UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM
INSTITUTO DE CIÊNCIAS EXATAS - ICE
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

CONVERGÊNCIA GLOBAL DO MÉTODO DE DESCIDA PARA MINIMIZAÇAO DE FUNÇÕES QUASECONVEXAS

Sílvia Dias de Souza

MANAUS - 2008

UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM
INSTITUTO DE CIÊNCIAS EXATAS - ICE
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Sílvia Dias de Souza

CONVERGÊNCIA DO MÉTODO DE DESCIDA PARA MINIMIZAÇAO DE FUNÇÕES QUASECONVEXAS

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Matemática, na área de concentração em Otimização.

Orientador: Prof^o. Dr. Roberto Cristóvão Mesquita Silva

MANAUS - 2008

Sílvia Dias de Souza

CONVERGÊNCIA DO MÉTODO DE DESCIDA PARA MINIMIZAÇÃO DE FUNÇÕES QUASECONVEXAS

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Matemática, na área de concentração em Otimização.

Manaus, 02 de Dezembro de 2008.

BANCA EXAMINADORA

....

Prof^o Dr. Roberto Cristóvão Mesquita Silva, Presidente

Universidade Federal do Amazonas

....

Prof^o Dr. Mário Salvatierra Junior, Membro

Universidade Federal do Amazonas

....

Prof^o Dr. Orizon Pereira Ferreira, Membro

Universidade Federal de Goiás

AGRADECIMENTOS

Sempre a Deus por essa e todas as conquistas que tive em minha vida.

Ao maior presente que uma pessoa pode ter, a família: a minha mãe, meus irmãos, meus sobrinhos que enchem minha vida de alegria .

Ao meu amado Celso Rômulo por toda a atenção, compreensão e apóio incondicional.

A Allan Alves (in memorian).

Aos meus amigos, pelo companherismo e toda a paciência com que sempre me trataram, em especial minhas amigas de mestrado, Sílvia Viviane e Jeanne Moreira.

Aos meus queridos professores, por todo o conhecimento que me ensinaram, especialmente os da área de Otimização: Professor Mário Salvatierra, Flávia Morgana e Alfredo Wagner.

Ao meu orientador Roberto Cristóvão Mesquita Silva, que conduziu minha caminhada ao longo do mestrado, com responsabilidade, seriedade e profissionalismo e ao Professor Orizon Pereira Ferreira por dispensar tempo e me dar a honra de estar na banca examinadora de minha defesa.

RESUMO

CONVERGÊNCIA DO MÉTODO DE DESCIDA PARA MINIMIZAÇÃO DE FUNÇÕES QUASECONVEXAS

Neste trabalho, trazemos uma demonstração detalhada dos resultados obtidos por Kiwiel e Murty [4]. Mostraremos um forte resultado de convergência do método de descida para funções continuamente diferenciáveis e quaseconvexas, faremos discussões e comparações com outras linhas de pesquisas e com resultados já existentes. Poderemos observar também que este trabalho é bastante relevante pois gereraliza o de Burachik *et al.* [2] e é usado como base fundamental para outros artigos.

ABSTRACT

CONVERGENCE OF THE STEEPEST DESCENT METHOD FOR QUASICONVEX FUNCTIONS MINIMIZATION

In this work we present a detailed proof of the results obtained by Kiwiel e Murty [4]. We prove a strong result about the convergence of the steepest descent method for quasiconvex and continuously differentiable functions, proceed discussions and comparisons with other research lines and existing results. We can note that this work has great relevance because it generalizes an earlier work of Burachik *et al.* [2] and is used as the fundamental basis in other articles.

Sumário

Introdução

Em 1995, Burachik *et al.* [2], provaram a convergência completa do método de descida com busca linear exata e inexata para funções psedo-convexas e convexas, usando para isto dois algoritmos onde um deles exige que a função além de todas as hipóteses já citadas, possua o gradiente satisfazendo a condição de lipschitz e que se conheça essa constante. Verificamos então a exigência de hipóteses muito fortes para que o resultado seja válido.

Este trabalho é baseado no artigo de Kiwiel e Murty [4], onde é demonstrado a convergência do método de descida gradiente com busca linear pela regra de Armijo para funções quaseconvexas, sendo $f : \mathbb{R}^n \to \mathbb{R}$ e, se pede como hipótese que a função seja continuamente diferenciável. Algumas hipóteses são consideradas e citadas neste trabalho e que inclusive generalizam a condição do parágrafo acima. O teorema principal é demonstrado de maneira simples depois que se demonstra os seguintes lemas :

Lema 1 : Se $f(x^k) \geq f(\widetilde{x})$, para algum $\widetilde{x}$ fixo e para todo k, então

$$\sum_{k=0}^{\infty} t_k^2 \|g^k\|^2 \leq [f(x^0) - f(\widetilde{x})]/\beta.$$

Com a sequência $\{x^k\}$ gerada pelo método do gradiente, t_k o tamanho de passo e $g^k = \nabla f(x^k)$.

Além disso, provaremos que $x^k \to \overline{x}$ para algum $\overline{x}$.

Lema 2 : Se $\overline{x}$ é um ponto de acumulação de $\{x^k\}$, então temos que $\overline{x} \in X = \{x : \nabla f(x) = 0\}$, isto é, $\nabla f(\overline{x}) = 0$.

Demonstrado esses dois lemas temos então o teorema principal:

Teorema principal :(Convergência Global). Considerando o conjunto de pontos de mínimo da f como sendo o conjunto $\check{X}$, temos :

$$x^k \to \overline{x} \in X = \{x : \nabla f(x) = 0\}, \text{ ou } \check{X} = \emptyset, \; \|x^k\| \to \infty, \text{ e } f(x^k) \to \inf f.$$

Além disso, separamos um capítulo para fazermos discurssões em cima do resultado demonstrado, considerando algumas outras linhas de pesquisa. Também comparamos este trabalho com os resultados citados no primeiro parágrafo, mostrando que o artigo publicado por Burachik *et al.* [2], pode ser provado usando as hipóteses deste trabalho, tornando-o na realidade um caso particular deste e portanto mais simples de ser demonstrado, verificamos que o artigo escrito por Kiwiel e Murty é na realidade um resultado mais forte que o artigo escrito por Burachik *et al.* [2].

Capítulo 1

Generalidades

Neste capítulo apresentaremos as definições e os resultados da teoria de funções quaseconvexas, regra de Armijo, e algumas definições importantes para desenvolvimento deste trabalho, e fixaremos a notação a ser utilizada nos capítulos posteriores. Algumas demonstrações serão inseridas e outras apenas referenciadas.

1.1 Definições

Definição 1.1.1. (Funções Quaseconvexas) ([6]) $\theta : \Gamma \subset \mathbb{R}^n \to \mathbb{R}$ é dita uma função quaseconvexa sobre $\overline{x} \in \Gamma$ (com respeito a Γ), se para cada x tal que $\theta(x) \leq \theta(\overline{x})$, a função θ assume um valor não maior que $\theta(\overline{x})$ sobre cada ponto da intersecção do segmento fechado $[\overline{x}, x]$, ou equivalentemente*:*

$$\begin{cases} x \in \Gamma \\ \theta(x) \leq \theta(\overline{x}) \\ 0 \leq \lambda \leq 1 \\ (1-\lambda)\overline{x} + \lambda x \in \Gamma \end{cases}$$

implica em $\theta[(1-\lambda)\overline{x} + \lambda x] \leq \theta(\overline{x})$.

