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

CONVERGÊNCIA COMPLETA DO MÉTODO DO GRADIENTE COM BUSCA LINEAR EXATA E INEXATA

Jeanne Moreira de Sousa

MANAUS

2008

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

Jeanne Moreira de Sousa

CONVERGÊNCIA COMPLETA DO MÉTODO DO GRADIENTE COM BUSCA LINEAR EXATA E INEXATA

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, área de concentração em Otimização.

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

MANAUS
2008

Aos amores da minha vida: Iran, Raquel e Daniel.

Jeanne Moreira de Sousa

CONVERGÊNCIA COMPLETA DO MÉTODO DO GRADIENTE COM BUSCA LINEAR EXATA E INEXATA

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, área de concentração em Otimização.

Manaus, 29 de dezembro de 2008.

BANCA EXAMINADORA

....

Prof^o Dr. Roberto Cristóvão Mesquita Silva, Presidente
Universidade Federal do Amazonas

....

Prof^a Dra. Flávia Morgana de Oliveira Jacinto, Membro
Universidade Federal do Amazonas

....

Prof^a Dra. Sissy da Silva Souza , Membro
Universidade Federal do Piauí.

AGRADECIMENTOS

Primeiramente agradeço à Deus por ter enviado o Seu único Filho Jesus Cristo para dar-me a salvação, ensinar-me a combater o bom combate, completar a carreira e guardar a fé.

Ao meu esposo Iran, pelo amor verdadeiro, incontestável apoio e por terme presenteado com dois lindos filhos Raquel e Daniel, que renovavam as minhas forças, enchendo-me de beijos, a cada dia que chegava em casa.

Aos meus amados pais, Davi e Nazaré, pelo amor, estrutura familiar e por me ensinarem que para sonhar é preciso estar bem acordado.

Queridos irmãos, pela amizade, confiança no meu trabalho e o compartilhar de momentos de alegria e tristeza.

A querida tia Iranice, por cuidar com todo amor, carinho e dedicação da minha valiosa herança: Raquel e Daniel.

Aos amigos do mestrado, em especial Silvia Dias e Silvia Viviane, pelo companherismo e amizade sincera.

Aos meus professores pelo conhecimento compartilhado.

Ao meu orientador, professor Roberto Cristóvão pelo incentivo, paciência, dedicação e auxilio durante o desenvolvimento desta dissertação.

As professoras Flavia Morgana e Sissy Sousa, pelo tempo disponibilizado para a leitura deste trabalho e todas as sugestões de melhorias à ele propostas.

Aos incentivadores financeiros deste projeto: Secretaria Municipal de Educação - SEMED e Fundação de Amparo à Pesquisa do Amazonas - FAPEAM.

RESUMO

CONVERGÊNCIA COMPLETA DO MÉTODO DO GRADIENTE COM BUSCA LINEAR EXATA E INEXATA

Neste trabalho utilizamos o método do gradiente para minimizar, sem restrições, funções continuamente diferenciáveis pseudo-convexas e convexas. Um termo considerado importante é o cálculo do comprimento do passo. Na minimização de funções pseudo-convexas a busca linear é exata. Neste caso, apresentamos o primeiro algoritmo para o cálculo do comprimento do passo, onde é acrescentado um termo de regularização quadrático no sentido do método do ponto proximal. Posteriormente, na minimização de funções convexas, a busca linear é inexata. Para o cálculo do comprimento do passo apresentamos dois algoritmos: um necessita que o gradiente da função objetivo satisfaça uma condição de Lipschitz com constante $L > 0$ conhecida, e o outro é baseado no trabalho desenvolvido por Dennis-Schnabel (ver [4]). Os três processos baseiam-se na noção da quase-Fejér convergência. Embora os métodos de descida necessitem que a função objetivo a ser minimizada possua conjuntos de níveis limitados a fim de estabelecer que os pontos de acumulação sejam estacionários, nesta abordagem é garantida a convergência completa de toda sequência para um minimizador da função sem a hipótese de limitação do conjunto de nível.

ABSTRACT

FULL CONVERGENCE OF THE GRADIENT METHOD WITH EXACT AND INEXACT LINE SEARCHES

In this work we use the gradient method to minimize, without restrictions, convex and pseudoconvex continuously differentiable functions. An important theme considered is the path length determination. We have that, when minimizing pseudoconvex functions, the linear search is exact. In this case, we present the first algorithm to obtain the path length, where will be included a quadratic regularization term, in the proximal point method sense. When dealing with the minimization of convex functions case, we have that the linear search is not exact. To obtain the path length, two algorithms are presented: the former needs that the gradient of the objective function satisfies a Lipschitz condition with a known constant $L > 0$. The latter is based on the work of Dennis-Schnabel (see [4]). The three process are based on the quasi-Fejér convergence principle. Although these descent methods need that the objective functions to be minimized have bounded level sets, in order to establish that the limit points are stationary, this approach guarantees the complete convergence of every sequence to a minimizer of the function without the hypothesis of bounded level sets.

Sumário

Introdução

O método do gradiente (também conhecido como método de Cauchy ou método de máxima descida) utilizado para a minimização de funções reais definidas no $\mathbb{R}^n$, embora seja um procedimento simples e com convergência lenta, é considerado um ponto de partida para muitos outros métodos de otimização mais sofisticados.

Um ponto considerado importante no método do gradiente, além da escolha da direção de descida que é o anti-gradiente, é a escolha do comprimento do tamanho do passo. Este procedimento é chamado de busca linear e divide-se em exata e inexata.

A busca linear exata consiste em obter λ_k como um minimizador de f na semi-reta $\{x^k - \lambda \nabla f(x^k) \ ; \ \lambda > 0\}$ enquanto que, na busca linear inexata, os valores são executados, isto é, λ_k é dado de forma pré-determinada ou é obtido através de um procedimento finito. Geralmente, os métodos que utilizam a busca linear exata são bem mais eficientes mas, na implementação computacional de algoritmos, em sua grande maioria, é utilizada a busca linear inexata.

Os resultados clássicos de convergência exigem que o ponto inicial pertença a um conjunto de nível limitado da função f e, por consequência, que f tenha ao menos um conjunto de nível limitado. Dessa forma, o método prova que a sequência gerada converge para um ponto estacionário de f, estabelecendo ainda que apenas seus pontos de acumulação sejam estacionários.

No método do gradiente, a hipótese de limitação do conjunto de nível é primordial não só para a convergência mas também para a sua boa definição. Mesmo quando a função f é convexa (neste caso, os pontos estacionários são minimizadores de f) a hipótese de limitação do conjunto de nível é necessária.

O problema principal considerado neste trabalho é o de garantir a existência de um minimizador da função $f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciável (pseudo-convexa ou convexa), sem a hipótese de limitação dos conjuntos de nível da função f. Escreveremos tal problema como

$$\min f(x) \text{ sujeito a } x \in \mathbb{R}^n.$$

Para resolver este problema, utilizamos o método do gradiente com a busca linear exata para a minimização da função pseudo-convexa e, para o caso da função convexa, a busca linear será inexata.

A organização desta dissertação é descrita como segue:

No **Capítulo 1** listamos alguns resultados teóricos de Otimização e Análise necessários à abordagem proposta neste trabalho.

O **Capítulo 2** é baseado em [5]. Iusem e Svaiter mostraram que a inclusão de um termo de regularização quadrático no algoritmo do ponto proximal, utilizado na busca linear do método do gradiente, remove a exigência da hipótese de limitação do conjunto de nível. No caso da função objetivo pseudo-convexa, é provada a convergência completa da sequência gerada por esse algoritmo para um minimizador de f. Este é o primeiro resultado principal desta dissertação e seu enunciado segue abaixo:

Teorema *Seja f pseudo-convexa e continuamente diferenciável. Se f atinge seu mínimo em $\mathbb{R}^n$ então, a sequência $\{x^k\}$ gerada por:*

$$x^0 \in \mathbb{R}^n,$$

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

com

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

converge para um minimizador de f.

O **Capítulo 3**, por sua vez, é baseado em [2], onde Burachik *et al.* mostraram a convergência completa da sequência $\{x^k\}$ descrita pelo método do gradiente para um minimizador de f, onde f é uma função objetivo convexa. A busca linear é inexata e neste caso, o comprimento do tamanho de passo será calculado segundo dois algoritmos.

No primeiro, denominado Algoritmo A, λ_k é fixado e exige-se que o gradiente de f satisfaça a condição de Lipschitz com constante $L > 0$ conhecida.

O segundo, denominado Algoritmo B, é baseado no trabalho desenvolvido por Dennis-Schnabel (ver [4]). Neste caso, λ_k é encontrado usando somente uma das inequações (em vez das duas) exigidas em [4].

Para os dois algoritmos a convergência do método do gradiente é assegurada sem a exigência de limitação do conjunto de nível. Este é o segundo resultado principal desta dissertação e seu enunciado segue abaixo:

Teorema *Seja f convexa e continuamente diferenciável. Então, a sequência $\{x^k\}$ gerada por:*

$$x^0 \in \mathbb{R}^n,$$

$$x^{k+1} = x^k - \lambda_k \nabla f(x^k),$$

onde λ_k é calculado utilizando-se um dos Algoritmos A ou B, converge para um minimizador de f.

Nas **Considerações Finais**, além dos nossos comentários sobre os resultados obtidos, citamos os trabalhos desenvolvidos que tiveram como base os apresentados nesta dissertação.

É importante citar que os resultados de convergência estabelecidos neste trabalho estão baseados na noção de quase-Fejér convergência, estabelecida pela primeira vez em [18].

Capítulo 1

Preliminares

Neste capítulo citamos alguns resultados e definições da Análise Matemática e da Otimização, necessários ao desenvolvimento deste trabalho. Fixamos a notação utilizada nos capítulos posteriores e estabelecemos que a função $\|.\| : \mathbb{R}^n \to \mathbb{R}$ é a norma euclidiana.

1.1 Sequências no espaço euclidiano

Nesta seção, além das definições de sequências, subsequências e limites de sequências, apresentamos alguns resultados de convergência. As demonstrações não serão exibidas mas é possível encontrá-las em [8], [9] e [10].

Definição 1.1.1. Uma *sequência* em $\mathbb{R}^n$ é uma aplicação $x : \mathbb{N} \to \mathbb{R}^n$, definida no conjunto $\mathbb{N}$ dos números naturais.

Usamos a notação $\{x^k\}$ para indicar a sequência cujo k-ésimo termo é $x^k \in \mathbb{R}^n$.

Definição 1.1.2. Uma *subsequência* de $\{x^k\}$ é a restrição da sequência a um subconjunto infinito $N' = \{k_1 < k_2 < ... < k_i < ...\} \subset \mathbb{N}$.

A subsequência é indicada pela notação $\{x^{k_j}\}$, $j \in \mathbb{N}$.

Dizemos que a sequência $\{x^k\}$ é limitada quando o conjunto de seus termos é limitado em $\mathbb{R}^n$, ou seja, quando existe um número real $c > 0$ tal que $|x^k| \leq c$ para todo $k \in \mathbb{N}$.

Definição 1.1.3. O ponto $a \in \mathbb{R}^n$ é chamado *limite da sequência de pontos* $x^k \in \mathbb{R}^n$ quando $\lim_{k\to\infty} x^k = a$.

Quando existe o limite $a = \lim_{k\to\infty} x^k$ diz-se que a sequência $\{x^k\}$ é convergente. Caso contrário, diz-se que $\{x^k\}$ é divergente.

Em termos de bolas, tem-se $\lim_{k\to\infty} x^k = a$ se, e somente se, qualquer bola aberta de centro a contém todos os termos x^k, salvo, possivelmente para um número finito de índices k.

Para verificarmos que uma sequência não converge, basta verificarmos que ela não é limitada. O resultado abaixo mostra que

Teorema 1.1.1. ([9], teorema 3, pg. 110) *Toda sequência convergente é limitada.*

A recíproca é falsa, como exemplo podemos citar a sequência $(0, 1, 0, 1, ...)$. Esta sequência é limitada mas não é convergente porque possui duas subsequências que convergem para limites diferentes, a saber, $(0, 0, ...)$ e $(1, 1, ...)$.

Uma sequência $\{x^k\}$ chama-se crescente quando $x^1 < x^2 < x^3 < ...$ isto é, quando $x^k < x^{k+1}$ para todo $k \in \mathbb{N}$. Se vale $x^k \leq x^{k+1}$ para todo k, a sequência diz-se não-decrescente.

Analogamente, quando $x^1 > x^2 > x^3 > ...$ ou seja, quando $x^k > x^{k+1}$ para todo $k \in \mathbb{N}$, a sequência $\{x^k\}$ diz-se decrescente. Ela é chamada não-crescente quando $x^k \geq x^{k+1}$ para todo $k \in \mathbb{N}$.

Definição 1.1.4. As sequências crescente, não-decrescente, decrescente e não-crescente são chamadas *sequências monótonas.*

O teorema seguinte dá uma condição suficiente para que uma sequência convirja.

Teorema 1.1.2. ([9], teorema 4, pg. 111) *Toda sequência monótona limitada é convergente.*

Corolário 1.1.1. ([9], corolário, pg. 111) *Se uma sequência monótona $\{x^k\}$ possui uma subsequência convergente, então $\{x^k\}$ é convergente.*

Apresentamos abaixo a definição de valor de aderência.

Definição 1.1.5. Um número real a chama-se *valor de aderência* de uma sequência $\{x^k\}$ quando a é limite de uma subsequência de $\{x^k\}$.

Dada a sequência $\{x^k\}$ de números reais, limitada, temos que o conjunto de seus valores de aderência é não vazio e entre eles, existe um que é o menor de todos e um outro que é maior. Listamos abaixo alguns resultados que garantem a convergência da sequência $\{x^k\}$ se, e somente se, possui um único valor de aderência.

Teorema 1.1.3. ([9], corolário 1, pg. 123) *Toda sequência limitada de números reais possui uma subseqência convergente.*

Teorema 1.1.4. ([9], corolário 2, pg. 123) *Uma sequência limitada de números reais $\{x^k\}$ é convergente se, e somente se, possui um único valor de aderência.*

De forma geral verificamos que

Teorema 1.1.5. ([10], teorema 5, pg. 17)(**Teorema de Bolzano-Weierstrass**) *Toda sequência limitada em $\mathbb{R}^n$ possui uma subsequência convergente.*

Vejamos agora a definição de ponto de acumulação.

Definição 1.1.6. Um ponto $a \in \mathbb{R}^n$ chama-se *ponto de acumulação* do conjunto $X \subset \mathbb{R}^n$ quando toda bola aberta de centro a contém algum ponto de X diferente do ponto a.

