x^*$. Contudo, observamos que essa condição pode ser retirada, pois por (4.8), temos*

$$||x^{k+1} - x^*||^2 \leq ||x^k - x^*||^2 - \gamma(2-\gamma)[(f(x^k) - f^*)/L]^{2/\beta}, \;\; \textit{com}\; \gamma \in (0,2).$$

Como $f^* \leq f(x^k)$, para todo $k \in \mathbb{N}$, temos que

$$\gamma(2-\gamma)[(f(x^k) - f^*)/L]^{2/\beta} \geq 0.$$

Logo,

$$||x^{k+1} - x^*||^2 \leq ||x^k - x^*||^2,$$

donde $\{x^k\}$ *é Féjer convergente a* D^**. Pelo corolário 3,* $\{x^k\}$ *tem ponto de acumulação em* D^**. Portanto, pela proposição 5,*

$$\lim_{k\to\infty} x^k = x^*.$$

O teorema também nos permite tirar várias taxas de convergência para o Algoritimo 2. Por exemplo, se o majorante e minorante da função $f(x) - f^*$ *coincidem, então o algoritmo converge para a solução geometricamente, entretanto, o grau* β *de ambas as funções pode ser arbitrário.*

Corolário 4. *Suponha que a condição (G) se verifica com* $\alpha = 1$*. Então existe um número* $\theta \in (0, 1)$*, tal que,*

$$||x^{k+1} - x^*|| \leq \theta ||x^k - x^*||.$$

Demonstração. De (4.10), com $\alpha = 1$, segue que

$$||x^{k+1} - x^*||^2 \leq \left[1 - \gamma(2-\gamma)\left(\frac{l}{L}\right)^{2/\beta}\right] ||x^k - x^*||^2,$$

assim tomando

$$\theta = \left[1 - \gamma(2-\gamma)\left(\frac{l}{L}\right)^{2/\beta}\right]^{1/2},$$

verifica-se o corolário. □

Portanto, se f *satisfaz a condição (G), com* $\alpha = 1$*, temos que o Algoritmo 2 converge geometricamente para a solução do problema* (1.4)*. Outros casos, para* $\alpha \neq 1$*, podem ser verificados em [12].*

4.3 Versões Inexatas

Nesta seção, consideraremos resultados de convergência para a versão inexata do método do subgradiente no Algoritmo 2, que admite computação inexata dos elementos de $Q(x)$ *ou uma avaliação inexata de* f^**.*

Primeiramente, considere o seguinte algoritimo ϵ*-subgradiente:*

Algoritmo 3

$$x^{k+1} := x^k - \lambda_k q^k, \quad q^k \in Q_{\epsilon_k}(x^k), \quad f(x^k) - f^* - \epsilon_k \geq 0 \tag{4.11}$$

$$\epsilon_k \geq 0, \quad \lambda_k = \gamma[(f(x^k) - f^* - \epsilon_k)/L]^{1/\beta}, \quad \gamma \in (0, 2). \tag{4.12}$$

Por simplicidade, considere $\Delta_k = f(x^k) - f^*$. Podemos estabelecer um resultado análogo ao corolário 3 para o Algoritimo 3.

Teorema 31. *Seja a sequência $\{x^k\}$ construída pelo Algoritimo 3. Suponha que f satisfaz a condição de $H\ddot{o}lder$ com constante L e grau β no ponto $x^* \in D^*$ e que*

$$\lim_{k\to\infty} \epsilon_k = 0. \tag{4.13}$$

Então:

i) $\lim_{k\to\infty} f(x^k) = f^*$;

ii) *A sequência $\{x^k\}$ tem ponto de acumulação e todos estes pontos pertencem a D^*.*

Demonstração. Usando (4.11) e (4.12), temos

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

Aplicando a proposição 6, temos

$$\begin{aligned} ||x^{k+1} - x^*||^2 &\leq ||x^k - x^*||^2 - 2\gamma\left(\frac{\Delta_k - \epsilon_k}{L}\right)^{2/\beta} + \gamma^2\left(\frac{\Delta_k - \epsilon_k}{L}\right)^{2/\beta} \\ &\leq ||x^k - x^*||^2 - \gamma(2-\gamma)\left(\frac{\Delta_k - \epsilon_k}{L}\right)^{2/\beta}. \end{aligned}$$