Note que esta definição de uma função quaseconvexa é um pouco mais geral que a definição habitual, pois :

i. Definimos quaseconvexidade primeiro sobre um ponto e então depois sobre um conjunto;

ii. Não necessita-se que Γ seja convexo. Esta generalização permite muitas vezes se trabalhar com uma classe de problemas um pouco mais ampla, pois não exigi-se a convexidade do conjunto em questão.

Segue da definição acima que se o conjunto Γ for convexo, a função θ é quaseconvexa sobre Γ, se, e somente se :

$$\begin{cases} x^1, x^2 \in \Gamma \\ \theta(x^2) \leq \theta(x^1) \\ 0 \leq \lambda \leq 1 \end{cases}$$

implicando em $\theta[(1-\lambda)x^1 + \lambda x^2] \leq \theta(x^1)$.

Podemos utilizar ainda a seguinte definição abaixo de acordo com [10]:

Definição 1.1.2. (Funções Quaseconvexas) ([9]) Dizemos que uma função $\theta : \Gamma \subset \mathbb{R}^n \to \mathbb{R}$ é quaseconvexa em um conjunto convexo Γ se, e somente se,

$$\theta(\alpha x + (1-\alpha)y) \leq max\{\theta(x), \theta(y)\}, \forall x, y \in \Gamma, \forall \alpha \in [0,1].$$

Podemos observar que esta definição é um caso particular da definição anterior onde dizemos ser o caso mais geral, pois se tivermos que $\theta(x) \leq \theta(y)$ teremos que o máximo será $\theta(y)$ e então temos a situação do caso mais geral.

Definição 1.1.3. (Funções Pseudo-Convexas) Consideremos a função $\theta : D \subset \mathbb{R}^n \to \mathbb{R}$, com D aberto e contendo Γ. θ é dita pseudo-convexa sobre $\overline{x} \in \Gamma$ (com respeito a Γ), se é diferenciável sobre $\overline{x}$ e dado $x \in \Gamma$ tal que $\nabla\theta(\overline{x})(x - \overline{x}) \geq 0$, implicar $\theta(x) \geq \theta(\overline{x})$.
θ é dita pseudo-convexa sobre Γ, se é pseudo-convexa sobre cada $x \in \Gamma$.

Definição 1.1.4. (Sequência Fejér convergente) Dizemos que uma sequência $\{y^k\} \subset \mathbb{R}^n$ é Fejér convergente a um conjunto $U \neq \emptyset$, $U \subset \mathbb{R}^n$, se

$$\|y^{k+1} - u\| \leq \|y^k - u\|,$$

para todo $k \geq 0$ e todo $u \in U$.

Definição 1.1.5. (Sequência quase Fejér convergente) Uma sequência $\{x^k\} \subset \mathbb{R}^n$ é quase Fejér convergente a um conjunto $\Gamma \subseteq \mathbb{R}^n$ se para todo $u \in \Gamma$, existe uma sequência $\{\xi_k\} \subseteq \mathbb{R}$, tal que

$$\xi_k \geq 0,\ \sum_{k=0}^{\infty} \xi_k < \infty,\ \|x^{k+1} - u\|^2 \leq \|x^k - u\|^2 + \xi_k,\ \forall\, k.$$

Definição 1.1.6. (Diferencial de uma função) Seja $f : U \to \mathbb{R}$ definida no aberto $U \subset \mathbb{R}^n$, diferenciável no ponto $a \in U$. A diferencial de f no ponto a é o funcional linear $df(a) : \mathbb{R}^n \to \mathbb{R}$, cujo valor no vetor $v = (\alpha_1, ..., \alpha_n)$ é dado por

$$df(a).v = \frac{\partial f}{\partial v}(a) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a).\alpha_i.$$

Quando $f : \mathbb{U} \rightarrow \mathbb{R}$ é diferenciável em todo ponto de U, obtemos uma aplicação $df : \mathbb{U} \rightarrow \mathfrak{L}(\mathbb{R}^n; \mathbb{R})$, que associa a cada ponto $x \in \mathbb{U}$ o funcional $df(x)$, cuja matriz é

$$\Big(\frac{\partial f}{\partial x_1}(x), ..., \frac{\partial f}{\partial x_n}(x)\Big).$$

A aplicação df é contínua se, e somente se, cada uma das suas funções coordenadas $\partial f/\partial x_i : U \rightarrow \mathbb{R}$ é contínua, isto é, se, e somente se, f é de classe C^1.

Definição 1.1.7. (Gradiente de uma função diferenciável) Dada a função diferenciável $f : U \rightarrow \mathbb{R}$, definida no aberto $U \subset \mathbb{R}^n$, temos que o gradiente de f em $a \in U$ é o vetor $\nabla f(a)$ (ou grad $f(a)$), definido como,

$$\nabla f(a) = \Big(\frac{\partial f}{\partial x_1}(a), ..., \frac{\partial f}{\partial x_n}(a)\Big).$$

Definição 1.1.8. (Direção de Descida) Dizemos que $d \in \mathbb{R}^n$ é uma direção de descida de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ no ponto $\overline{x} \in \mathbb{R}^n$, se existe $\varepsilon > 0$ tal que

$$f(\overline{x} + td) < f(\overline{x}), \forall t \in (0, \varepsilon].$$

1.2 Teoremas

Teorema 1.2.1. (Teorema do Valor Médio -T.V.M.)([6], Teorema 2, pp. 204). Seja a função $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$, diferenciável sobre o conjunto aberto e convexo U e sejam $x^1, x^2 \in U$. Então,

$$f(x^2) - f(x^1) = df[(x^1 + \gamma(x^2 - x^1))](x^2 - x^1),$$

para algum número real γ, $0 < \gamma < 1$.

Teorema 1.2.2. Seja a função $\theta : \Gamma \subset \mathbb{R}^n \to \mathbb{R}$, definida no conjunto aberto Γ. Considere $\overline{x} \in \Gamma$.

Se θ é diferenciável em $\overline{x}$, então θ é contínua em $\overline{x}$, $\nabla\theta(\overline{x})$ existe, e

$$\theta(\overline{x} + x) = \theta(\overline{x}) + \langle \nabla\theta(\overline{x}), x \rangle + \alpha(\overline{x}, x)\|x\|.$$

Onde α é uma função de x tal que $\lim_{x \to 0} \alpha(\overline{x}, x) = 0$.
Para $\overline{x} + x \in \Gamma$.([6], Teorema 3, pp.201).

Teorema 1.2.3. Seja Γ um conjunto aberto do $\mathbb{R}^n$, e considere a função $\theta : \Gamma \to \mathbb{R}$. Dados $x^1, x^2 \in \Gamma$, θ diferenciável e quaseconvexa sobre x^1, se $\theta(x^2) \leq \theta(x^1)$ então $\nabla\theta(x^1)(x^2 - x^1) \leq 0$.

Demonstração: Se $x^1 = x^2$, é óbvia a demonstração.