De forma simbólica temos que

$$\forall\, \varepsilon > 0\; \exists\, x \in X;\; 0 < \|x - a\| < \varepsilon.$$

Teorema 1.1.6. ([10], teorema 9, pg. 20) *Dados $X \subset \mathbb{R}^n$ e $a \in \mathbb{R}^n$, as seguintes afirmações são equivalentes*

(a) *a é ponto de acumulação de X;*

(b) *Existe uma sequência de pontos $x^k \in X$, com $\lim_{k\to\infty} x^k = a$ e $x^k \neq a$ para todo $k \in \mathbb{N}$;*

(c) *Toda bola aberta de centro a contém uma infinidade de pontos de X.*

Uma série é uma soma $s = a_1 + a_2 + ... + a_n + ...$ com um número infinito de parcelas. Para que isso faça sentido, poremos $s = \lim_{n\to\infty}(a_1 + ... + a_n)$. Como todo limite, este pode existir ou não. Por isso há séries convergentes e séries divergentes.

Dada uma sequência $\{a_n\}$ de números reais, a partir dela formamos uma nova sequência $\{s_n\}$ onde

$$s_1 = a_1, s_2 = a_1 + a_2, ..., s_n = a_1 + a_2 + ... + a_n, etc.$$

Os números s_n chamam-se as *reduzidas* ou *somas parciais* da série $\sum a_n$. A parcela a_n é o *n-ésimo termo* ou *termo geral* da série.

Se existir o limite $s = \lim_{n\to\infty} s_n$, diremos que a série $\sum a_n$ é convergente. Caso contrário, diremos que $\sum a_n$ é divergente.

Às vezes é conveniente considerar séries do tipo $\sum_{n=0}^{\infty} a_n$, que começam com a_0 em vez de a_1.

Citamos abaixo, um resultado que nos ajuda a classificar uma série em convergente ou divergente.

Teorema 1.1.7. ([9], teorema 16, pg. 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 $n \in \mathbb{N}$.*

Outro fato importante que devemos citar é

Teorema 1.1.8. ([8], teorema 2, pg. 39): *O termo geral de uma série convergente tem limite zero.*

1.2 Aplicações contínuas

A noção de função contínua, nesta seção, é estudada em seus aspectos mais básicos e utilizada como instrumento para aplicação nos capítulos seguintes.

Definição 1.2.1. Seja $f : X \to \mathbb{R}^n$ definida no conjunto $X \subset \mathbb{R}^m$. Diz-se que f *é contínua no ponto* $a \in X$ quando,

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

Se $f : X \to \mathbb{R}^n$ é contínua em todos os pontos do conjunto X, diz-se simplesmente que f é uma *aplicação contínua.*

Seja $a \in X$ um ponto de acumulação de X. Dizemos que $f : X \to \mathbb{R}$ é contínua no ponto a se, e somente se, $\lim_{x \to a} f(x) = f(a)$.

De forma geral temos que

Teorema 1.2.1. ([10], teorema 13, pg. 26) *Uma aplicação $f : X \to \mathbb{R}^n$ definida no conjunto $X \subset \mathbb{R}^m$, é contínua no ponto $a \in X$ se, e somente se, para toda sequência de pontos $x^k \in X$ com $\lim_{k \to \infty} x^k = a$, tem-se $\lim_{k \to \infty} f(x^k) = f(a)$.*

Vejamos agora a definição de uma aplicação Lipschitz-contínua.

Definição 1.2.2. Dado $X \subset \mathbb{R}^m$, uma aplicação $f : X \to \mathbb{R}^n$ diz-se *Lipschitz-contínua* quando existe uma constante $L > 0$ (constante de Lipschitz) tal que

$$x, y \in X \Rightarrow \|f(x) - f(y)\| \leq L\|x - y\|.$$

Observação: No Capítulo 3 desta dissertação, dado o primeiro algoritmo para o cálculo do comprimento do passo, consideramos $f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciável e com gradiente Lipschitz-continuo. Isto significa que existe $L > 0$ tal que

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$$

para todo $x, y \in \mathbb{R}^n$.

1.3 Aplicações diferenciáveis

Para efeito de diferenciabilidade, onde se compara o acréscimo $f(a+h)-f(a)$ da função f, com o acréscimo h dado ao ponto a, o domínio mais adequado para a função é um subconjunto aberto $U \subset \mathbb{R}^n$. Pois neste caso, dado $a \in U$, tem-se ainda $a + h \in U$ para todo acréscimo suficientemente pequeno h.

Definição 1.3.1. Seja $U \subset \mathbb{R}^n$ um conjunto aberto. Dizemos que uma aplicação $f : U \to \mathbb{R}^n$ *é diferenciável no ponto $x \in U$* quando existe uma transformação linear $T : \mathbb{R}^m \to \mathbb{R}^n$ tal que

$$f(x + h) = f(x) + T.h + r(h), \text{ onde } \lim_{h \to 0} \frac{r(h)}{\|h\|} = 0.$$

A transformação linear $T : \mathbb{R}^m \to \mathbb{R}^n$ que dá a melhor aproximação para f perto de x é única e será chamada de *derivada de f no ponto x* e indicada por $f'(x)$ ou $Df(x)$.

Dado $U \subset \mathbb{R}^n$ aberto, dizemos que uma aplicação $f : U \to \mathbb{R}^n$ é diferenciável em U quando ela é diferenciável em todos os pontos de U. Podemos observar também que, se f é diferenciável no ponto x, então f é contínua neste ponto.

Seja f uma função real definida em um subconjunto do espaço $\mathbb{R}^n$. Definimos abaixo, a derivada parcial de f que tem propriedades análogas às da derivada de uma função definida num intervalo real.

Definição 1.3.2. Seja $f : U \to \mathbb{R}$ uma função real definida num subconjunto aberto $U \subset \mathbb{R}^n$. Dado o ponto $a \in U$, a *i-ésima derivada parcial de f* no ponto a (onde $1 \leq i \leq n$) é o limite

$$\frac{\partial f}{\partial x_i}(a) = \lim_{h \to 0} \frac{f(a + he_i) - f(a)}{h},$$

quando tal limite existe.

Como e_i é o i-ésimo vetor da base canônica do $\mathbb{R}^n$, temos que as derivadas parciais fornecem informações sobre a função ao longo de retas paralelas aos eixos. A definição abaixo estende a noção de derivada parcial a outras direções.

Definição 1.3.3. Sejam $f : U \to \mathbb{R}$ definida no aberto $U \subset \mathbb{R}^n$, $a \in U$ e $v \in \mathbb{R}^n$. A *derivada direcional de f no ponto a segundo o vetor v*, é, por definição, o limite

$$\frac{\partial f}{\partial v}(a) = \lim_{h \to 0} \frac{f(a + hv) - f(a)}{h},$$

quando tal limite existe.

Sejam $f : U \to \mathbb{R}$ diferenciável no aberto $U \subset \mathbb{R}^n$ e $c \in \mathbb{R}$. Dizemos que o ponto $x \in U$ *está no nível c*, ou *tem nível c*, relativamente à f, quando $f(x) = c$. Fixado c, o conjunto dos pontos de U que estão no nível c é a imagem inversa $f^{-1}(c)$, a qual é chamada a *superfície de nível c da função f*. Quando $n = 2$, $f^{-1}(c)$ chama-se a *curva de nível c de f*.

Apresentamos abaixo a definição de vetor gradiente.

Definição 1.3.4. Dada a função diferenciável $f : U \to \mathbb{R}$, 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

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

Destacamos abaixo as três propriedades mais importantes do gradiente de uma função diferenciável f.

Fixamos um ponto a e supomos que $\nabla f(a) \neq 0$. Então:

1) O gradiente aponta para uma direção segundo a qual a função f é crescente;

2) Dentre todas as direções ao longo das quais a função f cresce, a direção do gradiente é a de crescimento mais rápido;

3) O gradiente de f no ponto a é perpendicular à superfície de nível de f que passa por esse ponto.

A definição abaixo nos mostra que nem sempre uma função diferenciável precisa ser contínua.

Definição 1.3.5. Dizemos que $f : U \to \mathbb{R}^n$ é *continuamente diferenciável*, ou que f é de classe C^1, e escrevemos $f \in C^1$, quando f for diferenciável e, além disso, $T = f' : U \to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ for contínua.

Analogamente, se $f' : U \to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ tem derivada no ponto $x \in U$, dizemos que f é duas vezes diferenciável no ponto x e escrevemos $f'' : \mathbb{R}^m \to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ para indicar a derivada de f' no ponto x. Se além disso, a aplicação $f'' : U \to \mathcal{L}(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n))$ for contínua, dizemos que f é *duas vezes continuamente diferenciável* em U e escrevemos $f \in C^2$.

De maneira geral, dizemos que f é uma aplicação de classe C^k, quando $f^{(k)}$ for contínua. Por conveniência, C^0 indicará o conjunto das aplicações contínuas.

Definimos a importante classe C^∞ das aplicações infinitamente diferenciáveis como sendo a interseção de todas as classes C^k: $C^\infty = C^0 \cap C^1 \cap C^2 \cap ...$. Assim, $f \in C^\infty$ se, e somente se, possuir derivadas de todas as ordens para cada ponto de U e as derivadas forem contínuas.

1.4 Existência de soluções globais

Nesta seção estudamos alguns fatos clássicos de Otimização tais como a existência de soluções globais e condições de otimalidade. Para alguns resultados não exibimos as demosntração mas, é possível encontrá-las em [15] e [16], como também, as demais definições aqui apresentadas.

Sejam dados um conjunto $D \subset \mathbb{R}^n$ e uma função $f : D \to \mathbb{R}$. Consideramos o problema de achar um minimizador de f no conjunto D. Tal problema é escrito por

$$\min f(x) \text{ sujeito a } x \in D. \tag{1.1}$$

O conjunto D é chamado *conjunto viável*, os pontos de D são chamados *pontos viáveis* e f é chamada *função objetivo*. O problema acima é chamado *restrito*.

Definição 1.4.1. Dizemos que um ponto $\overline{x} \in D$ é

(a) *minimizador global* de (1.1), se

$$f(\overline{x}) \leq f(x) \ \forall\ x \in D. \tag{1.2}$$

(b) *minimizador local* de (1.1), se existe uma vizinhança U de $\overline{x}$ tal que

$$f(\overline{x}) \leq f(x) \ \forall\ x \in D \cap U. \tag{1.3}$$

Se para todo $x \neq \overline{x}$ a desigualdade (1.2) ou (1.3) é estrita, $\overline{x}$ é chamado minimizador *estrito* (global ou local, respectivamente).

Os critérios abaixo garantem a existência de soluções para problemas com restrições.

Teorema 1.4.1. ([15], teorema 1.2.1, pg. 7)(**Teorema de Weierstrass**) *Seja $D \subset \mathbb{R}^n$ um conjunto compacto não vazio e $f : D \to \mathbb{R}$ uma função contínua. Então, o problema (1.1) têm solução global.*

Demonstração: Como a imagem de um conjunto compacto por uma função contínua é compacta, o conjunto $\{v \in \mathbb{R} \mid v = f(x) \text{ para algum } x \in D\}$ é compacto. Em particular, este conjunto é limitado inferiormente, ou seja,

$$-\infty < \overline{v} = \inf_{x \in D} f(x).$$

Pela definição de ínfimo, para todo $k \in \mathbb{N}$ existe um $x^k \in D$ tal que

$$\overline{v} \leq f(x^k) \leq \overline{v} + \frac{1}{k}.$$

Passando ao limite quando $k \to \infty$, concluimos que

$$\lim_{k \to \infty} f(x^k) = \overline{v}. \tag{1.4}$$

Como $\{x^k\} \subset D$ e D é compacto, segue-se que $\{x^k\}$ é uma sequência limitada. Logo, ela possui uma subsequência $\{x^{k_j}\}$ que converge a um ponto de D isto é,

$$\lim_{j \to \infty} x^{k_j} = \overline{x} \in D.$$

Pela continuidade de f,

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

Usando (1.4), temos que $f(\overline{x}) = \overline{v}$, isto é, f assume o valor mínimo em D no ponto $\overline{x} \in D$. Em outras palavras, $\overline{x}$ é um minimizador global do problema (1.1). □

Definição 1.4.2. O *conjunto de nível* da função $f : D \to \mathbb{R}$ associado a $c \in \mathbb{R}$, é o conjunto dado por

$$L_{f,D}(c) = \{x \in D \mid f(x) \leq c\}.$$

Em particular, temos o seguinte resultado

Corolário 1.4.1. ([15], corolário 1.2.1, pg. 9) *Sejam $D \subset \mathbb{R}^n$ e $f : D \to \mathbb{R}$ contínua no conjunto D. Suponhamos que existe $c \in \mathbb{R}$ tal que o conjunto de nível $L_{f,D}(c)$ seja não vazio e compacto. Então, o problema (1.1) possui uma solução global.*

Demonstração: Pelo Teorema de Weierstrass, o problema

$$\min f(x) \text{ sujeito a } x \in L_{f,D}(c)$$

tem uma solução global, digamos $\overline{x}$. Para todo $x \in D \backslash L_{f,D}(c)$, temos que

$$f(x) > c \geq f(\overline{x}),$$

o que mostra que $\overline{x}$ é um minimizador global de f não só em $L_{f,D}(c)$, mas também em D. □

Definição 1.4.3. Dizemos que uma sequência $\{x^k\} \subset \mathbb{R}^n$ é *crítica* em relação ao conjunto D, se $\{x^k\} \subset D$ e $\|x^k\| \to \infty$ ou $\{x^k\} \to x \in \text{cl } D \backslash D$ $(k \to \infty)$, onde cl D é o fecho do conjunto D.

Definição 1.4.4. Dizemos que a função $f : D \to \mathbb{R}$ é *coerciva no conjunto* D, quando para toda sequência $\{x^k\}$ crítica em relação a D, tem-se

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

Exemplo 1. A função $f(x) = \frac{1}{x}$ não é coerciva em $(0, +\infty]$, mas é coerciva em $(0, t]$ para $t > 0$ fixo qualquer. A função $f(x) = x^2 + \frac{1}{x}$ é coerciva em $(0, +\infty]$.

Corolário 1.4.2. ([15], corolário 1.2.3, pg. 12) *Sejam* $D \subset \mathbb{R}^n$ *e* $f : D \to \mathbb{R}$ *uma função contínua coerciva em* D*. Então, o problema (1.1) possui uma solução global.*

Demonstração: Dado $\overline{x} \in D$, definamos o seguinte conjunto de nível

$$L_{f,D}(f(\overline{x})) = \{x \in D;\ f(x) \leq f(\overline{x})\}.$$

Notemos que $L_{f,D}(f(\overline{x})) \neq \emptyset$, pois $\overline{x} \in L_{f,D}(f(\overline{x}))$. Se mostrarmos que $L_{f,D}(f(\overline{x}))$ é compacto e, como f é contínua, o resultado seguirá do Corolário 1.4.1.