Donde segue que a sequência $\{x^k\}$ é limitada e que

$$\lim_{k\to\infty} (\Delta_k - \epsilon_k) = 0.$$

Por conta de (4.13), obtemos

$$\lim_{k\to\infty} \Delta_k = 0,$$

isto é, a firmação (i) é verdadeira. Além disso, a afirmação (ii) também é verdade pela continuidade da f. □

Portanto, o algoritimo ϵ-subgradiente converge para a solução sob as hipóteses como na seção 4.1.

Agora, consideraremos outro algoritimo que admite uma avaliação inexata de f^*. Por simplicidade, ficaremos restrito ao caso Lipschitz, isto é, f é uma função Lipschitz contínua em $\mathbb{R}^n$. A sequência será construída como segue:

Algoritmo 4

$$x^{k+1} := x^k - \lambda_k q^k, \quad q^k \in Q_{\epsilon_k}(x^k), \quad \epsilon_k \geq 0; \tag{4.14}$$

$$\lambda_k = \sigma_k / L, \quad ||\sigma_k - \Delta_k + \epsilon_k|| = \delta_k, \quad \sigma_k \geq 0. \tag{4.15}$$

Logo, o algoritimo admite estimativa superior e inferior de $\Delta_k - \epsilon_k$. Dessa forma, o algotitmo converge sob hipóteses adicionais bastantes suaves.

Teorema 32. *Seja a sequência $\{x^k\}$ construída pelo Algoritmo 4. Suponha que f é Lipschitz contínua com constante L e que*

$$\lim_{k \to \infty} \epsilon_k = 0, \quad \sum_{k=0}^{\infty} {\delta_k}^2 < \infty. \tag{4.16}$$

Então:

i) $\lim_{k \to \infty} f(x^k) = f^*$*;*

ii) *A sequência $\{x^k\}$ tem ponto de acumulação e todos estes pontos pertencem a D^*.*

Demonstração. Fixado um ponto $x^* \in D^*$, de (4.14) e (4.15) temos

$$||x^{k+1} - x^*||^2 \leq ||x^k - \lambda_k q^k - x^*||^2 = ||x^k - x^*||^2 - 2\lambda_k \langle q^k, x^k - x^* \rangle + {\lambda_k}^2.$$

Aplicando a proposição 6 temos

$$\begin{aligned} ||x^{k+1} - x^*||^2 &\leq ||x^k - x^*||^2 - \frac{2\sigma_k}{L^2}(\Delta_k - \epsilon_k) + \frac{{\sigma_k}^2}{L} \\ &= ||x^k - x^*||^2 - \frac{\sigma_k}{L^2}[2(\Delta_k - \epsilon_k) - \sigma_k] \\ &= ||x^k - x^*||^2 - \frac{1}{L^2}(\Delta_k - \epsilon_k - \delta_k)(\Delta_k - \epsilon_k + \delta_k) \\ &= ||x^k - x^*||^2 - \frac{1}{L^2}(\Delta_k - \epsilon_k)^2 + \frac{1}{L^2}{\delta_k}^2. \end{aligned}$$

De (4.16) vemos que a sequência $\{x^k\}$ deve ser limitada, logo tem ponto limite. Além disso, temos

$$\sum_{k=0}^{\infty} (\Delta_k - \epsilon_k)^2 < \infty,$$

então

$$\lim_{k \to \infty} \Delta_k = 0$$

por conta de (4.16). Portanto, a afirmação (i) é verdadeira. Novamente, a afirmação (ii) segue da continuidade da f. □

4.4 Implementações computacionais

Conforme vimos na seção 4.2, o algoritmo

$$x^{k+1} = x^k - \lambda_k g^k,$$

onde $\lambda_k = \gamma[(f(x^k) - f^*)/L]^{1/\beta}$, *com* $\gamma \in (0,2)$ *e* $g^k \in Q(x^k)$, *converve para a solução do problema* (1.4), *onde* f *é uma função quase-convexa e* $Hölder$, *mas não necessariamente diferenciável. Nesta seção, utilizaremos o software livre Scilab 4.1.2 em sua versão Linux, para ilustarmos alguns exemplos de implementação deste algoritmo.*