Tome $x^1 \neq x^2$. Como Γ é aberto, existe uma bola aberta $B_\delta(x^1)$, de centro em x_1 e raio δ, que está contida em Γ. Então para algum $\overline{\mu}$ tal que $0 < \overline{\mu} < 1$ e $\overline{\mu} < \frac{\delta}{\|x^2 - x^1\|}$, nós temos :

$$\widetilde{x} = x^1 + \overline{\mu}(x^2 - x^1) = (1 - \overline{\mu})x^1 + \overline{\mu}x^2 \in B_\delta(x^1) \subset \Gamma,$$

pois,

$$\|\widetilde{x} - x^1\| = \overline{\mu}\|x^2 - x^1\| < \frac{\delta}{\|x^2 - x^1\|}.\|x^2 - x^1\| = \delta.$$

Então como θ é quaseconvexa sobre x^1, temos por definição que

$$\theta(x^2) \leq \theta(x^1) \Rightarrow \theta(\widetilde{x}) \leq \theta(x^1),$$

e como $\widetilde{x}, x^1 \in B_\delta(x^1)$, e pela convexidade de $B_\delta(x^1) \subset \Gamma$, para $0 < \lambda \leq 1$ obtemos,

$$\theta[(1 - \lambda)x^1 + \lambda\widetilde{x}] \leq \theta(x^1).$$

Usando a hipótese de que θ é diferenciável sobre $x^1 \in \Gamma$, teremos que pelo Teorema 1.2.2 temos,

$$\theta(x^1+\lambda(\widetilde{x}-x^1)) = \theta(x^1)+\langle\nabla\theta(x^1),\lambda(\widetilde{x}-x^1)\rangle+\alpha(x^1,\lambda(\widetilde{x}-x^1)).\lambda\|\widetilde{x}-x^1\| \leq \theta(x^1),$$

onde $\lim\limits_{\lambda\to 0}\alpha(x^1,\lambda(\widetilde{x}-x^1)) = 0$.

Dividindo a inequação por λ, temos:

$$\langle\nabla\theta(x^1),\widetilde{x}-x^1\rangle + \alpha(x^1,\lambda(\widetilde{x}-x^1))\|\widetilde{x}-x^1\| \leq 0.$$

Para $0 < \lambda \leq 1$. Passando o limite com $\lambda \to 0$ e lembrando que temos $\widetilde{x}-x^1 = \overline{\mu}(x^2-x^1)$, obtemos:

$$\langle\nabla\theta(x^1),\widetilde{x}-x^1\rangle \leq 0 \Rightarrow \overline{\mu}\langle\nabla\theta(x^1),x^2-x^1\rangle \leq 0.$$

Dividindo ambos os lados da segunda desigualdade por $\overline{\mu}$, temos:
$\langle\nabla\theta(x^1),x^2-x^1\rangle \leq 0$. □

Teorema 1.2.4. Seja o conjunto aberto $D \subset \mathbb{R}^n$, contendo o conjunto Γ, e a função $\theta : D \to \mathbb{R}$. Dado $\overline{x} \in \Gamma$:

i. Se Γ um conjunto convexo, θ diferenciável sobre $\overline{x}$ e $\theta(\overline{x}) = min_{x\in\Gamma}\theta(x)$, então

$$\nabla\theta(\overline{x})(x-\overline{x}) \geq 0, \forall\, x \in \Gamma.$$

ii. Se θ pseudo-convexa sobre $\overline{x}$, e $\forall\, x \in \Gamma, \nabla\theta(\overline{x})(x-\overline{x}) \geq 0$, então

$$\theta(\overline{x}) = min_{x\in\Gamma}\theta(x).$$

Demonstração: Vamos demonstrar apenas o item [ii], pois ele será usado posteriormente, para a demonstração do item [i] veja [6], teorema 3, pp.142. Temos que θ é pseudo-convexa sobre $\overline{x} \in \Gamma$, então por definição, para qualquer $x \in \Gamma$ e tendo $\nabla\theta(\overline{x})(x-\overline{x}) \geq 0$, temos que $\theta(x) \geq \theta(\overline{x})$, logo $\theta(\overline{x}) = min_{x\in\Gamma}\theta(x)$. □

Teorema 1.2.5. Toda sequência monótona limitada de números reais é convergente.([5], Teorema 4, pg. 25.)

Corolário 1.2.1. Uma sequência monótona de números reais é convergente se, e somente se, possui uma subsequência limitada.

Teorema 1.2.6. Se $\{x^k\}$ é quase Ferjér convergente para um conjunto não vazio $\Gamma \subseteq \mathbb{R}^n$, então $\{x^k\}$ é limitada. Se além disso, um ponto de acumulação $\overline{x}$ de $\{x^k\}$ pertence a Γ, então $\lim\limits_{k\to\infty} x^k = \overline{x}$.

Demonstração: Tome $\widetilde{x} \in \Gamma$, aplicando a definição da quase Fejér convergência de $\{x^k\}$ iterativamente, obtemos:

$$\|x^k - \widetilde{x}\|^2 \leq \|x^{k-1} - \widetilde{x}\| + \xi_{k-1} \leq ... \leq \|x^0 - \widetilde{x}\|^2 + \xi_0$$
$$\leq \|x^0 - \widetilde{x}\|^2 \ + \sum_{l=0}^{k-1} \xi_l \leq \|x^0 - \widetilde{x}\|^2 + \sum_{k=0}^{\infty} \xi_k$$

Deste modo, temos que $\{x^k\}$ é limitada (está contida dentro de uma bola de centro $\widetilde{x}$, e raio $\|x^0 - \widetilde{x}\|^2 + \sum\limits_{k=0}^{\infty} \xi_k$).

Seja $\overline{x}$ um ponto de acumulação de $\{x^k\}$, e tome $\delta > 0$. Tome $\{x^{l_k}\}$ uma subsequência de $\{x^k\}$ tal que $\lim_{k\to\infty} x^{l_k} = \overline{x}$.

Assim, tomando $\overline{x} \in \Gamma$, existe $\{\xi_k\}$ satisfazendo a Definição 1.1.5 (quase Fejér convergência).

Tome k_0 tal que $\sum\limits_{k=k_0}^{\infty} \xi_k \leq \dfrac{\delta}{2}$, $\overline{k}$ tal que $l_{\overline{k}} \geq k_0$, e $\|x^{l_{\overline{k}}} - \overline{x}\|^2 \leq \frac{\delta}{2}$.

Então, para qualquer $k \geq l_{\overline{k}}$ temos :

$$\|x^k - \overline{x}\|^2 \leq \|x^{l_{\overline{k}}} - \overline{x}\|^2 + \xi_{l_{\overline{k}}} \leq \frac{\delta}{2} + \sum_{i=k_0}^{\infty} \xi_i \leq \frac{\delta}{2} + \frac{\delta}{2} = \delta.$$

Como δ é arbitrário, segue-se que $\lim\limits_{k\to\infty} x^k = \overline{x}$. $\square$

Teorema 1.2.7. ([3]) Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função continuamente diferenciável e seja

$$\alpha_k = argmin\{f(x^k - \alpha\nabla f(x^k)) + \alpha^2\lambda_k\|\nabla f(x^k)\|^2, \alpha \geq 0\}$$

e

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

Se f é uma função pseudo-convexa e atinge seu mínimo sobre $\mathbb{R}^n$, então a sequência gerada por α_k e x^{k+1} converge para um mínimo da f.

Teorema 1.2.8. ([5], Teorema 16, pp. 137): Seja $a_n \geq 0$ para todo $n \in \mathbb{N}$. A série $\sum a_n$ converge se, e somente se, as reduzidas $S_n = a_1 + ... + a_n$ formam uma sequência limitada, isto é, se, e somente se, existe $K > 0$ tal que $a_1 + ... + a_n < K$ para todo $n \in \mathbb{N}$.