(1) Suponhamos, por absurdo, que $L_{f,D}(f(\overline{x}))$ seja ilimitado. Então vai existir uma sequência $\{x^k\} \subset L_{f,D}(f(\overline{x}))$ tal que

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

Como $\{x^k\} \subset L_{f,D}(f(\overline{x}))$ então $f(x^k) \leq f(\overline{x})$. Pela coercividade da f, temos que $\limsup\limits_{k\to\infty} \|x^k\| = +\infty$. Dessa forma

$$\limsup_{k\to\infty} f(x^k) \leq \limsup_{k\to\infty} f(\overline{x}),$$

implicando

$$+\infty \leq f(\overline{x}).$$

O que é um absurdo. Portanto, $L_{f,D}(f(\overline{x}))$ é limitado.

(2) Suponhamos, por absurdo, que $L_{f,D}(f(\overline{x}))$ não seja fechado. Então vai existir uma sequência $\{x^k\} \subset L_{f,D}(f(\overline{x}))$ tal que

$$\lim_{k \to \infty} x^k = \widehat{x} \in \mathrm{cl} L_{f,D}(f(\overline{x})) \backslash L_{f,D}(f(\overline{x})).$$

Como $\widehat{x} \notin L_{f,D}(f(\overline{x}))$, temos que $f(\widehat{x}) > f(\overline{x})$. Por outro lado, como $\{x^k\} \subset L_{f,D}(f(\overline{x}))$, então $f(x^k) \leq f(\overline{x})$. Pela continuidade de f, temos que

$$f(\lim_{k \to \infty} x^k) \leq f(\lim_{k \to \infty} \overline{x}),$$

e

$$f(\widehat{x}) \leq f(\overline{x}),$$

implicando $\widehat{x} \in L_{f,D}(f(\overline{x}))$. O que é absurdo. Portanto $L_{f,D}(f(\overline{x}))$ é fechado.

Dessa forma, concluímos que $L_{f,D}(f(\overline{x}))$ é compacto. Logo, segue-se o resultado. □

1.5 Condições de otimalidade

Apresentamos a seguir as condições de otimalidade para o problema de minimização irrestrita. Cabe notar que estes resultados são também verdadeiros para problemas de minimização restrita, desde que o ponto de interesse esteja no interior do conjunto viável D.

Seja $f : D \to \mathbb{R}$. Quando $D = \mathbb{R}^n$, dizemos que o problema (1.1) é irrestrito. Dessa forma, reescrevemos o problema (1.1) para

$$\min f(x) \text{ sujeito a } x \in \mathbb{R}^n. \tag{1.5}$$

Teorema 1.5.1. ([15], teorema 1.3.1, pg. 15)(**Condições de otimalidade no caso irrestrito**)

(a) *Suponhamos que a função $f : \mathbb{R}^n \to \mathbb{R}$ seja diferenciável no ponto $\overline{x} \in \mathbb{R}^n$. Suponhamos também que $\overline{x}$ seja um minimizador local do problema (1.5). Então*

$$\nabla f(\overline{x}) = 0. \tag{1.6}$$

Se f é duas vezes diferenciável em $\overline{x}$, então além de (1.6) tem-se que a matriz Hessiana de f no ponto $\overline{x}$ é semidefinida positiva, isto é,

$$\langle \nabla^2 f(\overline{x}) d, d \rangle \geq 0, \ \forall\, d \in \mathbb{R}^n. \tag{1.7}$$

(b) *Suponhamos que* $f : \mathbb{R}^n \to \mathbb{R}$ *seja duas vezes diferenciável no ponto* $\overline{x} \in \mathbb{R}^n$. *Se* $\overline{x}$ *satisfaz (1.6) e se a matriz Hessiana de* f *em* $\overline{x}$ *é definida positiva, isto é, se existe* $\gamma > 0$ *tal que*

$$\langle \nabla^2 f(\overline{x})d, d\rangle \geq \gamma \|d\|^2 \ \forall \ d \in \mathbb{R}^n, \tag{1.8}$$

então $\overline{x}$ *é um minimizador local estrito do problema (1.5).*

A condição (1.6) se chama a *condição necessária de primeira ordem* para o problema (1.5), e os pontos que a satisfazem se chamam *pontos estacionários* deste problema. A combinação de (1.6) e (1.7) se chama a *condição necessária de segunda ordem*, e a combinação de (1.6) e (1.8) é a *condição suficiente de segunda ordem* para o problema (1.5).

1.6 Convexidade

A noção de convexidade dentro da Otimização é de grande importância. Com as hipóteses de convexidade, as condições necessárias de otimalidade tornam-se suficientes. Em outras palavras, todo ponto estacionário é uma solução do problema. Em particular, qualquer minimizador local é global.

No Capítulo 3 desta dissertação, mostramos que o problema

$$\min f(x) \text{ sujeito a } x \in \mathbb{R}^n, \tag{1.9}$$

onde a função f é convexa, possui solução global.

No Capítulo 2, mostramos que mesmo enfraquecendo a hipótese de convexidade, a saber pseudo-convexidade, ainda é garantida a existência de um minimizador para a função f.

Definição 1.6.1. Um conjunto $D \subset \mathbb{R}^n$ é chamado *conjunto convexo* se para quaisquer $x \in D, y \in D$ e $\alpha \in [0,1]$, tem-se $\alpha x + (1-\alpha)y \in D$.

Definição 1.6.2. Se $D \subset \mathbb{R}^n$ é um conjunto convexo, diz-se que a *função* $f : D \to \mathbb{R}$ *é convexa* em D quando para quaisquer $x \in D, y \in D$ e $\alpha \in [0,1]$, tem se

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

A função f diz-se *estritamente convexa* quando a desigualdade acima é estrita para todos $x \neq y$ e $\alpha \in (0,1)$.

A função f diz-se *fortemente convexa com módulo* $\gamma > 0$, quando para quaisquer $x \in D$, $y \in D$ e $\alpha \in [0,1]$, tem-se

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

Podemos observar que uma função fortemente convexa é estritamente convexa e, uma função estritamente convexa é convexa.

Exemplo 2. A função $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$ é fortemente convexa. A função $f : \mathbb{R} \to \mathbb{R}$, $f(x) = e^x$ é estritamente (mas não fortemente) convexa. A função $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x$ é convexa (mas não estritamente).

Dizemos que

$$\min f(x) \text{ sujeito a } x \in D,$$

é um problema de minimização convexo quando $D \subset \mathbb{R}^n$ é um conjunto convexo e $f : D \to \mathbb{R}$ é uma função convexa no conjunto D.

A importância da convexidade pode ser vista no resultado seguinte.

Teorema 1.6.1. ([15], teorema 3.1.5, pg. 69)(**Teorema da minimização convexa**) *Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e $f : D \to \mathbb{R}$ uma função convexa em D. Então todo minimizador local é global. Além disso, o conjunto de minimizadores é convexo. Se f é estritamente convexa, não pode haver mais de um minimizador.*

Demonstração: Suponha que $\overline{x} \in D$ seja um minimizador local que não é global. Então existe $y \in D$ tal que $f(y) < f(\overline{x})$.

Definimos $x(\alpha) = \alpha y + (1-\alpha)\overline{x}$. Pela convexidade de D, $x(\alpha) \in D$ para todo $\alpha \in [0,1]$. Agora, pela convexidade de f, para todo $\alpha \in (0,1]$, tem-se

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

Tomando $\alpha > 0$ suficientemente pequeno, podemos garantir que o ponto $x(\alpha)$ é arbitrariamente próximo ao ponto $\overline{x}$, e ainda tem-se que $f(x(\alpha)) < f(\overline{x})$ e $x(\alpha) \in D$. Isto contradiz o fato de que $\overline{x}$ é minimizador local do problema (1.9). Portanto, qualquer solução local deve ser global.

Sejam $S \subset D$, o conjunto de minimizadores globais e, $\overline{v} \in \mathbb{R}$ o valor ótimo do problema ($f(x) = \overline{v}$ para qualquer $x \in S$). Para quaisquer $x \in S$, $\overline{x} \in S$ e $\alpha \in [0,1]$, pela convexidade de f obtemos

$$f(\alpha x + (1-\alpha)\overline{x}) \leq \alpha f(x) + (1-\alpha)f(\overline{x}) = \alpha\overline{v} + (1-\alpha)\overline{v} = \overline{v},$$

o que implica $f(\alpha x + (1-\alpha)\overline{x}) = \overline{v}$ e portanto, $\alpha x + (1-\alpha)\overline{x} \in S$. Dessa forma, S é convexo.

Suponhamos agora que f seja estritamente convexa e que existam $x \in S$ e $\overline{x} \in S$, $x \neq \overline{x}$. Seja $\alpha \in (0,1)$. Como x e $\overline{x}$ são minimizadores globais e $\alpha x + (1-\alpha)\overline{x} \in D$, pela convexidade de D, segue-se que

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

No entanto, pela convexidade estrita,

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

o que resulta em contradição. Concluimos que neste caso o minimizador é único. □

A seguir mostramos que os conjuntos de nível de uma função convexa são convexos.

Teorema 1.6.2. ([15], teorema 3.4.1, pg. 133)(**Convexidade de conjuntos de nível de funções convexas**) *Suponhamos que o conjunto $D \subset \mathbb{R}^n$ seja convexo e a função $f : D \to \mathbb{R}$ seja convexa em D. Então o conjunto de nível*

$$L_{f,D}(c) = \{x \in D \mid f(x) \leq c\}$$

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

Demonstração: Tomamos $c \in \mathbb{R}$ arbitrário. Se $L_{f,D}(c) = \emptyset$, a conclusão segue, pois o conjunto vazio é convexo trivialmente.

Sejam $x \in L_{f,D}(c)$ e $y \in L_{f,D}(c)$, isto implica, $x \in D$, $f(x) \leq c$ e $y \in D$, $f(y) \leq c$. Pela convexidade de D, $\alpha x + (1-\alpha)y \in D$. Pela convexidade de f em D,

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

o que mostra que $(\alpha x + (1-\alpha)y) \in L_{f,D}(c)$. □

A convexidade de todos os conjuntos de nível de uma função não é suficiente para garantir que ela é convexa.

Definição 1.6.3. Seja $D \subset \mathbb{R}^n$ um conjunto convexo. Dizemos que $f : D \to \mathbb{R}$ é *quase-convexa* em D quando os conjuntos de nível $L_{f,D}(c)$ são convexos para todo $c \in \mathbb{R}$.

Exemplo 3. A função $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$, possui conjuntos de nível convexos, mas não é convexa. Dessa forma, $f(x) = x^3$ é pseudo-convexa.

Uma função convexa é contínua em qualquer subconjunto aberto do seu domínio. Além do mais, ela é Lipschitz-contínua no interior do seu domínio.

Teorema 1.6.3. ([15], teorema 3.4.2, pg. 136)**(Continuidade de funções convexas)** *Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e aberto e $f : D \to \mathbb{R}$ uma função convexa em D. Então f é localmente Lipschitz-continua em D. Em particular, f é continua em D.*

Dada uma função f diferenciável, a convexidade admite algumas caracterizações úteis para determinar se f é convexa ou não.

Teorema 1.6.4. ([15], teorema 3.4.7, pg. 146)**(Caracterização de funções convexas)** *Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e aberto e $f : D \to \mathbb{R}$ uma função diferenciável em D. Então, as seguintes propriedades são equivalentes:*

(a) A função f é convexa em D.

(b) Para todo $x \in D$ e todo $y \in D$,

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

(c) Para todo $x \in D$ e todo $y \in D$,

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

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

(d) A matriz hessiana de f é semidefinida positiva em todo ponto de D:

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

Daremos agora, a definição de uma função pseudo-convexa.

Definição 1.6.4. ([12], definição 1, pg. 140) Seja f uma função real definida sobre algum conjunto aberto do $\mathbb{R}^n$ contendo o conjunto D. Dizemos que f

é *pseudo-convexa sobre* $\overline{x} \in D$ (com respeito a D) se é diferenciável em $\overline{x}$ e dado $x \in D$ com

$$\nabla f(\overline{x})(x - \overline{x}) \geq 0$$

implicar

$$f(x) \geq f(\overline{x}).$$

Dessa forma, f é dita *pseudo-convexa sobre* D, se é pseudo-convexa sobre cada $x \in D$.

Vale ressaltar que toda função convexa diferenciável é pseudo-convexa pois, para todo $x \in D$ e todo $y \in D$ com $\nabla f(x)(y - x) \geq 0$ vale que

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

implicando

$$f(y) \geq f(x).$$

Exemplo 4. A função $f : \mathbb{R} \to \mathbb{R}$, $f(x) = e^x$ é convexa e diferenciável para todo $x \in \mathbb{R}$. Logo, $f(x) = e^x$ é pseudo-convexa para todo $x \in \mathbb{R}$.

1.7 Método de descida

Os *métodos de descida* são uma estratégia natural para resolver problemas irrestritos tais como

$$\min f(x) \text{ sujeito a } x \in \mathbb{R}^n.$$

Seu desenvolvimento é: Dados $f : \mathbb{R}^n \to \mathbb{R}$ e $x^k \in \mathbb{R}^n$ uma aproximação da solução do problema. Encontramos o ponto $x^{k+1} \in \mathbb{R}^n$ tal que

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

Obviamente, isto pode ser feito de várias maneiras. Uma delas é a seguinte: toma-se uma direção $d^k \in \mathbb{R}^n$ tal que f é decrescente, preferencialmente à passos curtos, a partir do ponto x^k nessa direção. Calcula-se o comprimento de passo $\alpha_k > 0$ que fornece um valor de f menor do que no ponto x^k. Isto é,

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

Assim, obtemos o próximo iterado $x^{k+1} = x^k + \alpha_k d^k$. Repete-se o processo para o novo ponto x^{k+1}, etc.

Definição 1.7.1. Dizemos que $d \in \mathbb{R}^n$ é uma *direção de descida da função* $f : \mathbb{R}^n \to \mathbb{R}$ *no ponto* $x \in \mathbb{R}^n$, se existe $\varepsilon > 0$ tal que

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

Denotamos por $\mathcal{D}_f(x)$ o conjunto de todas as direções de descida da função f no ponto x.

No Lema abaixo apresentamos duas caracterizações para as direções de descida.

Lema 1.7.1. ([15], lema 1.4.1, pg. 22) (**Direções de descida**) *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função diferenciável no ponto* $x \in \mathbb{R}^n$. *Então:*

a) Para todo $d \in \mathcal{D}_f(x)$, *tem-se* $\langle \nabla f(x), d \rangle \leq 0$.

b) Se $d \in \mathbb{R}^n$ *satisfaz* $\langle \nabla f(x), d \rangle < 0$, *tem-se que* $d \in \mathcal{D}_f(x)$.