Considere a função

$$f(x) = \begin{cases} -9x + 54, & \text{se } x \leq -3 \\ x^3, & \text{se } -3 < x < 3 \\ 9x, & \text{se } x \geq 3 \end{cases} .$$

A função f, *assim definida é quase-convexa, pois possui todos os conjuntos de níveis convexos, é* $Hölder$, *com constante* $L = 27$ *e grau* $\beta = 1$. *Conforme podemos observar na figura abaixo,* f *não é diferenciável em* $x = -3$ *e* $x = 3$, *e* $x = 0$ *é ponto estacionário, mas não é ponto de mínimo.*

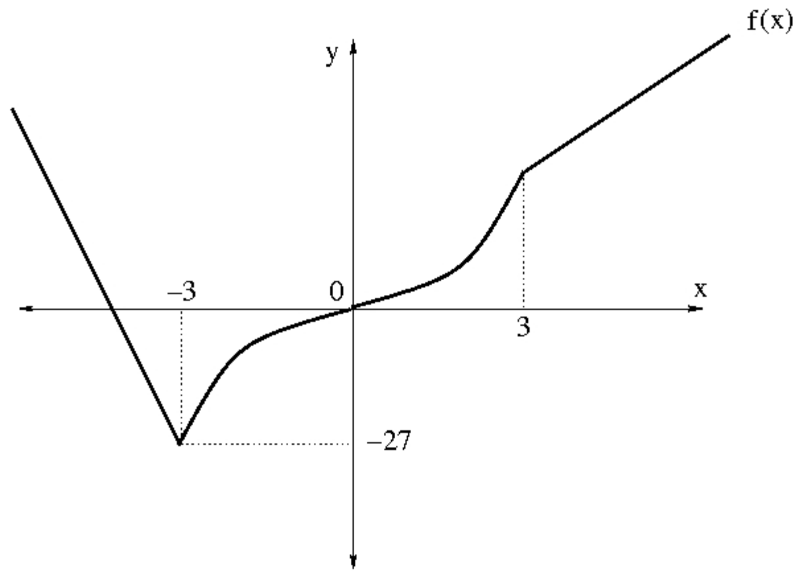

Figura 4.1: Gráfico da função f.

Dessa forma, iniciaremos a implementação do algoritmo com o ponto $x^0 = 4$*, para observar que o ponto onde* f *não é diferenciável e o ponto estacionário não interferem na convergência. Neste exemplo, temos que* $Q(x^k) = \{1\}$*, para todo* $k \in \mathbb{N}$*, ou seja,* $q^k = 1$*,* $\forall k \in \mathbb{N}$*. Sendo* $f^* = -27$ *e tomando* $\gamma = 0,2$*, aplicando estes dados no software mencionado, com o critério de parada* $|x^k + 3| < 10^{-20}$*, obtemos o seguinte resultado:*

k	x^k	$f(x^k)$
0	4	36
1	3,5333333	31,7999997
10	0,8907814	0,7068274
20	-1,1081916	-1,3609574
30	-2,5568934	-16,716212
40	-2,9436257	-25,506316
50	-2,9938181	-26,833433
60	-2,9993347	-26,982041
70	-2,9999285	-26,998071
80	-2,9999923	-26,999793
90	-2,9999992	-26,999978
100	-2,9999999	-26,999998

A dificuldade de implementação do método está em encontrar a constante, o grau de $Hölder$ *e o cone normal do conjunto de nível da função. Em alguns casos, a computação desses elementos pode ser complicada.*

Como os mesmos dados, mudando somente $\gamma = 0,7$*, obtemos agora o seguinte resultado:*

k	x^k	$f(x^k)$
0	*4*	*36*
1	*2,3666667*	*13,255963*
2	*1,3229936*	*2,3156514*
3	*0,5629581*	*0,17841384*
4	*-0,1416674*	*-0,0028432*
6	*-1,5261397*	*-3,5545353*
7	*-2,133985*	*-9,7179377*
8	*-2,5820385*	*-17,214251*
10	*-2,9445424*	*-25,530155*
18	*-2,9999961*	*-26,999896*
19	*-2,9999988*	*-26,999969*
22	*-3*	*-27*