1.3 Regra de Armijo

Nesta seção, falaremos sobre a regra de Armijo para o cálculo do tamanho de passo (para maiores detalhes veja [1] e [9]).

Uma estratégia natural para o cálculo do tamanho de passo para um método descida é obtermos um comprimento de passo que resulte em um decréscimo suficiente da função em relação a iterada seguinte. A idéia é que isto pode ser bem mais econômico do ponto de vista computacional, do que métodos onde se procure obter o máximo decréscimo da função, e ainda assim, possamos garantir sua convergência.

Suponhamos que f seja uma função diferenciável no ponto x^k. Fixamos os parâmetros $\widehat{\alpha} > 0$, e $\sigma, \theta \in (0, 1)$, e seja d^k uma direção de descida. Tomamos

$\alpha := \widehat{\alpha}$.

1. Verificamos se a desigualdade

$$f(x^k + \alpha d^k) \leq f(x^k) + \sigma\alpha\langle \nabla f(x^k), d^k\rangle \tag{1.1}$$

se satisfaz ou não.

2. Se (1.1) não é satisfeita, tomamos $\alpha := \theta\alpha$ e retornamos ao Passo 1.

Caso contrário, aceitamos $\alpha^k = \alpha$, como o valor do comprimento de Passo.

Em outras palavras, α^k é o maior número entre todos os números da forma $\alpha = \widehat{\alpha}\theta^i$, i=0,1,2..., que satisfaz a desigualdade (1.1).

O valor $\alpha\langle \nabla f(x^k), d^k\rangle$ no lado direito de (1.1), representa o valor de redução de f, previsto pela sua aproximação linear para o passo de comprimento α na direção d^k. Portanto, a desigualdade (1.1) significa que o decréscimo real de f deve ser pelo menos a fração (determinada por $\sigma \in (0,1)$) do previsto.

A implementação computacional da regra de Armijo é fácil : reduzimos o valor inicial $\widehat{\alpha}$, multiplicando ele cada vez pelo parâmetro $\theta \in (0,1)$, até que a desigualdade (1.1) seja satisfeita. A demonstração de que a regra de Armijo está bem definida pode ser encontrada na referência [9], Lema 3.1.2, pp. 66.

1.4 Métodos Computacionais

Nesta seção descreveremos alguns métodos computacionais (para maiores detalhes veja [9], e [7]) que utilizaremos posteriormente.

1.4.1 Método do Gradiente: Suponhamos que a função f seja diferenciável no $\mathbb{R}^n$, e seja o esquema iterativo geral de métodos de descida $x^{k+1} = x^k + \alpha_k d^k$, falamos do método do gradiente quando a direção de descida d^k se escolhe como o anti-gradiente da função f no ponto x^k, isto é, $d^k = -\nabla f(x^k)$. Se $\nabla f(x^k) = 0$ para algum k, x^k é um ponto estacionário do problema $min f(x), x \in \mathbb{R}^n$, e o método pára (no entanto, para fins de análise de convergência pode ser conveniente considerar que $x^j = ... = x^{k+1} = x^k$, para todo $j \geq k$, o que é válido, já que $x^{k+1} = x^k - \alpha_k \nabla f(x^k) = x^k$). Em particular, o esquema de descida se reduz a :

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k), \quad k = 0, 1, 2, ...,$$

para uma escolha conveniente do passo α_k.

1.4.2 Método da Bisseção I: Este método é descrito nas literaturas mais antigas e é usado para resolver problemas de se minimizar uma função f real definida em um aberto $X \supset [a, b]$, continuamente diferenciável. Para encontrarmos o mínimo, iniciaremos com intervalo $[a, b]$ satisfazendo $f'(a) < 0$ e $f'(b) > 0$. Este método será exposto de acordo com [7]. Avaliaremos $f'(a + b/2)$.

Se $f'[(a + b/2)] = 0$, o ponto $(a + b)/2$ satisfaz a condição necessária de primeira ordem para mínimo local.

Se $f'(a+b/2) > 0$, tome $[a, a+b/2]$ como o novo intervalo e avaliaremos novamente a f' no ponto médio deste intervalo.

Se $f'[(a + b/2)] < 0$, tomamos como o novo intervalo $[a + b/2, b]$ e avaliaremos novamente a f'.

Desde que o intervalo seja dividido pela metade a cada nova avaliação, o comprimento deste intervalo converge para zero linearmente. Quando seu comprimento se torna menos que uma tolerância especificada ϵ, qualquer ponto no intervalo final poderia ser levado a uma aproximação do mínimo. Uma desvantagem desse método é que ele confia totalmente nos valores da derivada e não usa valores da função f na minimização.

1.4.3 Método da Bisseção II: Apresentaremos agora um método da bisseção que encontramos nas literaturas mais atuais para resolver o problema :

$$\begin{cases} min f(x) \\ s.a\, x \in [a, b] \end{cases}$$

onde $a, b \in \mathbb{R}, a < b, f : [a, b] \to \mathbb{R}$. Este método é sequencial e relativamente eficiente. Ele permite aproximar uma solução global do problema (e, como consequência, também o valor ótimo), mas se aplicam a uma classe de estrutura mais específica.

Dizemos que uma função $f : [a, b] \to \mathbb{R}$, é `unimodal` (em [a,b]), quando ela tem um único minimizador global $\overline{x}$ em [a,b], e é estritamente decrescente em $[a, \overline{x}]$ e estritamente crescente em $[\overline{x}, b]$, isto é,

$$f(y) > f(z),\ \forall\, y, z \in\ [a, \overline{x}], y < z,$$

$$f(y) < f(z),\ \forall\, y, z \in\ [\overline{x}, b], y < z.$$

Precisaremos ainda de um lema que enunciaremos para o desenvolvimento do método da bisseção:

Lema: Sejam $f : [a,b] \to \mathbb{R}$ uma função unimodal e $\overline{x}$ o minimizador global de f em $[a,b]$. Segue-se que para quaisquer pontos $y, z \in [a,b]$ tais que $y < z$, vale o seguinte :

(a) Se $f(y) \leq f(z)$, então $\overline{x} \in [a,z]$.

(b) Se $f(y) \geq f(z)$, então $\overline{x} \in [y,b]$.

Como consequência, comparando os valores de uma função unimodal f em dois pontos de $[a,b]$, podemos "localizar"a solução global do problema, $minf(x)$, sujeito a $x \in [a,b]$, num intervalo menor. Repetindo este procedimento, podemos diminuir cada vez mais o intervalo que contém a solução. Os métodos baseados nesta estratégia podem ser diferentes na maneira de escolher pontos para a comparação. Ressaltamos que estamos aproximando ao mesmo tempo o valor ótimo do problema e a solução correspondente.

O algoritmo deste método pode ser encontrado na ref. [9]

Capítulo 2

Convergência Global do Método de Descida Gradiente para Funções Quaseconvexas

Neste capítulo vamos provar a convergência global do método de descida. Para isto, consideraremos uma função $f : \mathbb{R}^n \to \mathbb{R}$ quaseconvexa e continuamente diferenciável, algumas hipóteses adicionais e dois resultados auxiliares. Ressaltamos que este é o resultado principal deste trabalho.