1.8 Busca Linear

Na estratégia descrita acima para os métodos de descida, além da importância na escolha da direção de descida temos a escolha do comprimento do passo, que pode ser calculado observando o comportamento da função f ao longo da semi-reta a partir de x^k na direção d^k, definido à priori sob determinadas condições, ou encontrado por meio de um procedimento numérico finito. Tal procura chama-se *busca linear*. Apresentamos agora duas das principais regras de busca linear.

Regra da minimização uni-dimensional

O procedimento é minimizar a função objetivo na semi-reta $x^k + \alpha d^k$, $\alpha \geq 0$. O comprimento do passo α_k vem dado pela condição

$$f(x^k + \alpha_k d^k) = \min_{\alpha \geq 0} f(x^k + \alpha d^k),$$

isto é, α_k é uma solução do problema

$$\min \varphi_k(\alpha), \; \alpha \in \mathbb{R}_+,$$

onde

$$\varphi_k : \mathbb{R}_+ \to \mathbb{R}, \; \varphi_k(\alpha) = f(x^k + \alpha d^k).$$

Sejam as funções f diferenciável no ponto x^{k+1} e, ψ_k diferenciável em α_k. Como α_k é mínimo da ψ_k então, pela condição necessária de primeira ordem (Teorema 1.5.1 (a)) segue que

$$0 = \varphi_k'(\alpha_k) = \langle \nabla f(x^k + \alpha_k d^k), d^k \rangle = \langle \nabla f(x^{k+1}), d^k \rangle.$$

Dessa forma,

$$\langle \nabla f(x^{k+1}), d^k \rangle = 0.$$

Portanto, para encontrarmos α_k basta solucionarmos a igualdade anterior.

Regra de Armijo

O procedimento é calcular um comprimento do passo que resulte no decréscimo *suficiente* da função f em relação ao valor $f(x^k)$. Supondo f diferenciável no ponto x^k, fixamos os parâmetros $\widehat{\alpha}$, σ, $\theta \in (0,1)$. Tomamos $\alpha := \widehat{\alpha}$.

P1. Verificamos se a desigualdade

$$f(x^k + \alpha d^k) \leq f(x^k) + \sigma\alpha \langle \nabla f(x^k), d^k \rangle \tag{1.10}$$

é satisfeita.

P2. Em caso afirmativo, aceitamos $\alpha_k = \alpha$ como o valor do comprimento do passo. Caso contrário, tomamos $\alpha := \theta\alpha$ e retornamos à P1.

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 (1.10).

1.9 Método gradiente

Seja $f : \mathbb{R}^n \to \mathbb{R}$ diferenciável. Quando tomamos d^k como o antigradiente da função f no ponto x^k isto é, $d^k = -\nabla f(x^k)$, o método de descida será chamado *método do gradiente*. Dessa forma, o processo iterativo se reduz a

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

O método do gradiente com a minimização unidimensional se chama o *método de máxima descida* e, uma propriedade importante deste método é a igualdade

$$\langle \nabla f(x^{k+1}), \nabla f(x^k) \rangle = 0.$$

O resultado abaixo garante a convergência global do método do gradiente quando utilizamos a minimização unidimensional ou a regra de Armijo para o cálculo do comprimento do passo e, a hipótese de continuidade para a derivada de f.

Teorema 1.9.1. ([16], teorema 3.1.2, pg. 78) (**Convergência global do método do gradiente**) *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no $\mathbb{R}^n$ com derivada continua. Suponhamos que o algoritmo geral do método do gradiente utiliza a minimização unidimensional ou a regra de Armijo. Então, cada ponto de acumulação de qualquer sequência $\{x^k\}$ gerada por tal algoritmo é um ponto estacionário do problema*

$$\min f(x) \text{ sujeito a } x \in \mathbb{R}^n.$$

Se a sequência é limitada, então vale

$$\{\nabla f(x^k)\} \to 0 \ (k \to \infty).$$

1.10 Método do ponto proximal

Sejam $f : \mathbb{R}^n \to \mathbb{R}$ convexa e diferenciável e o problema de minimização irrestrita

$$\min f(x) \text{ sujeito a } x \in \mathbb{R}^n. \tag{1.11}$$

Para resolver tal problema, é possível utilizarmos o *método do ponto proximal clássico* que consiste em descrever um algoritmo da seguinte forma:

$$x^0 \in \mathbb{R}^n \tag{1.12}$$

$$x^{k+1} = \arg\min_{x \in \mathbb{R}^n} [f(x) + \lambda_k \|x - x^k\|^2]. \tag{1.13}$$

Temos que $\|.\|$ é a norma Euclidiana e $\{\lambda_k\} \subset \mathbb{R}$ satisfaz $0 < \lambda_k < \overline{\lambda}$, sendo $\overline{\lambda}$ uma constante positiva.

O método do ponto proximal utilizado nesta dissertação está baseado em [6].

Mostramos abaixo que a sequência x^k gerada pelo algoritmo do ponto proximal é convergente ao conjunto de minimizadores de f, denominado S^*.

Teorema 1.10.1. ([6], teorema 2.1, pg. 4) *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa e continuamente diferenciável. Suponhamos que o conjunto solução do problema (1.11), denominado de S^*, seja não vazio. Então a sequência $\{x^k\}$ gerada por (1.12) e (1.13) converge a um ponto $\overline{x} \in S^*$.*

Demonstração: Dividiremos a prova em quatro passos:

Passo 1: Vamos mostrar que a sequência $\{x^k\}$ está bem definida.

Seja

$$f_k(x) = f(x) + \lambda_k \|x - x^k\|^2.$$

Como S^* é não vazio, tomemos $\beta \in S^*$ tal que $f(x) > \beta > -\infty$ para todo $x \in \mathbb{R}^n$. Assim, f é limitada inferiormente e

$$f_k(x) > \beta + \lambda_k \|x - x^k\|^2.$$

Como

$$\begin{aligned} \|x - x^k\|^2 &= \langle x - x^k, x - x^k \rangle \\ &= \|x\|^2 - 2\langle x, x^k \rangle + \|x^k\|^2 \\ &\geq \|x\|^2 - 2\|x\|.\|x^k\| + \|x^k\|^2 \\ &= (\|x\| - \|x^k\|)^2, \end{aligned}$$

então,

$$f_k(x) \geq \beta + \lambda_k (\|x\| - \|x^k\|)^2.$$

Assim,

$$\lim_{\|x\| \to \infty} f_k(x) = \infty.$$

Como f_k é continua e coerciva, temos que a mesma admite um ponto de mínimo, o qual é único, pois f_k é estritamente convexa.

Passo 2: Vamos mostrar que vale a desigualdade

$$\|x^{k+1} - \overline{x}\|^2 \leq \|x^k - \overline{x}\|^2 - \|x^{k+1} - x^k\|^2 \ \forall\, k \in \mathbb{N} \text{ e } \forall\, \overline{x} \in S^*.$$

Dada a igualdade abaixo,

$$\begin{aligned} \|x^k - \overline{x}\|^2 &= \|x^k - x^{k+1} + x^{k+1} - \overline{x}\|^2 \\ &= \|x^{k+1} - x^k\|^2 + \|x^{k+1} - \overline{x}\|^2 + 2\langle x^k - x^{k+1}, x^{k+1} - \overline{x} \rangle. \end{aligned}$$

Como x^{k+1} resolve (1.13) temos que

$$0 = \nabla f_k(x^{k+1}) = \nabla f(x^{k+1}) + 2\lambda_k(x^{k+1} - x^k).$$

Assim,

$$\nabla f(x^{k+1}) = -2\lambda_k(x^{k+1} - x^k). \tag{1.14}$$

Pela convexidade de f, o fato de que $\overline{x} \in S^*$ e por (1.14), segue-se que

$$\begin{aligned}
\|x^k - \overline{x}\|^2 - \|x^k - x^{k+1}\|^2 - \|x^{k+1} - \overline{x}\|^2 &= 2\langle x^k - x^{k+1}, x^{k+1} - \overline{x}\rangle \\
&= 2\langle \frac{\nabla f(x^{k+1})}{2\lambda_k}, x^{k+1} - \overline{x}\rangle \\
&= \frac{1}{\lambda_k}\langle \nabla f(x^{k+1}), x^{k+1} - \overline{x}\rangle \\
&\geq \frac{1}{\overline{\lambda}}[f(x^{k+1}) - f(\overline{x})] \geq 0.
\end{aligned}$$

Logo, a desigualdade é verificada.

Passo 3: Abaixo, provamos que $\lim\limits_{k\to\infty}(x^{k+1} - x^k) = 0$.

Do **Passo 2** temos que

$$0 \leq \|x^{k+1} - x^k\|^2 \leq \|x^k - \overline{x}\|^2 - \|x^{k+1} - \overline{x}\|^2. \tag{1.15}$$

Dessa forma,

$$\|x^{k+1} - \overline{x}\|^2 \leq \|x^k - \overline{x}\|^2, \text{ para todo } k \in \mathbb{N}, \tag{1.16}$$

o que implica

$$\|x^k - \overline{x}\|^2 \leq ... \leq \|x^0 - \overline{x}\|^2. \tag{1.17}$$

Da desigualdade acima a sequência $\{\|x^k - \overline{x}\|\}$ é decrescente, não negativa e limitada. Portanto $\{\|x^k - \overline{x}\|\}$ é convergente. Passando ao limite em (1.15) o lado direito é igual a zero, logo

$$\lim_{k\to\infty} \|x^{k+1} - x^k\|^2 = 0.$$

Assim, temos que

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

Passo 4: Finalmente, obtemos que os pontos de acumulação da sequência $\{x^k\}$ pertencem a S^*.

Temos que $\{x^k\}$ é limitada, logo possui subsequência $\{x^{k_j}\} \subset \{x^k\}$ convergente. Seja $\overline{x}$ um ponto de acumulação de $\{x^k\}$ tal que $\lim_{j\to\infty} x^{k_j} = \overline{x}$. Do **Passo 3**, reduzindo os termos à subsequência, obtemos que

$$\lim_{j\to\infty} x^{k_j} = \lim_{j\to\infty} x^{k_j+1} = \overline{x}.$$

Passando ao limite em (1.14) e usando o fato de f ser continuamente diferenciável com $\lambda_k \leq \widetilde{\lambda}$, conclui-se que $\nabla f(\overline{x}) = 0$, daí $\overline{x} \in S^*$.

Dessa forma, como $\overline{x} \in S^*$ e

$$\|x^{k_j+1} - \overline{x}\|^2 \leq \|x^{k_j} - \overline{x}\|^2,$$

temos que a subsequência $\{\|x^{k_j} - \overline{x}\|\}$ converge para zero.

A sequência $\{\|x^k - \overline{x}\|\}$ é decrescente, não negativa e possui uma subsequência convergindo a zero, então ela toda converge a zero. Portanto, a sequência $\{x^k\}$ converge a $\overline{x}$. $\square$

1.11 Fejér convergência

A noção da quase-Fejér convergência, utilizada nos artigos estudados nesta dissertação, além da garantia da limitação das sequências descritas, generaliza os resultados apresentados em [3] e [4], de fraca à convergência completa.

Definição 1.11.1. ([6], definição 4.2, pg. 10) Seja $U \subset \mathbb{R}^n$.

(i) Uma sequência $\{y^k\} \subset U$ é dita *convergente fortemente* para $y \in U$ ($y^k \longrightarrow y$) se e somente se,

$$0 = \lim_{k\to\infty} \|y^k - y\|^2 = \lim_{k\to\infty} \langle y^k - y, y^k - y\rangle.$$

(ii) Uma sequência $\{y^k\} \subset U$ é dita *convergente fracamente* para $y \in U$ ($y^k \xrightarrow{w} y$) se e somente se,

$$0 = \lim_{k\to\infty} \langle y^k - y, x\rangle \text{ para todo } x \in U.$$

Em outras palavras, uma sequência $\{y^k\} \subset U$ é dita fracamente convergente para $y \in U$ se $\{y^k\}$ é limitada, $y^{k+1} - y^k$ converge para zero e, cada ponto de acumulação de $\{y^k\}$ pertence a U.

Temos que $y^k \longrightarrow y$ implica trivialmente (via Cauchy-Schwartz) em $y^k \xrightarrow{w} y$ mas, a implicação oposta não é assegurada.

Exemplo 5. A sequência $\{y^k\} \subset \mathbb{R}_+$, $y^k = (1+\frac{1}{k})^k$ é fracamente convergente ao conjunto $\mathbb{R}_+$.

Definição 1.11.2. ([6], definição 1, pg. 4) Uma sequência $\{y^k\} \subset \mathbb{R}^n$ é dita *Fejér convergente* para um conjunto $U \neq \emptyset$ ($U \subset \mathbb{R}^n$), com respeito a distância Euclidiana se,

$$\|y^{k+1} - u\| \leq \|y^k - u\|, \forall\, u \in U.$$

Exemplo 6. A sequência $\{y^k\} \subset \mathbb{R}_+$, $y^k = \sqrt[k]{k}$ é Fejér convergente ao conjunto $\mathbb{R}_+$.

O resultado abaixo é de grande importância pois garante que toda sequência Fejér convergente é limitada. Um resultado análogo é provado posteriormente para sequências quase-Fejér convergente, caso este, de grande importância para este trabalho.

Proposição 1.11.1. ([6], proposição 2.1, pg. 4) *Se $\{y^k\}$ é Fejér convergente a um conjunto não vazio U, então $\{y^k\}$ é limitada. Se, além disso, um ponto de acumulação y de $\{y^k\}$ pertence a U, então $y = \lim\limits_{k\to\infty} y^k$.*

Demonstração: Tome $u \in U$. Ajustando iterativamente a Definição 1.11.2, obtemos:

$$\begin{aligned} \|y^k - u\| &\leq \|y^{k-1} - u\| \\ &\leq \|y^{k-2} - u\| \\ &\leq \ldots \leq \|y^0 - u\|. \end{aligned}$$

Assim segue que $\{y^k\}$ é limitada, já que

$$y^k \in B(u, \|y^0 - u\|).$$

Seja agora $y \in U$ um ponto de acumulação de $\{y^k\}$ e, $\{y^{k_j}\}$ uma subsequência de $\{y^k\}$ tal que $\lim\limits_{j\to\infty} y^{k_j} = y$. Dessa forma, a subsequência $\{\|y^{k_j} - y\|\}$ converge para zero. Como a sequência $\{\|y^k - y\|\}$ é decrescente, não negativa e possui uma subsequência convergindo para zero, então toda ela converge para zero, isto é,

$$0 = \lim_{k\to\infty} \|y^k - y\|.$$

Portanto

$$\lim_{k\to\infty} y^k = y.$$

□

Definição 1.11.3. ([5], definição 1, pg. 128) Uma sequência $\{y^k\}$ é dita *quase-Fejér convergente* a um conjunto $U \neq \emptyset$ ($U \subset \mathbb{R}^n$) se para todo $u \in U$, existe uma sequência $\{\varepsilon_k\} \subseteq \mathbb{R}$, tal que

$$\varepsilon_k \geq 0, \ \sum_{k=0}^{\infty} \varepsilon_k < \infty \ \text{ e } \ \|y^{k+1} - u\|^2 \leq \|y^k - u\|^2 + \varepsilon_k.$$

Exemplo 7. Dadas as sequências $\{y^k\} \subset \mathbb{R}_+$, $y^k = \sqrt[k]{k}$ e $\{\varepsilon_k\} \subset \mathbb{R}$, $\varepsilon_k = \frac{1}{2^k}$, teremos que $\{y^k\}$ é quase-Fejér convergente ao conjunto $\mathbb{R}_+$.