Fazendo algumas implementações, para diferentes valores de γ, observamos que a convergência com o menor número de iterações ocorre quando $\gamma = 1$, conforme quadro abaixo:

k	x^k	$f(x^k)$
0	*4*	*36*
1	*1,6666667*	*4,6296296*
2	*0,4951989*	*0,1214336*
3	*-0,5092986*	*-0,1321045*
4	*-1,5044059*	*-3,4048271*
5	*-2,3783012*	*-13,452424*
6	*-2,8800632*	*-23,889446*
7	*-2,995269*	*-26,872463*
8	*-2,9999925*	*-26,999799*
9	*-3*	*-27*

Referências Bibliográficas

[1] *Lima, E.L., Curso de análise vol. 2.* Projeto Euclides, *IMPA, Rio de Janeiro, Nona edição, 2006.*

[2] *Izmailov, A., Solodov, M., Otimização - volume 1 - Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade, IMPA, Rio de Janeiro. 2005.*

[3] *Izmailov, A., Solodov, M., Otimização - volume 2 - Métodos computacionais, IMPA, Rio de Janeiro. 2007.*

[4] *Tiel, J. van, Convex Analysis - An Introductory Text, New York. 1984.*

[5] *Bazaraa, M.S., Sherali, H.D., Shetty, C.M., Nonlinear Programming - Theory and Algorithms, New York, Segunda Edição, 1993.*

[6] *Mangasarian, O.L., Nonlinear Programming, New York,1969.*

[7] *Hiriart-Urruty, J.B., Lemaréchal, C., Convex Analysis and Minimization Algotithms I, New York, Segunda Edição, 1996.*

[8] *Hiriart-Urruty, J.B., Lemaréchal, C., Convex Analysis and Minimization Algotithms II, New York, Segunda Edição, 1996.*

[9] *Rockafellar, R.T., Convex Analysis, Princeton University, Princeton, 1970.*

[10] *Oliveira, P.R., Introdução à Programação Não Linear, Escola Brasileira de Otimização, Rio de Janeiro, 1988.*

[11] *Burachik, R.S., Generalized proximal point methods for the variational inequality problem, PhD Thesis, IMPA, Rio de Janeiro, 1995.*

[12] *Konnov, I.V., On convergence properties of a subgradient method, Otimization Methods and Soft, 18, 53-62, 2003.*

[13] Plastria,F., Lower subdifferenciable functions and their minimization by cutting planes, J. Optim. Theory and Appl., 46, 37-53, 1985.

[14] Censor, Y., Segal, A.N., Algorithms for the quasiconvex fasibility problem, Journal of Computational and Applied Mathematics, 185, 34-50, 2006.

[15] Da Cruz Neto, J.X., Lopes, J.O., Travaglia, M.V., Algorithms for quasiconvex mimimization, Preprint, 2009.

[16] Burachik, R.S., Graña Drummond, L.M., Iusem, A.N., Svaiter, B.F., Full convergence of the steepest descent method with inexact searches, Optimization, 32, 137-146, 1995.

[17] Iusem, A.N., On convergence properties of the projected gradient method for convex optimization, Journal of Computational and Applied Mathematics, 22, 37-52, 2003.

[18] Luc, D.T., Caracterizations of quasiconvex functions, Bull. Austral. Math. Soc., 48, 393-406, 1993.

[19] Da Cruz Neto, J.X., Oliveira, P.R. Soubeyran, A., Souza, S.S., A proximal method with separable Bregman distances for quasiconvex minimization over the nonnegative orthant, European Journal of Operational Research., 201, 365-376, 2010.

[20] Allen, E., Helgason, R., Kennington, J., Shetty, B., A generalization of Polyak's convergence result for subgradient optimization, Math. Progr., 37, 309-317, 1987.

[21] Kim, S., Ahn, H., Cho, S.-C, Variable target value subgradient method. Math Progr., 49, 359-379, 1990.