2.1 Hipóteses Preliminares

O método de descida de Cauchy ou também conhecido como do gradiente, gera uma sequência x^k via,

$$x^{k+1} = x^k - t_k g^k,\ g^k = \nabla f(x^k),\ k = 0, 1, 2, ... \tag{2.1}$$

onde,

$$t_k = argmax\{t : f(x^k - tg^k) \leq f(x^k) - \alpha t\|g^k\|^2\},\ t = 2^{-i}, i = 0, 1, 2, ... \quad (2.2)$$

é obtido pela regra de Armijo, com $\alpha \in (0, 1)$.

A seguir vamos considerar uma função $\phi : \mathbb{R}_+ \to \mathbb{R}_+$, satisfazendo as seguintes hipóteses:

(A1) $\exists\, \alpha \in (0, 1), \tau_\alpha > 0,\ \forall\, t \in (0, \tau_\alpha] : \phi(t) \leq \alpha t$,

(A2) $\exists\, \beta > 0, \tau_\beta \in (0, \infty], \forall\, t \in (0, \tau_\beta) \bigcap \mathbb{R} : \phi(t) \geq \beta t^2$,

(A3) $\forall\, k, f(x^{k+1}) \leq f(x^k) - \phi(t_k)\|g^k\|^2$, e $0 < t_k \leq \tau_\beta$ em (2.1),

(A4) $\exists\, \gamma > 1, \tau_\gamma > 0, \forall\, k : t_k \geq \tau_\gamma$, ou
$[\exists\, \widetilde{t}_k \in [t_k, \gamma t_k] : f(x^k - \widetilde{t}_k g^k) \geq f(x^k) - \phi(\widetilde{t}_k)\|g^k\|^2]$.

Note que a condição (2.2) é um caso particular dessas hipóteses. Para vermos isto basta fazermos :

$$\phi(t) = \alpha t, \quad \beta = \alpha, \quad \gamma = 2, \quad \tau_\gamma = \tau_\beta = \tau_\alpha = 1.$$

Além das hipóteses descritas anteriormente, vamos supor ainda a seguinte condição:

$$f(x^k) \geq f(\widetilde{x}), \quad (2.3)$$

para todo k e para algum $\widetilde{x}$ fixo.

É muito importante observarmos que a condição (2.3) é válida se o conjunto de pontos de mínimo da f que denotaremos por $\check{X}$ é não vazio, ou $\widetilde{x}$ é um ponto de acumulação de $\{x^k\}$. De fato,

i) consideraremos que $\check{X} \neq \emptyset$. Dessa forma temos que existe $\widetilde{x} \in \check{X}$ tal que $f(x^k) \geq f(\widetilde{x}), \forall\, k$.

ii) suponhamos que $\widetilde{x}$ é ponto de acumulação de $\{x^k\}$, então existe uma subsequência $\{x^{k_j}\} \subset \{x^k\}$ tal que, $\lim_{j\to\infty} x^{k_j} = \widetilde{x}$. Pela continuidade da f temos que $\lim_{j\to\infty} f(x^{k_j}) = f(\widetilde{x})$, portanto esta subsequência é limitada.

De (A3) temos que para todo k

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

logo $\{f(x^k)\}$ é monótona não crescente, e como possui uma subsequência limitada pelo Corolário 1.2.1, a sequência converge, e será para o mesmo limite da subsequência . Pelo Teorema 1.2.5

$$\lim_{k\to\infty} f(x^k) = f(\widetilde{x}) = \inf\{f(x^k), k = 1, 2, ...\}.$$

Logo $f(x^k) \geq f(\widetilde{x})$, $\forall k$. $\square$

O conjunto de pontos que satisfazem a desigualdade (2.3), será denotado por

$$\Gamma = \{\widetilde{x} \in \mathbb{R}^n; f(x^k) \geq f(\widetilde{x}), \forall k\}.$$

2.2 Lemas Auxiliares

Lema 2.2.1. Se $\Gamma \neq \emptyset$, então para todo $\widetilde{x} \in \Gamma$, temos

$$\sum_{k=0}^{\infty} t_k^2 \|g^k\|^2 \leq [f(x^0) - f(\widetilde{x})]/\beta. \tag{2.4}$$

Além disso, $x^k \to \overline{x}$ para algum $\overline{x} \in \Gamma$.

Demonstração: De (A2) e (A3) temos :

$$\beta t_k^2 \leq \phi(t_k) \text{ e } \phi(t_k)\|g^k\|^2 \leq f(x^k) - f(x^{k+1}),$$

unindo estas duas desigualdades, obtemos

$$\beta t_k^2 \|g^k\|^2 \leq \phi(t_k)\|g^k\|^2 \leq f(x^k) - f(x^{k+1}),$$

como $\beta > 0$,

$$t_k^2 \|g^k\|^2 \leq [f(x^k) - f(x^{k+1})]/\beta, \forall k.$$

Fazendo k variar de 0 até n,

$$\begin{aligned} t_0^2\|g^0\| &\leq [f(x^0) - f(x^1)]/\beta \\ t_1^2\|g^1\| &\leq [f(x^1) - f(x^2)]/\beta \\ &\vdots \\ t_n^2\|g^n\| &\leq [f(x^n) - f(x^{n+1})]/\beta, \end{aligned}$$

e adicionando as inequações

$$\sum_{k=0}^{n} t_k^2 \|g^k\|^2 \leq \frac{[f(x^o) - f(x^{n+1})]}{\beta} \leq \frac{[f(x^0) - f(\widetilde{x})]}{\beta},$$

onde o extremo direito da desigualdade vem da hipótese que vale (2.3), em que $f(x^k) \geq f(\widetilde{x})$ para todo k, e como as reduzidas formam uma sequência limitada pelo Teorema 1.2.8 a série converge, e portanto existe o limite. Assim passando o limite na última desigualdade quando $k \to \infty$, temos :

$$\sum_{k=0}^{\infty} t_k^2 \|g^k\|^2 \leq \frac{[f(x^0) - f(\widetilde{x})]}{\beta},$$

logo vale a primeira parte do Lema 2.2.1.

Pela quaseconvexidade da f, levando em conta que vale (2.3) e por f ser continuamente diferenciável no $\mathbb{R}^n$, temos pelo Teorema 1.2.3 que

$$\langle \nabla f(x^k), \widetilde{x} - x^k \rangle \leq 0, \quad k = 0, 1, 2, ...,$$

lembrando que $\nabla f(x^k) = g^k$ obtemos

$$\langle g^k, \widetilde{x} - x^k \rangle \leq 0, \quad k = 0, 1, 2, ...$$

Sendo $x^{k+1} = x^k - t_k g^k \Rightarrow x^k - x^{k+1} = t_k g^k$, deduzimos que :

$$\begin{aligned} 0 \leq \|\widetilde{x} - x^{k+1}\|^2 &= \|\widetilde{x} - x^k + x^k - x^{k+1}\|^2 \\ &= \|\widetilde{x} - x^k\|^2 + 2\langle \widetilde{x} - x^k, x^k - x^{k+1} \rangle + \|x^k - x^{k+1}\|^2 \\ &= \|\widetilde{x} - x^k\|^2 + 2t_k\langle \widetilde{x} - x^k, g^k \rangle + \|t_k g^k\|^2 \\ &\leq \|\widetilde{x} - x^k\|^2 + t_k^2\|g^k\|^2. \end{aligned}$$