O resultado seguinte foi provado em ([1], teorema 4.1) para distâncias mais gerais. Apresentamos aqui a prova para a distância Euclidiana.

Proposição 1.11.2. ([5], proposição 4, pg. 128) *Se $\{y^k\}$ é quase-Fejér convergente a um conjunto não vazio U, então $\{y^k\}$ é limitada. Se, além disso, um ponto de acumulação y de $\{y^k\}$ pertence a U, então $y = \lim_{k\to\infty} y^k$.*

Demonstração: Tome $u \in U$. Ajustando iterativamente a Definição 1.11.3, obtemos:

$$\begin{aligned} \|y^k - u\|^2 &\leq \|y^{k-1} - u\|^2 + \varepsilon_{k-1} \\ &\leq \|y^{k-2} - u\|^2 + \varepsilon_{k-1} + \varepsilon_{k-2} \\ &\leq \ ... \leq \|y^0 - u\|^2 + \sum_{l=0}^{k-1} \varepsilon_l \\ &\leq \|y^0 - u\|^2 + \beta \end{aligned}$$

onde $\beta = \sum_{k=0}^{\infty} \varepsilon_k < \infty$. Assim segue que $\{y^k\}$ é limitada, já que

$$\|y^k - u\| \leq \sqrt{\|y^0 - u\|^2 + \beta}$$

e portanto,

$$y^k \in B(u, \sqrt{\|y^0 - u\|^2 + \beta}).$$

Seja agora $y \in U$ um ponto de acumulação de $\{y^k\}$ e tome algum $\delta > 0$. Seja $\{y^{k_j}\}$ uma subsequência de $\{y^k\}$ tal que $\lim\limits_{j\to\infty} y^{k_j} = y$. Visto que $y \in U$, existe $\{\varepsilon_k\}$ satisfazendo as propriedades da Definição 1.14.3.

Como $\displaystyle\sum_{k=0}^{\infty} \varepsilon_k < \infty$, tomemos k_0 tal que $\displaystyle\sum_{k=k_0}^{\infty} \varepsilon_k \leq \frac{\delta}{2}$. Seja $\overline{k}$ tal que $l_{\overline{k}} \geq k_0$ e $\|y^{l_{\overline{k}}} - y\|^2 \leq \frac{\delta}{2}$. Então, para algum $k \geq l_{\overline{k}}$,

$$\begin{aligned}
\|y^k - y\|^2 &\leq \|y^{l_{\overline{k}}} - y\|^2 + \sum_{i=l_{\overline{k}}}^{k-1} \varepsilon_i \\
&\leq \frac{\delta}{2} + \sum_{i=k_0}^{\infty} \varepsilon_i \\
&\leq \frac{\delta}{2} + \frac{\delta}{2} = \delta.
\end{aligned}$$

Como δ é arbitrário, isto segue que

$$y = \lim_{k\to\infty} y^k.$$

$\square$

Observação: Toda sequência Fejér convergente é quase-Fejér convergente. Para isso, basta tomar $\varepsilon_k = 0$.

1.12 Resultados Auxiliares

O teorema abaixo é uma versão, sem diferenciabilidade na primeira variável, do Teorema da Função Implícita. Este resultado é aplicado na Seção 3.2, Proposição 3.2.1.

Teorema 1.12.1. ([14], pg. 163) *Seja $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ tal que*

(i) *Existe $(x_0, u_0) \in \mathbb{R}^n \times \mathbb{R}$ tal que $F(x_0, u_0) = 0$;*

(ii) *F é continua em uma vizinhança de (x_0, u_0);*

(iii) *F é diferenciável com relação a variável u em (x_0, u_0) e $\frac{\partial F}{\partial u}(x_0, u_0) \neq 0$;*

Então existe uma vizinhança $V(x_0, u_0)$ de x_0 e, pelo menos uma função $u : V(x_0) \to \mathbb{R}$ tal que $u(x_0) = u_o$ e,

$$F(x, u(x)) = 0 \textit{ para algum } x \in V(x_0).$$

Se, além disso,

(iv) $\frac{\partial F}{\partial u}$ *é continua em (x_0, u_0),*
então, a função u é a única que satisfaz (1.14) e é continua em x_0.

O resultado seguinte é aplicado na Seção 3.3, Proposição 3.3.1.

Teorema 1.12.2. ([16], teorema 1.5.2, pg. 163)**(Fórmula de Newton-Leibnitz)** *Seja $F : \mathbb{R}^n \to \mathbb{R}^l$ uma função continuamente diferenciável. Então para todo $x \in \mathbb{R}^n$ tem-se que*

$$F(x + y) = F(x) + \int_0^1 F'(x + ty)y \, dt.$$

Capítulo 2

Funções pseudo-convexas e busca linear exata

Iniciamos este capítulo relembrando o método do gradiente. Em seguida, é feita a composição da semi-reta $x^k - \alpha\nabla f(x^k)$ com o passo iterativo do algoritmo do ponto proximal. Com o auxílio da busca linear unidimensional mostramos que, para uma função f pseudo-convexa ocorre a convergência completa da sequência (gerada após a composição) para um minimizador de f. Observado que a condição de limitação do conjunto de nível nesta situação é relevante.

2.1 Regularização proximal

Dada uma função $f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciável, o método de máxima descida para resolver o problema

$$\min f(x) \text{ sujeito a } x \in \mathbb{R}^n$$

é dado por

Inicialização:

$$x^0 \in \mathbb{R}^n$$

Passo Iterativo:
Se $\nabla f(x^k) = 0$, PARE. Caso contrário, calcule

$$\alpha_k = \arg\min_{\alpha \geq 0} f(x^k - \alpha\nabla f(x^k)) \tag{2.1}$$

e faça

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k). \tag{2.2}$$

A regularização proximal consiste em compor a semi-reta $x^k - \alpha \nabla f(x^k)$ com o passo iterativo do método do ponto proximal $f(x) + \lambda_k \|x - x^k\|^2$. Dessa forma, tomando $x = x^k - \alpha \nabla f(x^k)$ obtemos

$$\begin{aligned} f(x) + \lambda_k \|x - x^k\|^2 &= f(x^k - \alpha \nabla f(x^k)) + \lambda_k \|x^k - \alpha \nabla f(x^k) - x^k\|^2 \\ &= f(x^k - \alpha \nabla f(x^k)) + \lambda_k \| - \alpha \nabla f(x^k)\|^2 \\ &= f(x^k - \alpha \nabla f(x^k)) + \lambda_k \alpha^2 \|\nabla f(x^k)\|^2. \end{aligned}$$

Logo, a regularização proximal do método de máxima descida substitui (2.1) e (2.2) por

Inicialização:

$$x^0 \in \mathbb{R}^n$$

Passo Iterativo:
Se $\nabla f(x^k) = 0$, PARE. Caso contrário, calcule

$$\alpha_k = \arg \min_{\alpha \geq 0} \{ f(x^k - \alpha \nabla f(x^k)) + \lambda_k \alpha^2 \|\nabla f(x^k)\|^2 \} \tag{2.3}$$

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k), \tag{2.4}$$

onde $\{\lambda_k\} \subset \mathbb{R}$ satisfaz

$$\widehat{\lambda} \leq \lambda_k \leq \widetilde{\lambda} \tag{2.5}$$

para algum $\widehat{\lambda}$, $\widetilde{\lambda}$ tal que $0 < \widehat{\lambda} \leq \widetilde{\lambda}$.

2.2 Boa definição da regularização proximal

Nesta seção mostramos que a regularização proximal está bem definida. Para tal, assumimos que o conjunto de minimizadores de f, denominado $\overline{S}$ é diferente do vazio, isto é, o problema

$$\min f(x) \text{ sujeito a } x \in \mathbb{R}^n$$

tem solução e seja $f^* = \min_{x \in \mathbb{R}^n} f(x)$.

Proposição 2.2.1. ([5], proposição 1, pg. 126) *Seja $f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciável. A sequência $\{x^k\}$, gerada por (2.3) e (2.4), está bem definida e para todo $k \in \mathbb{N}$,*

$$\nabla f(x^{k+1})^T \nabla f(x^k) = 2\alpha_k \lambda_k \|\nabla f(x^k)\|^2. \tag{2.6}$$

Demonstração: Afirmamos que o problema (2.3) tem solução. Definamos $\varphi_k : \mathbb{R}_+ \to \mathbb{R}_+$ como

$$\varphi_k(\alpha) = f(x^k - \alpha \nabla f(x^k)) + \alpha^2 \lambda_k \|\nabla f(x^k)\|^2.$$

Se $\nabla f(x^k) = 0$, então x^k será ponto estacionário e candidato a mínimo. Assim

$$\varphi_k(\alpha) = f(x^k) \in \mathbb{R},$$

e $\varphi_k(\alpha)$ é constante. Como $\nabla f(x^k) = 0$, de (2.4) temos que $x^{k+1} = x^k$ para qualquer escolha de α_k. Logo, o problema (2.3) tem solução e os termos da sequência são todos constantes e iguais à x^k.

Por outro lado, se $\nabla f(x^k) \neq 0$,

$$\begin{aligned} \varphi_k(\alpha) &= f(x^k - \alpha \nabla f(x^k)) + \alpha^2 \lambda_k \|\nabla f(x^k)\|^2 \\ &\geq f^* + \alpha^2 \widehat{\lambda} \|\nabla f(x^k)\|^2, \end{aligned}$$

pois $f^* = \min\limits_{x \in \mathbb{R}^n} f(x)$ e $\widehat{\lambda} < \lambda_k < \widetilde{\lambda}$.

Ao aplicarmos o limite com $\alpha \to \infty$, obtemos

$$\lim_{\alpha \to \infty} \varphi_k(\alpha) \geq \lim_{\alpha \to \infty} [f^* + \alpha^2 \widehat{\lambda} \|\nabla f(x^k)\|^2] = \infty.$$

Desta forma, a minimização em (2.3) reduz-se à um intervalo limitado. Como φ_k é continua e coerciva, a existência de α_k é garantida pelo Corolário 1.4.2.

Logo, a afirmação é verificada e a sequência gerada por (2.3) e (2.4) está bem definida.

Derivando $\varphi_k(\alpha)$ com respeito a α, temos que

$$\varphi_k'(\alpha) = -\nabla f(x^k - \alpha \nabla f(x^k))^T \nabla f(x^k) + 2\alpha \lambda_k \|\nabla f(x^k)\|^2.$$

(i) Para $\alpha = 0$

$$0 = \varphi_k'(0) = -\|\nabla f(x^k)\|^2 \leq 0$$

implicando $\nabla f(x^k) = 0$. Neste caso, (2.6) é verificado trivialmente.

(ii) Para $\alpha = \alpha_k > 0$ qualquer

$$\begin{aligned} 0 &= \varphi_k'(\alpha_k) \\ &= -\nabla f(x^k - \alpha_k \nabla f(x^k))^T \nabla f(x^k) + 2\alpha_k \lambda_k \|\nabla f(x^k)\|^2. \end{aligned}$$

Pela igualdade (2.4) temos que

$$0 = -\nabla f(x^{k+1})^T \nabla f(x^k) + 2\alpha_k \lambda_k \|\nabla f(x^k)\|^2.$$

Então,

$$\nabla f(x^{k+1})^T \nabla f(x^k) = 2\alpha_k \lambda_k \|\nabla f(x^k)\|^2.$$

□

2.3 Análise de convergência

Nesta seção, faremos uma análise de convergência com respeito a regularização proximal proposta para o cálculo do comprimento do passo no método de máxima descida, utilizado para a minimização da função $f : \mathbb{R}^n \to \mathbb{R}$ pseudo-convexa e continuamente diferenciável. Dentre os tópicos abaixo, podemos destacar a estimativa para o valor do comprimento do passo α_k e, o decrescimento dos valores da função objetivo.

Proposição 2.3.1. ([5], proposição 2, pg. 126) *Sejam $f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciável e $\{x^k\}$ a sequência gerada por (2.3) e (2.4). Então*

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

(ii) $\{f(x^k)\}$ *é decrescente e convergente,*

(iii) $\displaystyle\sum_{k=0}^{\infty} \|x^{k+1} - x^k\|^2 < \infty$,

(iv) $\displaystyle\lim_{k\to\infty} (x^{k+1} - x^k) = 0$

Demonstração:
(i) Tomando φ_k como definida na Proposição 2.2.1 temos que

$$\begin{aligned} \varphi_k(\alpha_k) &= f(x^k - \alpha_k \nabla f(x^k)) + \alpha_k^2 \lambda_k \|\nabla f(x^k)\|^2 \\ &= f(x^k - \alpha_k \nabla f(x^k)) + \lambda_k \|\alpha_k \nabla f(x^k)\|^2. \end{aligned}$$

Utilizando (2.4) nas parcelas da soma acima, obtemos

$$\varphi_k(\alpha_k) = f(x^{k+1}) + \lambda_k \|x^{k+1} - x^k\|^2.$$

Por (2.3), segue que

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

e dessa forma, $\varphi_k(\alpha) \geq \varphi_k(\alpha_k)$ para todo α.

Em particular, para $\alpha = 0$, vale que $\varphi_k(0) \geq \varphi_k(\alpha_k)$. Assim

$$f(x^k) = \varphi_k(0) \geq \varphi_k(\alpha_k) = f(x^{k+1}) + \lambda_k \|x^{k+1} - x^k\|^2.$$

Logo,

$$f(x^k) \geq f(x^{k+1}) + \lambda_k \|x^{k+1} - x^k\|^2, \tag{2.7}$$

o que comprova o lado direito da inequação.

O lado esquerdo é trivial, pois para $\lambda_k \geq 0$ e $f^* = \min_{x \in \mathbb{R}^n} f(x)$, vale que

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

Portanto,

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

(ii) De (2.7) temos que

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

Como $\lambda_k > 0$, então

$$0 < \lambda_k \|x^{k+1} - x^k\|^2 \leq f(x^k) - f(x^{k+1}).$$

Dessa forma,

$$f(x^{k+1}) \leq f(x^k) \leq ... \leq f(x^0).$$

Assim, $\{f(x^k)\}$ é decrescente (portanto, monótona) e limitada. Logo, pelo Teorema 1.1.2, temos que $\{f(x^k)\}$ é convergente.