Observamos que $\{x_k\}$ é quase Fejér convergente sobre Γ, pela Definição 1.1.5 basta tomarmos $\xi_k = t_k^2\|g^k\|^2 \geq 0$ e por (2.4)

$$\sum_{k=0}^{\infty} \xi_k = \sum_{k=0}^{\infty} t_k^2\|g^k\|^2 < \infty,$$

logo

$$\|\widetilde{x} - x^{k+1}\|^2 \leq \|\widetilde{x} - x^k\|^2 + \xi_k, \forall\, k.$$

Pelo Teorema 1.2.6 temos que $\{x^k\}$ é limitada (está contida dentro de uma bola de centro $\widetilde{x}$, e raio $\|x^0 - \widetilde{x}\|^2 + \sum_{k=0}^{\infty} \xi_k$). Segue que $\{x^k\}$ possui um ponto de acumulação que chamaremos de $\overline{x}$. Como o conjunto dos pontos de acumulação de $\{x^k\}$ é representado por Γ, então $\overline{x} \in \Gamma$. Pela segunda parte do Teorema 1.2.6, obtemos $\lim_{k \to \infty} x^k = \overline{x}$. $\square$

Lema 2.2.2. Se $\overline{x}$ é um ponto de acumulação de $\{x^k\}$, então temos que $\overline{x} \in X = \{x : \nabla f(x) = 0\}$.

Demonstração: Seja $\overline{x}$ um ponto de acumulação de x^k, pela última parte do Lema 2.2.1 temos que $x^k \to \overline{x}$ quando $(k \to \infty)$. Suponhamos por absurdo que $\overline{g} = \nabla f(\overline{x}) \neq 0$. Como f é continuamente diferenciável, quando $k \to \infty$,

$$g^k = \nabla f(x^k) \to \nabla f(\overline{x}) = \overline{g} \neq 0 \;\text{ e }\; f(x^k) \to f(\overline{x}).$$

De (A2) e (A3) deduzimos que $t_k \to 0$ quando $k \to \infty$ pois :

$$0 \leq \beta t_k^2 \|g^k\|^2 \leq \phi(t_k)\|g^k\|^2 \leq f(x^k) - f(^{k+1}) \Rightarrow$$
$$0 \leq \beta t_k^2 \|g^k\|^2 \leq f(x^k) - f(x^{k+1}).$$

Como $\beta > 0$ temos ainda que :

$$0 \leq t_k^2 \|g^k\|^2 \leq \frac{f(x^k) - f(x^{k+1})}{\beta}.$$

Passando o limite quando $k \to \infty$:

$$0 \leq \lim_{k \to \infty} t_k^2 \|g^k\|^2 \leq 0.$$

Como $g^k \to \overline{g} \neq 0$, obtemos

$$0 \leq \lim_{k \to \infty} t_k^2 \leq 0.$$

Logo $t_k \to 0$ quando $k \to \infty$.

Assim para todo k suficientemente grande, de (A4) e (A1) temos :

$$f(x^k - \widetilde{t}_k g^k) - f(x^k) \geq -\phi(\widetilde{t}_k)\|g^k\|^2 \geq -\alpha \widetilde{t}_k \|g^k\|^2, \tag{2.5}$$

com $0 \leq t_k \leq \widetilde{t}_k \leq \gamma t_k \to 0$, quando $k \to \infty$.

Pelo Teorema 1.2.1 (Teorema do Valor Médio), o lado esquerdo da desigualdade (2.5) para $0 < \theta < 1$ será :

$$f(x^k - \widetilde{t}_k g^k) - f(x^k) = \langle \nabla f(x^k - \theta \widetilde{t}_k g^k), -\widetilde{t}_k g^k \rangle = -\widetilde{t}_k \langle g^k, \nabla f(x^k - \theta \widetilde{t}_k g^k) \rangle.$$

Então $-\widetilde{t}_k \langle g^k, \nabla f(x^k - \theta \widetilde{t}_k g^k) \rangle \geq -\alpha \widetilde{t}_k \|g^k\|^2$.

Dividindo ambos os lados da última desigualdade por $\widetilde{t}_k > 0$ e passando o limite com $k \to \infty$ teremos

$$-\langle \nabla f(\overline{x}), \nabla f(\overline{x}) \rangle = -\|\overline{g}\|^2 \geq -\alpha \|\overline{g}\|^2,$$

contradição, visto que $\alpha < 1$ e $\overline{g} \neq 0$, logo $\nabla f(\overline{x}) = 0$. $\square$

2.3 Teorema Principal

Teorema 2.3.1. Convergência Global. Temos que $x^k \to \overline{x} \in X$, ou $\check{X} = \emptyset, \|x^k\| \to \infty$, e $f(x^k) \to \inf f$.

Demonstração: Suponha que $\check{X} \neq \emptyset$, então existe pelo menos um mínimo global x^*, tal que $f(x^k) \geq f(x^*)$, para todo k. Assim a condição (2.3) é satisfeita. Aplicando o Lema 2.2.1, $x^k \to \overline{x}$, e pelo Lema 2.2.2, $\overline{x} \in X$.

Se $\|x^k\| \nrightarrow \infty$, teremos que $\{x^k\}$ é limitada tendo assim um ponto de acumulação, então novamente a condição (2.3) será satisfeita e pelos Lemas 2.2.1 e 2.2.2 $\{x^k\} \to \overline{x} \in X$.

Se $\lim\limits_{k\to\infty} f(x^k) > \inf f = f(\widetilde{x})$ então

$f(\widetilde{x}) < f(x^k), \forall k$, então vale (2.3), e os resultados anteriores. □

Capítulo 3

Discussões

Neste capítulo trataremos de discussões em cima de outras linhas de pesquisa, modificaremos algumas hipóteses deste trabalho e faremos uma comparação com resultados obtidos anteriormente a este trabalho .

(1) Analisemos primeiro para o caso em que f é convexa, e seja $\alpha \in (1/2, 1)$. Então a prova do Lema 2.2.1 com essas novas hipóteses torna-se simples, desde que (2.3) seja válida e já que,

$$f(\widetilde{x}) \geq f(x^k) + \langle \nabla f(x^k), \widetilde{x} - x^k \rangle \Rightarrow$$
$$\langle g^k, x^k - \widetilde{x} \rangle \geq f(x^k) - f(\widetilde{x}) \geq f(x^k) - f(x^{k+1}) \geq \alpha t_k \|g^k\|^2.$$

Se multiplicarmos ambos os lados da desigualdade por $-2t_k$:

$$-2t_k \langle g^k, x^k - \widetilde{x} \rangle \leq -2t_k^2 \alpha \|g^k\|^2.$$

Sabemos que,

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

$$= -2\langle x^k - \widetilde{x}, t_k g^k\rangle + t_k^2\|g^k\|^2 \leq -2\alpha t_k^2\|g^k\|^2 + t_k^2\|g^k\|^2$$
$$= t_k^2\|g^k\|^2.(-2\alpha + 1) = -(2\alpha - 1)t_k^2\|g^k\|^2 \leq 0.$$

Pois $\alpha \in (1/2, 1)$. Então,

$$\|\widetilde{x} - x^{k+1}\|^2 - \|\widetilde{x} - x^k\|^2 \leq 0 \Rightarrow$$
$$\|\widetilde{x} - x^{k+1}\|^2 \leq \|\widetilde{x} - x^k\|^2, \forall\, k \geq 0,$$

assim, $\{x^k\}$ é Fejér convergente, e aplicando esta definição sucessivas vezes teremos :