(iii) De (2.7)

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

Por (2.5), temos que

$$\widehat{\lambda}\|x^{k+1} - x^k\|^2 \leq \lambda_k\|x^{k+1} - x^k\|^2 \leq f(x^k) - f(x^{k+1}).$$

Logo

$$\widehat{\lambda}\|x^{k+1} - x^k\|^2 \leq f(x^k) - f(x^{k+1}). \tag{2.8}$$

Aplicando o somatório, segue que

$$\sum_{k=0}^{l}\widehat{\lambda}\|x^{k+1} - x^k\|^2 \leq \sum_{k=0}^{l}(f(x^k) - f(x^{k+1})).$$

Isto implica

$$\widehat{\lambda}\sum_{k=0}^{l}\|x^{k+1} - x^k\|^2 \leq [f(x^0) - f(x^1)] + [f(x^1) - f(x^2)] + ... + [f(x^l) - f(x^{l+1})].$$

Como $f^* = \min_{x \in \mathbb{R}^n} f(x)$, segue que

$$\widehat{\lambda}\sum_{k=0}^{l}\|x^{k+1} - x^k\|^2 \leq f(x^0) - f(x^{l+1}) \leq f(x^0) - f^*.$$

Logo

$$\sum_{k=0}^{l}\|x^{k+1} - x^k\|^2 \leq \frac{1}{\widehat{\lambda}}[f(x^0) - f^*] < \infty,$$

e portanto

$$\sum_{k=0}^{\infty}\|x^{k+1} - x^k\|^2 < \infty.$$

(iv) De (2.8) temos que

$$\|x^{k+1} - x^k\|^2 \leq \frac{1}{\widehat{\lambda}}[f(x^k) - f(x^{k+1})].$$

Aplicando ao limite, com $k \to \infty$ obtemos

$$\lim_{k\to\infty}\|x^{k+1} - x^k\|^2 \leq \frac{1}{\widehat{\lambda}}\lim_{k\to\infty}(f(x^k) - f(x^{k+1})).$$

Por (ii) segue que $\{f(x^k)\}$ é convergente e dessa forma, o lado direito da desigualdade converge para zero. Logo

$$\lim_{k\to\infty} \|x^{k+1} - x^k\|^2 = 0,$$

e isto implica

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

□

O resultado seguinte nos mostra que, se $\overline{x}$ é um ponto de acumulação da sequência $\{x^k\}$ então satisfaz a condição necessária de primeira ordem para o problema

$$\min f(x) \text{ sujeito a } x \in \mathbb{R}^n,$$

e claramente, será um candidato a minimo.

Proposição 2.3.2. ([5], proposição 3, pg. 127) *Seja $f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciável. Se $\overline{x}$ é um ponto de acumulação da sequência $\{x^k\}$ gerada por (2.3) e (2.4) então $\nabla f(\overline{x}) = 0$.*

Demonstração: Seja $\{x^{k_j}\}$ uma subsequência de $\{x^k\}$ tal que

$$\lim_{j\to\infty} x^{k_j} = \overline{x}.$$

Pela igualdade (2.6), reduzindo os termos à subsequência, obtemos

$$\begin{aligned} \nabla f(x^{k_j+1})^T \nabla f(x^{k_j}) &= 2\alpha_{k_j}\lambda_{k_j}\|\nabla f(x^{k_j})\|^2 \\ &= 2\alpha_{k_j}\lambda_{k_j}\|\nabla f(x^{k_j})\|.\|\nabla f(x^{k_j})\| \\ &= 2\lambda_{k_j}\|\ \alpha_{k_j}\nabla f(x^{k_j})\|.\|\nabla f(x^{k_j})\|. \end{aligned}$$

Utilizando a igualdade (2.4) obtemos

$$\nabla f(x^{k_j+1})^T \nabla f(x^{k_j}) = 2\lambda_{k_j}\|\ x^{k_j+1} - x^{k_j}\|.\|\nabla f(x^{k_j})\|.$$

Por (2.5) temos que $\widehat{\lambda} \leq \lambda_k \leq \widetilde{\lambda}$. Logo,

$$\nabla f(x^{k_j+1})^T \nabla f(x^{k_j}) \leq 2\widetilde{\lambda}\|x^{k_j+1} - x^{k_j}\|.\|\nabla f(x^{k_j})\|.$$

Como f é continuamente diferenciável, ao aplicarmos o limite com $j \to \infty$, obtemos

$$\nabla f(\lim_{j\to\infty} x^{k_j+1})^T \nabla f(\lim_{j\to\infty} x^{k_j}) \leq 2\widetilde{\lambda}\|\lim_{j\to\infty}(x^{k_j+1} - x^{k_j})\|.\|\nabla f(\lim_{j\to\infty} x^{k_j})\|.$$

Isto implica

$$\nabla f(\overline{x})^T \nabla f(\overline{x}) \leq 2\widetilde{\lambda} \| \lim_{j\to\infty} (x^{k_j+1} - x^{k_j}) \| . \| \nabla f(\overline{x}) \|. \tag{2.9}$$

Pela Proposição 2.3.1 (iv), ao reduzirmos os termos à subsequência, temos que

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

Aplicando este resultado em (2.9), observamos que o lado direito da desigualdade converge para zero. Como o lado esquerdo resulta em $\|\nabla f(\overline{x})\|^2$, segue que

$$\|\nabla f(\overline{x})\|^2 \leq 0.$$

Dessa forma,

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

□

Podemos observar que todos os resultados anteriores foram verificados utilizando somente a hipótese de continuamente diferenciável para a função f. No resultado abaixo, utilizamos a hipótese da pseudo-convexidade para f e mostramos a quase-Fejér convergência da sequência $\{x^k\}$ para o conjunto de minimizadores de f onde, pela Proposição 1.11.2 temos a garantia de limitação da sequência $\{x^k\}$.

Proposição 2.3.3. ([5], proposição 5, pg. 128) *Seja $f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciável. Se f é pseudo-convexa então a sequência $\{x^k\}$ gerada por (2.3) e (2.4), é quase-Fejér convergente ao conjunto de minimizadores da f.*

Demonstração: Seja $z \in \overline{S}$. Dada a soma abaixo, desenvolvendo o quadrado da norma em cada parcela, obtemos

$$\begin{aligned} \|z - x^{k+1}\|^2 - \|z - x^k\|^2 - \|x^k - x^{k+1}\|^2 &= \|z\|^2 + \|x^{k+1}\|^2 - 2\langle z, x^{k+1}\rangle \\ &\quad -\|z\|^2 - \|x^k\|^2 + 2\langle z, x^k\rangle \\ &\quad -\|x^k\|^2 - \|x^{k+1}\|^2 + 2\langle x^k, x^{k+1}\rangle \\ &= -2\|x^k\|^2 + 2\langle z, x^k - x^{k+1}\rangle \\ &\quad +2\langle x^k, x^{k+1}\rangle \end{aligned}$$

$$
\begin{aligned}
&= -2\langle x^k, x^k\rangle + 2\langle z, x^k - x^{k+1}\rangle \\
&\quad +2\langle x^k, x^{k+1}\rangle \\
&= -2\langle x^k, x^k - x^{k+1}\rangle + 2\langle z, x^k - x^{k+1}\rangle \\
&= 2\langle z - x^k, x^k - x^{k+1}\rangle.
\end{aligned}
$$

Pela igualdade (2.4) temos que

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

Isto implica

$$
\|z - x^{k+1}\|^2 - \|z - x^k\|^2 - \|x^k - x^{k+1}\|^2 = 2\alpha_k (z - x^k)^T \nabla f(x^k). \tag{2.10}
$$

Afirmamos que $(z - x^k)^T \nabla f(x^k) \leq 0$. Pois caso contrário, teríamos $(z - x^k)^T \nabla f(x^k) > 0$ e como f é pseudo-convexa, isto implica $f(z) \geq f(x^k)$. Dessa forma, x^k também é minimizador de f implicando $\nabla f(x^k) = 0$, o que contradiz $(z - x^k)^T \nabla f(x^k) > 0$.

Assim, de (2.10) temos que

$$
\|z - x^{k+1}\|^2 - \|z - x^k\|^2 - \|x^k - x^{k+1}\|^2 \leq 0.
$$

Logo,

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

Observamos que, para $\varepsilon_k = \|x^{k+1} - x^k\|^2$ teremos $\varepsilon_k \geq 0$. Pela Proposição 2.3.1 (iii) temos que

$$
\sum_{k=0}^{\infty} \|x^{k+1} - x^k\|^2 < \infty.
$$

Portanto, pela Definição 1.11.3, $\{x^k\}$ é quase-Fejér convergente ao conjunto de minimizadores de f. □

2.4 Primeiro teorema principal

Mostraremos agora que todo ponto de acumulação de $\{x^k\}$ pentence ao conjunto de minimizadores de f. Como consequência, temos a convergência completa da sequência $\{x^k\}$ ao conjunto de minimizadores de f.

Teorema 2.4.1. ([5], teorema 1, pg. 129) *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *continuamente diferenciável. Se* f *é pseudo-convexa e atinge seu mínimo em* $\mathbb{R}^n$ *então a sequência* $\{x^k\}$ *gerada por (2.3) e (2.4) converge para um minimizador de* f*.*

Demonstração: Pela Proposição 2.3.3, $\{x^k\}$ é quase-Fejér convergente ao conjunto de minimizadores de f. Logo, pela primeira parte da Proposição 1.11.2, a sequência $\{x^k\}$ é limitada e, dessa forma, possui uma subsequência $\{x^{k_j}\}$ convergente. Seja $\overline{x} \in \mathbb{R}^n$ um ponto de acumulação de $\{x^k\}$ tal que

$$\lim_{j\to\infty} x^{k_j} = \overline{x}.$$

Pela Proposição 2.3.2, $\overline{x}$ é ponto estacionário de f, isto é, $\nabla f(\overline{x}) = 0$. Por hipótese, f é pseudo-convexa e $\nabla f(\overline{x}) = 0$ satisfaz

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

Assim, segue por definição que

$$f(x) \geq f(\overline{x}).$$

Logo, $\overline{x}$ é mínimo de f.

Como $\{x^k\}$ é quase-Fejér convergente ao conjunto de minimizadores de f e $\overline{x}$ é ponto de acumulação de $\{x^k\}$, segue da segunda parte da Proposição 1.11.2. que

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

□

Capítulo 3

Funções convexas e busca linear inexata

Neste capítulo apresentamos dois algoritmos para minimizar uma função $f : \mathbb{R}^n \to \mathbb{R}$ convexa e continuamente diferenciável, utilizando para isto o método do gradiente com busca linear inexata. No primeiro algoritmo, o tamanho do comprimento do passo é pré-fixado e exigimos que o gradiente de f satisfaça a condição de Lipschitz com constante $L > 0$ conhecida. No segundo algoritmo, o cálculo do comprimento de passo é baseado no trabalho desenvolvido por Dennis-Schnabel, onde λ_k deve satisfazer somente uma das inequações propostas em [4]. Para os dois casos, não é necessário a hipótese de limitação dos conjuntos de nível da função objetivo e mesmo assim, provamos a convergência completa das sequências geradas para um minimizador de f.

3.1 Os Algoritmos

Dada uma função $f : \mathbb{R}^n \to \mathbb{R}$ convexa e continuamente diferenciável. Consideramos o conjunto $\widetilde{S}$ de minimizadores de f não vazio e $f^* = \min\limits_{x \in \mathbb{R}^n} f(x)$.

O método do gradiente para resolver o problema

$$\min f(x) \text{ sujeito a } x \in \mathbb{R}^n$$

é dado por

Inicialização:

$$x^0 \in \mathbb{R}^n \tag{3.1}$$

Passo Iterativo:

Se $\nabla f(x^k) = 0$, PARE. Caso contrário, calcule

$$x^{k+1} = x^k - \lambda_k \nabla f(x^k). \tag{3.2}$$

Para o cálculo do tamanho do passo λ_k utilizamos um dos seguintes critérios:

Algoritmo A (conhecido L)**:** Neste algoritmo partimos do pressuposto que o gradiente da função objetivo seja Lipschitz-contínuo com constante $L > 0$ conhecida.

Fixamos duas constantes positivas δ_1, δ_2, com $0 < \delta_1 < \delta_2 < 1$ tais que

$$\frac{L}{2}\,\delta_1 + \delta_2 < 1. \tag{3.3}$$

Assim:

Inicialização:

$$x^0 \in \mathbb{R}^n$$

Passo Iterativo:

Se $\nabla f(x^k) = 0$, PARE. Caso contrário, escolha λ_k satisfazendo

$$\delta_1 \leq \lambda_k \leq \frac{2}{L}\,(1 - \delta_2), \tag{3.4}$$

e faça

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

Algoritmo B: Neste algoritmo precisamos conhecer uma função $\psi : \mathbb{R}_+ \to \mathbb{R}$ satisfazendo as seguintes condições

B1) ψ é convexa e continuamente diferenciável;

B2) $\psi(0) = 0$ e $\psi'(0) < 1$;

B3) $\displaystyle\lim_{u \to 0^+} \frac{\psi(u)}{u^2} > 0$.

Observação: Temos que B3 implica $\psi'(0) \geq 0$ pois

$$\psi'(0) = \lim_{u \to 0^+} \frac{\psi(u) - \psi(0)}{u - 0} = \lim_{u \to 0^+} \frac{\psi(u)}{u} \;\geq\; \lim_{u \to 0^+} \frac{\psi(u)}{u^2} > 0.$$

Portanto, utilizando B2 obtemos $0 \leq \psi'(0) < 1$ e, por B1 temos que ψ é não-decrescente.

Além disso, devemos escolher duas constantes positivas δ_1, δ_2 tais que $0 < \delta_1 < \delta_2 < 1$. Dessa forma,

Inicialização:

$$x^0 \in \mathbb{R}^n$$

Passo Iterativo:

Se $\nabla f(x^k) = 0$, PARE. Senão, definamos t_j (para $j = 0, 1, 2...$). Dado

$$\delta_1 < t_0 < \delta_2, \tag{3.5}$$

devemos verificar se t_j satisfaz

$$f(x^k - t_j \nabla f(x^k)) \leq f(x^k) - \psi(t_j)\|\nabla f(x^k)\|^2 \tag{3.6}$$

Em caso afirmativo, tomemos $\lambda_k = t_j$,

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

e a iteração pára. Caso contrário, façamos

$$t_{j+1} = \frac{t_j}{2}.$$

3.2 Boa definição do Algoritmo B

Para o Algoritmo B estar bem definido deve ser estabelecido que a inequação (3.6) seja satisfeita após um número finito de passos.