$\|x^k - \widetilde{x}\|^2 \leq \|x^0 - \widetilde{x}\|^2$, para algum $\widetilde{x}$, ou seja a sequência $\{x^k\}$ está contida em alguma bola de centro $\widetilde{x}$ e raio $\|x^0 - \widetilde{x}\|$, logo $\{x^k\}$ é limitada. Seja $\{x^{k_j}\} \subset \{x^k\}$, tal que $\lim\limits_{j\to\infty} x^{k_j} = \overline{x}$. Assim, devido a sequência $\{x^k\}$ ser Fejér convergente teremos que a sequência $\{\|x^k - \overline{x}\|\}$ é decrescente e não negativa, e tem uma subsequência $\{\|x^{k_j} - \overline{x}\|\} \to 0$. Concluimos então que a sequência $\{\|x^k - \overline{x}\|\} \to 0$, ou seja, $\lim\limits_{k\to\infty} x^k = \overline{x}$. □

(2) É fácil verificar o Teorema 2.3.1 para alguma linha de pesquisa se vale o Lema 2.2.2, e também se $\alpha \in (0, 1)$ e para todo k, tivermos

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

e $t_k \in (0, t_{max}]$, para algum $t_{max} > 0$ fixado. Esse tamanho de passo é encontrado em muitos procedimentos. Note que as buscas lineares exatas não são admissíveis, mas uma pode ser usada, que é a do método da bisseção, que consiste em procurar o ponto de mínimo em um intervalo, que será dividido ao meio a cada nova iteração. Quando seu comprimento se tornar menor que uma tolerância especificada, cada ponto do intervalo poderá ser levado a uma aproximação do mínimo.

Tomaremos então o tamanho de passo que realiza o mínimo de $f(x^{k+1})$, ou seja, o tamanho de passo calculado da seguinte maneira :

$t_k \approx argmin\{f(x^k - tg^k) : f(x^k - tg^k) \leq f(x^k) - \alpha t \|g^k\|^2, 0 < t < t_{max}\}$.

onde denotaremos neste caso citado como $argmin$, o conjunto de pontos que minimizam a função f no ponto $(x^k - tg^k)$ e ainda satisfazem a desigualdade de Armijo.

(3) Se relaxarmos a regra de Armijo como no parágrafo anterior e em particular usarmos

$$t_k \approx \check{t}_k = argmin\{f(x^k - tg^k) + \alpha t^2 \|g^k\|^2 : t > 0\}, \tag{3.1}$$

onde t_k tomado dessa forma é o mesmo do Teorema 1.2.7 e, se vale que f é pseudo-convexa, então $X = \check{X}$.

Sendo $\check{X}$ o conjunto de minimizadores da f e $X = \{x : \nabla f(x) = 0\}$. Provaremos a afirmação acima.

i) $X \subset \check{X}$. Como f é pseudo-convexa sobre R^n então em particular é pseudo-convexa sobre X, e para todo $\widetilde{x} \in X$, pela Definição 1.1.3 e pelo Teorema 1.2.4 :

$f(\widetilde{x}) = min_{x \in X} f(x) \Rightarrow \widetilde{x} \in \check{X}$, logo $X \subset \check{X}$.

ii) $\check{X} \subset X$ pela condição necessária de primeira ordem para problemas irrestritos.

Assim, se $\check{X} \neq \emptyset$ e $t_k = \check{t_k}, \forall\, k$, então $x^k \to \overline{x} \in X = \check{X}$.

De fato, pois recorremos ao resultado da referência [3], que neste trabalho equivale ao Teorema 1.2.7 onde é garantido que se a função é pseudo-convexa e atinge seu mínimo sobre o $\mathbb{R}^n$ então a sequência $\{x^k\}$ gerada por (2.1) e por (3.1) converge para o mínimo da f (como vimos, $\overline{x}$). $\square$

(4) Podemos examinar as hipóteses (A1)-(A4) para os algoritmos de [2], onde o teorema principal dessa referência prova a convergência do método usando funções convexas continuamente diferenciáveis, tal demonstração utiliza dois algoritmos. Para o primeiro algoritmo é assumido que o gradiente da função satisfaz a condição de lipschitz e é usado o método do gradiente, descrito no item 1.4.1, com busca linear inexata. Desde que se conheça a constante de lipschitz L, o tamanho de passo é calculado da seguinte forma : considera-se dois números positivos δ_1 e δ_2 tal que

$$\frac{L}{2}\delta_1 + \delta_2 < 1 \Rightarrow 0 < \delta_1 < (1-\delta_2).\frac{2}{L}$$

O tamanho de passo t_k é tomado no intervalo $[\delta_1, \frac{2}{L}(1-\delta_2)]$.

Para verificarmos que o algoritmo se encaixa nas hipóteses (A1)-(A4) basta tomarmos $\phi(t) = \beta t^2$, onde na notação do trabalho de Burachik *et al.*, temos que

$$\beta = \frac{L\delta_2}{2(1-\delta_2)}, \quad \tau_\gamma = \delta_1(\text{satisfaz A4})$$

$$\tau_\beta = \frac{2}{L}(1-\delta_2) \text{ (satisfaz A2 e A3)},$$

$$\tau_\alpha = \frac{\alpha}{\beta} \text{ e } \alpha \in (0,1)(\text{satisfaz A1}).$$

Podemos observar que β satisfaz a hipótese (A2), onde $\beta > 0$ pois temos $0 < (1-\delta_2)\frac{2}{L} \Rightarrow 0 < [L.\delta_2/2(1-\delta_2)] = \beta$.

Para o segundo algoritmo da referência [2], é considerado uma função convexa e continuamente diferenciável, portanto, para nós basta fazermos esta função igual a ϕ deste trabalho. Logo assumiremos as hipóteses (A1)-(A4).

O Teorema 2.3.1 é um resultado mais forte que o teorema principal da referência [2], pois enfraquecemos as hipóteses do mesmo, visto que não exigimos que a função seja convexa e nem muito menos que ela possua seu gradiente satisfazendo a condição de Lipschitz e sua demonstração como já vimos é simples.

(5) Notemos que a quaseconvexidade da f é necessária para o Lema 2.2.1, e consequentemente o Teorema 2.3.1. Por exemplo, seja $n = 2$, e tome $f(x,y) = e^x - y^2, \ \ (x,y)^0 = (0,0)^T$.

Observemos que esta função não é quaseconvexa. De fato, pois temos que pela Definição 1.1.1, a função f será quaseconvexa sobre o $\mathbb{R}^2$, se, e somente se, dados $x, y \in \mathbb{R}^2$ e $f(x) \leq f(y)$ e $\lambda \in [0,1]$ implicar em $f[(1-\lambda)y + \lambda x] \leq f(y)$. Tomemos então $x = (k_1, k_2)$ e $y = (k_1, -k_2)$, com $k_2 \neq 0$.