Seja $G = \{x \in \mathbb{R}^n / \nabla f(x) \neq 0\}$. Pela continuidade e diferenciabilidade de f, G é aberto.

A proposição seguinte apresenta uma estimativa para o parâmetro t_j utilizado para o cálculo do comprimento do passo no Algoritmo B.

Proposição 3.2.1. ([2], proposição 1, pg. 142) *Seja $f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciável e ψ satisfazendo B1, B2 e B3. Então*

i) Para todo $x \in G$ existe um único $u(x) > 0$ tal que

$$f(x - u(x)\nabla f(x)) = f(x) - \psi(u(x))\|\nabla f(x)\|^2 \tag{3.7}$$

e

$$f(x - \overline{u}\nabla f(x)) \leq f(x) - \psi(\overline{u})\|\nabla f(x)\|^2, \tag{3.8}$$

se e somente se, $0 \leq \overline{u} \leq u(x)$.

ii) $u : G \to \mathbb{R}_+$ *é continua em* G.

Demonstração:

(i) Fixemos $x \in G$ e definamos $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ onde

$$F(x, \overline{u}) = f(x - \overline{u}\nabla f(x)) - f(x) + \psi(\overline{u})\|\nabla f(x)\|^2. \tag{3.9}$$

Como f é convexa e continuamente diferenciável e, por B1, ψ é convexa e continuamente diferenciável então, $F(x, .)$ é convexa e continuamente diferenciável.

De B2 temos que $\psi(0) = 0$ e $\psi'(0) < 1$. Assim,

$$F(x, 0) = 0, \tag{3.10}$$

$$\frac{\partial F}{\partial u}(x, 0) = \|\nabla f(x)\|^2(\psi'(0) - 1) < 0. \tag{3.11}$$

Como $f^* = \min_{x \in \mathbb{R}^n} f(x)$, de (3.9) segue que

$$F(x, \overline{u}) \geq f^* - f(x) + \psi(\overline{u})\|\nabla f(x)\|^2. \tag{3.12}$$

Observamos de (3.10) e (3.11) que $F(x, .)$ é negativa em algum intervalo à direita de zero. Aplicando ao limite com $\overline{u} \to \infty$ em (3.12), obtemos

$$\lim_{\overline{u} \to \infty} F(x, \overline{u}) = +\infty,$$

visto que ψ é convexa.

Como $F(x, .)$ é uma função real e convexa, segue que existe $u(x) > 0$ tal que $F(x, u(x)) = 0$. Dessa forma, (3.7) é verificada trivialmente. Como $F(x, 0) = F(x, u(x)) = 0$ e, por (3.11), zero não é o valor mínimo de $F(x, .)$ então $u(x)$ é único. Logo,(i) é estabelecida.

(ii) Seja $u_0 := u(x_0)$ dado por (i), para algum $x_0 \in G$. Então, temos que

$$F(x_0, u_0) = F(x_0, u(x_0)) = 0,$$

Sabemos que $F(., .)$ é continua em uma vizinhança de (x_0, u_0) e,

$$\frac{\partial F}{\partial u}(x_0, u_0) = -\nabla f(x_0 - u_0\nabla f(x_0))^T\nabla f(x_0) + \psi'(u_0)\|\nabla f(x_0)\|^2. \tag{3.13}$$

Como $F(x_0, .)$ é estritamente crescente em u_0, temos que

$$\frac{\partial F}{\partial u}(x_0, u_0) > 0.$$

A função f é continuamente diferenciável em (x_0, u_0) e a função ψ é continuamente diferenciável em (x_0, u_0). Concluímos de (13) que $\frac{\partial F}{\partial u}$ é continua em (x_0, u_0).

Dessa forma, todas as hipóteses do Teorema 1.12.1 (Teorema da Função Implicita) são verificadas e assim, u é continua em x_0. $\square$

Verificaremos agora que a sequência gerada pelo Algoritmo B está bem definida. Vamos considerar $u(x^k)$ como definida na Proposição 3.2.1.

Proposição 3.2.2. ([2], proposição 3, pg. 143) *O Algoritmo B, definido por (3.5) e (3.6) pára após um número finito de iterações com*

$$\min \{\delta_1, \frac{u(x^k)}{2}\} \leq \lambda_k \leq \min \{\delta_2, u(x^k)\}. \tag{3.14}$$

Demonstração: Consideremos dois casos para o valor de t_0

1) $t_0 \in (0, u(x^k))$

2) $t_0 \geq u(x^k)$

(1) Como $0 < t_0 < u(x^k)$, de (3.8) segue que

$$f(x - t_0 \nabla f(x)) \leq f(x) - \psi(t_0) \|\nabla f(x)\|^2,$$

que é a verificação de (3.6). Dessa forma, para $\lambda_k = t_0$, a iteração pára com $j = 0$.

Assim, (3.14) é verificada pois, em (3.5) temos que $\delta_1 < t_0 < \delta_2$ e, como consideramos $0 < t_0 < u(x^k)$ então $t_0 < \delta_2$ e $t_0 < u(x^k)$. Isto implica $\lambda_k = t_0 < \min\{\delta_2, u(x^k)\}$. Da mesma forma temos que $t_0 > \delta_1$ e $\lambda_k = t_0 \geq \min\{\delta_1, \frac{u(x^k)}{2}\}$.

(2) Existe um único $s \in \mathbb{N}$, $s \geq 1$, tal que

$$2^{s-1} u(x^k) < t_0 \leq 2^s u(x^k). \tag{3.15}$$

Então

$$\frac{u(x^k)}{2} \leq \frac{t_0}{2^s} \leq u(x^k). \tag{3.16}$$

Por (3.5) e (3.6), temos que $t_j = \frac{t_0}{2^j}$ pois

$$t_1 = \frac{t_0}{2} \Rightarrow t_2 = \frac{t_1}{2} = \frac{t_0}{2^2} \Rightarrow ... \Rightarrow t_j = \frac{t_0}{2^j}.$$

Assim, (3.16) nos mostra que

$$\frac{u(x^k)}{2} < t_s \leq u(x^k). \tag{3.17}$$

Afirmamos que $\lambda_k = t_s$. Pois, de (3.15) temos que

$$u(x^k) < \frac{t_0}{2^{s-1}} \leq 2u(x^k)$$

o que implica,

$$u(x^k) < t_{s-1} \leq 2u(x^k).$$

Dessa forma,

$$t_{s-1} > u(x^k).$$

De (3.16), temos que

$$\frac{u(x^k)}{2} < t_s \leq u(x^k),$$

implicando

$$t_s \leq u(x^k).$$

Logo, usando a Proposição 3.2.1, (3.6) é satisfeita para $\lambda_k = t_s$ mas não para $\lambda_k = t_{s-1}$. Logo, (3.14) segue de (3.17). $\square$

3.3 Análise de convergência

Nesta seção, faremos uma análise de convergência com respeito aos algoritmos propostos para o cálculo do comprimento do passo no método do gradiente, utilizado para a minimização da função $f : \mathbb{R}^n \to \mathbb{R}$ convexa e continuamente diferenciável. Dentre os tópicos abaixo, podemos destacar o decrescimento dos valores da função objetivo.

Proposição 3.3.1. ([2], proposição 4, pg. 143) *Seja $f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciável. Dados os Algoritmos A e B, temos que* *i) Existe* $\gamma > 0$ *tal que*

$$f(x^{k+1}) \leq f(x^k) - \gamma\|x^{k+1} - x^k\|^2, \text{ para todo } k; \tag{3.18}$$

ii) $\{f(x^k)\}$ *é decrescente e convergente;*

iii) $\displaystyle\sum_{k=0}^{\infty} \|x^{k+1} - x^k\|^2 < \infty$

Demonstração:

(i) (a) Para o Algoritmo A, usando o Teorema 1.12.2 (Fórmula de Newton-Leibniz) com $x = x^k$, $y = -\lambda_k \nabla f(x^k)$, $t = u$ e $F = f$, temos que

$$f(x^k - \lambda_k \nabla f(x^k)) = f(x^k) + \int_0^1 \nabla f(x^k - u\lambda_k \nabla f(x^k))^T(-\lambda_k \nabla f(x^k))du.$$

Utilizando a igualdade (3.2) obtemos

$$f(x^{k+1}) = f(x^k) - \lambda_k \int_0^1 \nabla f(x^k - u\lambda_k \nabla f(x^k))^T \nabla f(x^k) du.$$

Somando e subtraindo $\lambda_k\|\nabla f(x^k)\|^2$ no lado direito da igualdade acima, obtemos

$$\begin{aligned} f(x^{k+1}) &= f(x^k) - \lambda_k\|\nabla f(x^k)\|^2 - \\ &\quad \lambda_k \int_0^1 [\nabla f(x^k - u\lambda_k \nabla f(x^k))^T \nabla f(x^k) - \|\nabla f(x^k)\|^2]du \\ &= f(x^k) - \lambda_k\|\nabla f(x^k)\|^2 - \\ &\quad \lambda_k \int_0^1 [(\nabla f(x^k - u\lambda_k \nabla f(x^k)) - \nabla f(x^k))^T \nabla f(x^k)]du. \end{aligned}$$

Implicando

$$\begin{aligned} f(x^{k+1}) - f(x^k) + \lambda_k\|\nabla f(x^k)\|^2 &= -\lambda_k \int_0^1 [(\nabla f(x^k - u\lambda_k \nabla f(x^k)) - \\ &\quad \nabla f(x^k))^T \nabla f(x^k)]du \\ &\leq \lambda_k \int_0^1 |(\nabla f(x^k - u\lambda_k \nabla f(x^k)) - \\ &\quad \nabla f(x^k))^T \nabla f(x^k)|du \\ &\leq \lambda_k \int_0^1 \|\nabla f(x^k - u\lambda_k \nabla f(x^k)) - \\ &\quad \nabla f(x^k)\|.\|\nabla f(x^k)\|du. \end{aligned}$$

Como o gradiente de f é Lipschitz-contínuo,

$$\begin{aligned}
f(x^{k+1}) - f(x^k) + \lambda_k \|\nabla f(x^k)\|^2 &\leq \lambda_k \int_0^1 L\|x^k - u\lambda_k \nabla f(x^k) - \\
&\quad x^k\|.\|\nabla f(x^k)\| du \\
&\leq \lambda_k \int_0^1 L\|u\lambda_k \nabla f(x^k)\|.\|\nabla f(x^k)\| du \\
&\leq \lambda_k^2 L\|\nabla f(x^k)\|^2 \int_0^1 u du \\
&\leq \frac{1}{2}\lambda_k^2 L\|\nabla f(x^k)\|^2.
\end{aligned}$$

Assim,

$$\begin{aligned}
f(x^{k+1}) &\leq f(x^k) - \lambda_k\|\nabla f(x^k)\|^2 + \frac{1}{2}\lambda_k^2 L\|\nabla f(x^k)\|^2 \\
&\leq f(x^k) - \lambda_k(1 - \frac{L\lambda_k}{2})\|\nabla f(x^k)\|^2.
\end{aligned}$$

Da igualdade (3.2) temos que

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

Logo,

$$f(x^{k+1}) \leq f(x^k) - \frac{1}{\lambda_k}(1 - \frac{L\lambda_k}{2})\|x^{k+1} - x^k\|^2. \tag{3.19}$$

Em (3.4) vale que

$$\delta_1 \leq \lambda_k \leq \frac{2(1-\delta_2)}{L},$$

implicando

$$\frac{1}{\delta_1} \geq \frac{1}{\lambda_k} \geq \frac{L}{2(1-\delta_2)}.$$

Multiplicando a segunda desiguldade por $(1 - \frac{L\lambda_k}{2})$ obtemos

$$\frac{1}{\lambda_k}(1 - \frac{L\lambda_k}{2}) \geq \frac{L}{2(1-\delta_2)}(1 - \frac{L\lambda_k}{2}). \tag{3.20}$$

Novamente, de (3.4) temos que

$$\delta_1 \leq \lambda_k \leq \frac{2(1-\delta_2)}{L},$$

implicando

$$\frac{L\delta_1}{2} \leq \frac{L\lambda_k}{2} \leq 1-\delta_2.$$

Logo,

$$1 - \frac{L\delta_1}{2} \geq 1 - \frac{L\lambda_k}{2} \geq \delta_2. \tag{3.21}$$

De (3.20) e (3.21)

$$\frac{1}{\lambda_k}(1 - \frac{L\lambda_k}{2}) \geq \frac{L}{2(1-\delta_2)}(1 - \frac{L\lambda_k}{2}) \geq \frac{L}{2(1-\delta_2)}\delta_2. \tag{3.22}$$

Substituindo (3.22) em (3.19), obtemos

$$\begin{aligned} f(x^{k+1}) &\leq f(x^k) - \frac{1}{\lambda_k}(1 - \frac{L\lambda_k}{2})\|x^{k+1} - x^k\|^2 \\ &\leq f(x^k) - \delta_2 \frac{L}{2(1-\delta 2)}\|x^{k+1} - x^k\|^2. \end{aligned}$$

O que estabelece (3.18) para

$$\gamma = \delta_2 \frac{L}{2(1-\delta_2)}.$$

(b) Para o Algoritmo B, é verificado a validade de

$$f(x^{k+1}) \leq f(x^k) - \psi(\lambda_k)\|\nabla f(x^k)\|^2.$$

Dessa forma

$$f(x^k) - f(x^{k+1}) \geq \psi(\lambda_k)\|\nabla f(x^k)\|^2.$$

Pela igualdade (3.2) temos que

$$f(x^k) - f(x^{k+1}) \geq \psi(\lambda_k)\|\frac{x^{k+1} - x^k}{-\lambda_k}\|^2.$$

Assim,

$$f(x^k) - f(x^{k+1}) \geq \frac{\psi(\lambda_k)}{\lambda_k^2}\|x^{k+1} - x^k\|^2. \tag{3.23}$$

Por B3, temos que $\lim_{u \to 0^+} \dfrac{\psi(u)}{u^2} > 0$ logo, podemos tomar $\xi > 0$ tal que

$$0 < \xi < \lim_{u \to 0^+} \frac{\psi(u)}{u^2}.$$

Pela definição de ξ, existe $\theta > 0$ tal que, se $\lambda \in (0, \theta)$ então

$$\frac{\psi(\lambda)}{\lambda^2} > \xi. \tag{3.24}$$

Para cada k, temos duas possibilidades:

1) $\lambda_k \in (0, \theta)$. Neste caso, por (3.24) vale que $\frac{\psi(\lambda)}{\lambda^2} > \xi$.