$f(x) = f(k_1, k_2) = e^{k_1} - k_2^2 \leq f(y) = e^{k_1} - k_2^2$, por outro lado temos,

$$f[(1-\lambda)y + \lambda x] = f[k_1, k_2(2\lambda - 1)] = e^{k_1} - k_2^2(2\lambda - 1).$$

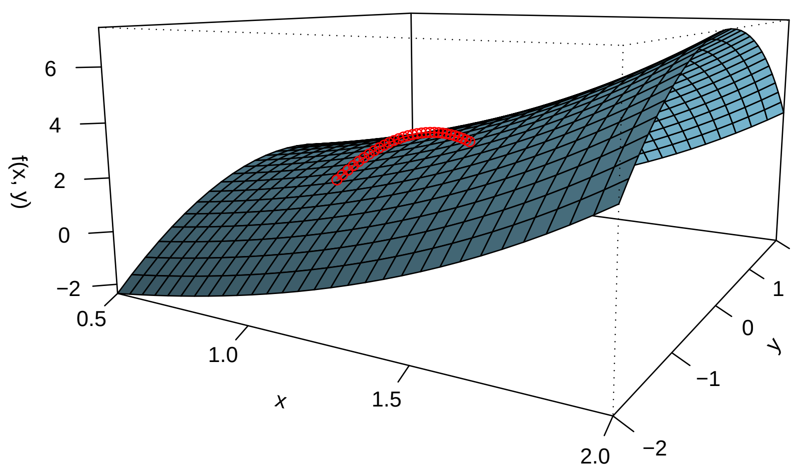
Se tormamos λ no interior do intervalo [0,1], por exemplo $\lambda = 1/2$, obtemos

$$f[(1-\lambda)y + \lambda x] = e^{k_1} > f(y) = e^{k_1} - k_2^2.$$

Concluímos assim que f não é quaseconvexa.

Analisemos na figura 3.1 o gráfico desta função para pontos $P_1 = (1,1)$ e $P_2 = (1,-1)$, sabemos que $f(1,1) = e - 1 \leq f(1,-1) = e - 1$. Analisando a função f no segmento $(1-\lambda)P_2 + \lambda P_1$, com $\lambda \in [0,1]$ (conforme a definição de quaseconvexidade), observamos que no gráfico da f, teremos que $f[(1-\lambda)P_2 + \lambda P_1] > f(P_2)$ para todo $\lambda \in (0,1)$, o que confirma que realmente a função em questão não é quaseconvexa. Os pontos do segmento que satisfazem a definição de quaseconvexidade da f são apenas quando particularizamos $\lambda = 0$ e $\lambda = 1$, que equivalem aos extremos que estão contidos na superfície.

Figura 3.1:



Em uma breve análise verificamos que a função acima, nem ao menos possui pontos estacionários, pois temos que

$$\nabla f(x) = (e^{x_1}, -2x_2).$$

Observamos então que $\nexists\, x_1$ tal que $e^{x_1} = 0$.

Analisaremos agora a segunda parte do teorema, pois concluimos que $\check{X} = \emptyset$. O algoritmo gerará então uma sequência $x^k = (x_1^k, 0)$ onde podemos observar que primeira coordenada $x_1^k \to -\infty$ quando $k \to \infty$

e temos que $f(x^k) = f(x_1^k, 0) = e^{x_1^k}$ quando $k \to \infty$ temos,

$$\lim_{k\to\infty} f(x_1^k, 0) = \lim_{k\to\infty} e^{x_1^k} - 0 = 0.$$

Logo $f(x^k) \to 0$.

Observemos que o $\inf f = -\infty$. De fato, basta tomarmos a seguinte sequência, $x^{k'} = (x_1, x_2) = (0, k)$, então $f(x^{k'}) = e^0 - (k)^2 = 1 - k^2$, ou seja, f é ilimitada inferiormente, o que implicará que a segunda parte do teorema também não é satisfeita, pois $f(x^k) \to 0$, enquanto $\inf f = -\infty$.

Conclusão

Mostramos de maneira simples e com hipóteses bastante razoáveis, a convergência do método de descida do gradiente com busca linear inexata pela regra de Armijo para funções quaseconvexas. Constatamos que é de suma importância a quaseconvexidade da função, pois a retirada da mesma impossibilita a demonstração do teorema principal, apresentamos um contra-exemplo de uma função que não era quaseconvexa, e verificamos que a mesma não permitiu que o teorema principal fosse satisfeito.

Notamos também que se alterarmos a hipótese da quaseconvexidade da função e sugerirmos que a mesma seja convexa restringindo o valor de α das hipóteses preliminares dentro do intervalo de $(1/2, 1)$, tornamos a prova do Lema 2.2.1 bem simples. Notamos em um dos comentários do capítulo 3, que podemos "relaxar"o tamanho de passo t_k, de maneira semelhante ao do Teorema 1.2.7 da parte de generalidades, um tamanho de passo em que utiliza-se uma função de regularização quadrática.

Comparamos este trabalho com o da referência [2] e mostramos que suas hipóteses se encaixam como um caso particular deste trabalho.

Gostaríamos de concluir descrevendo a relevância deste trabalho, que serviu de base fundamental para outros artigos como "Convergence of de Gradient Projection Method for Generalized Convex Minimization", referência [11], feito por Wang e Xiu publicado em 2000, onde é utilizado fortemente o teorema principal demonstrado aqui, para mostrar entre outros resultados a convergência método do gradiente projetado para funções pseudoconvexas e quaseconvexas . O artigo escrito por Quiroz *et al.* [8], "Steepest Descent Method with a Generalized Armijo Search for Quasiconvex Functions on Riemanian Manifolds", publicado em 2008, mostra este trabalho com todas a hipóteses, teoremas e lemas sobre a variedade riemaniana.

Referências Bibliográficas

[1] Armijo, L., *Minimization of Functions Having Continuous Partial Derivatives*, Pacific Journal of Mathematics, vol. 16, pp. 1-3, 1996.

[2] Burachik, R., Grana Drummond, L. M., Iusem,A. N., and Svaiter, B. F., *Full Convergence of the Steespest Descent Method with Inexact Line Searches*, Optimization, vol. 32, pp. 137-146, 1995.

[3] Iusem, A. N., and Svaiter, B.F., *A Proximal Regularization of the Steepest Descent Method*, Rairo Rechércher Opérationelle, vol. 29, pp.123-130, 1995.

[4] Kiwiel, K. C., and Murty, K., *Convergence of the Steepest Descent Method for Minimizing Quasiconvex Functions*, Journal of Optimization Theory and Applications, vol. 89, pp. 221-226, 1996.

[5] Lima, E. L., *Curso de Análise*, Instituto de Matemática Pura e Aplicada, vol. 1, terceira edição, Rio de Janeiro, 1995.

[6] Mangasarian, O. L., *Nonlinear Programming*, Mc-Graw, New York, 1969; Reimpresso por SIAM, Filadélfia, Pensylvania, 1994.

[7] Murty, K. G., *Linear Complementarity, Linear and Nonlinear Programming*, Heldermann Verlag,Berlin, Germany, 1988.

[8] Quiroz, E. A. P., Quispe,E. M., and Oliveira, P. R., *Steepest Descent Method with a Generalized Armijo Search for Quasiconvex Functions on Riemannian Manifolds*, Journal of Mathematical Analysis and Applications, vol.341, pp. 467-477, Rio de Janeiro, 2007.

[9] Solodov, M., e Izmailov,A., *Otimização : Métodos Computacionais*, Instituto de Matemática Pura e Aplicada , vol. 2, Rio de Janeiro, 2007.

[10] Solodov, M., e Izmailov, A., *Otimização : Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade*, Instituto de Matemática Pura e Aplicada , vol. 1,Rio de Janeiro, 2007.

[11] Wang, C., e Xiu, N., *Convergence of the Gradient Projection Method for generalized Convex Minimization*, Computational Optimization and Applications, vol. 16, pp. 111-120, Holanda, 2000.