2) $\lambda_k \geq \theta$. Neste caso, pela Proposição 3.2.2 temos que

$$\min \{\delta_1, \frac{u(x^k)}{2}\} \leq \lambda_k \leq \min \{\delta_2, u(x^k)\} \leq \delta_2$$

o que implica

$$\lambda_k \leq \delta_2$$

e portanto,

$$\frac{1}{\lambda_k^2} \geq \frac{1}{\delta_2^2}.$$

De B1 e B2 temos que ψ é crescente, implicando em

$$\psi(\lambda_k) \geq \psi(\theta).$$

Dessa forma

$$\frac{\psi(\lambda_k)}{\lambda_k^2} \geq \frac{\psi(\theta)}{\delta_2^2}.$$

Tomando $\gamma = \min\{\xi, \frac{\psi(\theta)}{\delta_2^2}\}$ e utilizando (3.23), provamos a desigualdade (3.18) para o Algoritmo B.

(ii) De (i), vale que

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

Como $\gamma > 0$ temos que

$$0 < \gamma \|x^{k+1} - x^k\|^2 \leq f(x^k) - f(x^{k+1}).$$

Dessa forma,

$$f(x^{k+1}) \leq f(x^k) \leq ... \leq f(x^0).$$

Assim, $\{f(x^k)\}$ é decrescente (portanto, monótona) e limitada. Logo, pelo Teorema 1.1.2, temos que $\{f(x^k)\}$ é convergente.

(iii) Por (i) existe $\gamma > 0$ tal que

$$\begin{aligned} f(x^{k+1}) &\leq f(x^k) - \gamma\|x^{k+1} - x^k\|^2 \\ &\leq f(x^{k-1}) - \gamma\|x^k - x^{k-1}\|^2 - \gamma\|x^{k+1} - x^k\|^2 \\ &\leq f(x^0) - \gamma\|x^1 - x^0\|^2 - ... - \gamma\|x^{k+1} - x^k\|^2 \\ &= f(x^0) - \gamma \sum_{l=0}^{k} \|x^{l+1} - x^l\|^2. \end{aligned}$$

Assim,

$$f(x^{k+1}) \leq f(x^0) - \gamma \sum_{l=0}^{k} \|x^{l+1} - x^l\|^2.$$

Dessa forma,

$$\gamma \sum_{l=0}^{k} \|x^{l+1} - x^l\|^2 \leq f(x^0) - f(x^{k+1}),$$

então,

$$\begin{aligned} \sum_{l=0}^{k} \|x^{l+1} - x^l\|^2 &\leq \frac{1}{\gamma}(f(x^0) - f(x^{k+1})) \\ &\leq \frac{1}{\gamma}(f(x^0) - f^*). \end{aligned}$$

Tomando $l \to \infty$, obtemos

$$\sum_{l=0}^{\infty} \|x^{l+1} - x^l\|^2 < \infty.$$

□

A desigualdade verificada na proposição seguinte é de grande importância para a demonstração da quase-Fejér convergência da sequência $\{x^k\}$, gerada pelo método do gradiente utilizando os Algoritmos A ou B, para o conjunto T definido abaixo.

Proposição 3.3.2. ([2], proposição 2, pg. 142) *Seja $f : \mathbb{R}^n \to \mathbb{R}$ convexa e continuamente diferenciável. Dado o conjunto*

$$T = \{z \in \mathbb{R}^n / f(z) \leq \lim_{k\to\infty} \inf f(x^k)\},$$

temos que

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

para qualquer $z \in T$, onde $\{x^k\}$ é a sequência gerada por (3.1) e (3.2), com qualquer $\lambda_k > 0$.

Demonstração: Seja $z \in T$. Dada a soma abaixo, desenvolvendo o quadrado da norma em cada parcela, obtemos

$$\begin{aligned}
\|x^{k+1} - z\|^2 - \|x^k - z\|^2 - \|x^{k+1} - x^k\|^2 &= \|x^{k+1}\|^2 + \|z\|^2 - 2\langle x^{k+1}, z\rangle \\
&\quad -\|x^k\|^2 - \|z\|^2 + 2\langle x^k, z\rangle \\
&\quad -\|x^{k+1}\|^2 - \|x^k\|^2 + 2\langle x^{k+1}, x^k\rangle \\
&= -2\|x^k\|^2 - 2\langle x^{k+1}, z\rangle + 2\langle x^k, z\rangle \\
&\quad +2\langle x^{k+1}, x^k\rangle \\
&= -2\langle x^k, x^k\rangle - 2\langle x^{k+1}, z\rangle + 2\langle x^k, z\rangle \\
&\quad +2\langle x^{k+1}, x^k\rangle \\
&= 2\langle x^k, z - x^k\rangle - 2\langle x^{k+1}, z - x^k\rangle \\
&= 2\langle x^k - x^{k+1}, z - x^k\rangle \\
&= 2(z - x^k)^T (x^k - x^{k+1})
\end{aligned}$$

Utilizando a igualdade (3.2) obtemos

$$\begin{aligned}
\|x^{k+1} - z\|^2 - \|x^k - z\|^2 - \|x^{k+1} - x^k\|^2 &= 2(z - x^k)^T \lambda_k \nabla f(x^k) \\
&= 2\lambda_k (z - x^k)^T \nabla f(x^k) \\
&\leq 2\lambda_k (f(z) - f(x^k)) \\
&\leq 0
\end{aligned}$$

devido à desigualdade do gradiente aplicada na primeira inequação, pois f é convexa e, o fato de $z \in T$ aplicado na segunda inequação.

Portanto

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

□

O resultado seguinte garante a convergência completa da sequência $\{x^k\}$ para o conjunto $T = \{z \in \mathbb{R}^n / f(z) \leq \lim_{k\to\infty} \inf f(x^k)\}$.

Proposição 3.3.3. ([2], proposição 5, pg. 144) *Seja $f : \mathbb{R}^n \to \mathbb{R}$ convexa e continuamente diferenciável. A sequência $\{x^k\}$ gerada por (3.1) e (3.2) é convergente para um ponto $\overline{x} \in T$.*

Demonstração: Pela Proposição 3.3.2, para $z \in T$, temos que

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

Dessa forma, chamando $\varepsilon_k = \|x^{k+1} - x^k\|^2$, temos que $\varepsilon_k \geq 0$. Pela Proposição 3.3.1 (iii) temos que

$$\sum_{k=0}^{\infty} \|x^{k+1} - x^k\|^2 < \infty.$$

Pela Definição 1.11.3, a sequência $\{x^k\}$ é quase-Fejér convergente ao conjunto T. Agora, precisamos provar que existe um ponto de acumulação de $\{x^k\}$ em T e aplicar a segunda afirmação da Proposição 1.11.2.

Pela primeira afirmação da Proposição 1.11.2, temos que $\{x^k\}$ é limitada e portanto, possui subsequência convergente. Seja $\overline{x}$ um ponto de acumulação de $\{x^k\}$, isto é

$$\lim_{j\to\infty} x^{k_j} = \overline{x}.$$

Como f é continua, temos que

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

Dessa forma, $f(\overline{x})$ é valor de aderência de $\{f(x^k)\}$. Pela Proposição 3.3.1 (ii), $\{f(x^k)\}$ é convergente, logo possui ponto um único valor de aderência. Então

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

Assim, temos que

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

implicando $\overline{x} \in T$. $\square$

3.4 Segundo teorema principal

Mostramos anteriormente que a sequência $\{x^k\}$ é convergente ao conjunto T. Com o auxilio deste resultado, provamos na proposição seguinte a convergência completa da sequência $\{x^k\}$, gerada pelo método do gradiente utilizando os Algoritmos A ou B, ao conjunto de minimizadores de f, denominado $\widetilde{S}$.

Teorema 3.4.1. ([2], teorema 3, pg. 144) *Seja $f : \mathbb{R}^n \to \mathbb{R}$ convexa e continuamente diferenciável. Então, a sequência $\{x^k\}$ gerada por (3.1) e (3.2) usando os Algoritmos A ou B, converge para um minimizador de f.*

Demonstração: Pela Proposição 3.3.3 temos que

$$\lim_{k\to\infty} x^k = \overline{x} \in T,$$

logo é suficiente provar que $\overline{x} \in \widetilde{S}$ que é o conjunto de minimizadores de f.

Para o Algoritmo A, por (3.4) temos que

$$\|x^{k+1} - x^k\|^2 = \lambda_k^2 \|\nabla f(x^k)\|^2 \geq \delta_1^2 \|\nabla f(x^k)\|^2.$$

Então, pela Proposição 3.3.1 (iii) e continuidade do $\nabla f(x)$, segue que $\nabla f(\overline{x}) = 0$. Como f é convexa, temos que $\overline{x}$ é um minimizador.

Para o Algoritmo B, suponhamos que $\overline{x} \notin \widetilde{S}$. Pela convexidadde de f, tomando $\overline{x} \in G$ temos que $\|\nabla f(\overline{x})\| > 0$. Pela Proposição 3.2.1, $u(x^*) > 0$ e $u(x^k)$ converege para $u(\overline{x})$. Dessa forma, existe k_0 tal que, para todo $k \geq k_0$

$$u(x^k) \geq \frac{1}{2} u(\overline{x}) \text{ e } \|\nabla f(x^k)\|^2 \geq \frac{1}{2}\|\nabla f(\overline{x})\|^2. \tag{3.25}$$

Seja

$$\sigma = (\min\{\delta_1, \frac{u(x^*)}{4}\})^2 \frac{\|\nabla f(x^*)\|^2}{2}.$$

Então, para algum $k \geq k_0$, utilizando (3.2) e a Proposição 3.2.2, teremos

$$\begin{aligned} \|x^{k+1} - x^k\|^2 &= \lambda_k^2 \|\nabla f(x^k)\|^2 \\ &\geq (\min\{\delta_1, \frac{u(x^k)}{2}\})^2 \|\nabla f(x^k)\|^2. \end{aligned}$$

De (3.25) segue que

$$\|x^{k+1} - x^k\|^2 \geq (\min\{\delta_1, \frac{u(x^*)}{4}\})^2 \frac{\|\nabla f(x^*)\|^2}{2} = \sigma > 0. \tag{3.26}$$

Como (3.26) contradiz a Proposição 3.3.1 (iii) provamos que $\overline{s} \in \widetilde{S}$. $\square$

Considerações finais

Os resultados clássicos que garantem a existência de solução local ou global para o problema

$$\min f(x), \text{ sujeito a } x \in D$$

com $f : D \to \mathbb{R}$ e $D \subseteq \mathbb{R}^n$, utilizam hipóteses relevantes de compacidade de conjuntos e/ou convexidade de funções. Resultados estes listados neste trabalho na forma dos Teorema 1.4.1, Corolário 1.4.1, Corolário 1.4.2 e Teorema 1.6.1.

Nos métodos de descida é necessário fazer uma escolha específica da direção de descida e da regra que define o comprimento de passo nessa direção, para obtermos um decrescimento dos termos da sequência $\{x^k\}$ para curvas cada vez menores. Pelo Lema 1.7.1 é observado que a escolha mais natural para a direção de descida é o anti-gradiente, isto é, $d^k = -\nabla f(x^k)$ considerando $\nabla f(x^k) \neq 0$. No entanto, esta escolha é longe de ser a mais eficiente na prática.

Com grande frequência, os métodos de descida cuja direção de descida é o anti-gradiente possuem convergência bastante lenta, principalmente quando os conjuntos de nível da função objetivo são alongados, a menos que f tenha conjuntos de níveis quase esféricos. O problema é que nos casos assim, a direção do anti-gradiente é quase ortogonal à direção $x^k - \overline{x}$ onde $\overline{x}$ é a solução do problema. Por isso, a sequência $\{x^k\}$ gerada pelo método do gradiente se aproxima da solução seguindo uma trajetória "zig-zag".

Na minimização unidimensional a direção de busca linear é exatamente ortogonal à direção do gradiente no iterando seguinte e dessa forma, a resolução de um problema unidimensional, em cada iteração de um método mesmo que seja inexata, é relativamente cara. Por isso na prática, outras regras de busca linear são mais úteis e mais usadas.

Iusem e Svaiter apresentam em [5] uma variante muito simples do método de máxima descida para que se tenha resultados de convergência mesmo com o enfraquecimento da hipótese sobre f, a saber pseudo-convexidade, e sem a hipótese de limitação do conjunto de nível.

Mesmo sabendo que a regularização proximal está sujeita às mesmas limitações e, a adição do termo de regularização quadrático para a busca linear represente um encargo computacional adicional, o procedimento numérico utilizado nessa busca será mais eficiente.

No processo desenvolvido, gradientes consecutivos formam um ângulo agudo, que em princípio parece melhor, e há uma maior liberdade na escolha do tamanho de passo λ_k que contola este ângulo. Dessa forma, obtém-se resultados teóricos melhores.

No trabalho de Burachik *et al.* (ver [2]), é feito uma reestruturação dos trabalhos apresentados por Polyak [3] e Dennis-Schnabel [4].

Em [3], Polyak demonstra que dada $f : \mathbb{R}^n \to \mathbb{R}$ convexa, com conjuntos de nível limitados e gradiente satisfazendo uma condição de Lipschitz com constante $L > 0$ conhecida, se na busca linear inexata $\lambda_k = \lambda \in (0, \frac{2}{L})$ então a sequência gerada pelo método de descida, converge fracamente para o conjunto de pontos estacionários de f.

Dennis-Schnabel propuseram em [4] uma estratégia de parada, onde sucessivos valores de λ são experimentados até que um é encontrado para satisfazer duas inequações pré-estabelecidas. Com a hipótese de limitação do conjunto de nível e da condição de Lipschitz para o gradiente de uma função convexa f, é provado que a sequência gerada pelo método de descida, com essa busca linear inexata, converge fracamente para o conjunto de pontos estacionários de f.

Burachik *et al.* comprovaram em [2] que, mesmo com o enfraquecimento das hipóteses anteriores, com relação à limitação do conjunto de nível para os dois casos e, a condição de Lipschitz para o segundo caso, os resultados de convergência são garantidos.

Vale ressaltar que a noção de quase-Fejér convergência foi primordial pois, além da garantia de limitação das sequências, tornou possível a generalização dos resultados apresentados por Polyak e Dennis-Schnabel, de convergência fraca à convergência completa.

Concluímos este trabalho destacando a importância do mesmo pois serviu de base fundamental para o desenvolvimento de muitos outros artigos. Entre eles, podemos citar o publicado por Kiwiel e Murty (ver [7]) em 1996, onde é provada a completa convergência do método de descida para a minimização de funções quase-convexas, generalizando o resultado de Burachik *et al.* Neste, também é retirada a hipótese de convexidade para f e não é necessário a condição de limitação dos conjuntos de nível. São adicionadas hipóteses bem mais gerais e a busca linear utilizada é a regra de Armijo.

O trabalho publicado por Wang e Xiu [17] em 2000 utiliza fortemente o teorema principal demonstrado por Kiwiel e Murty em [7], para mostrar entre outros resultados, a convergência do método gradiente projetado para funções pseudo-convexas e quase-convexas.

O artigo escrito por Quiroz *et al.* (ver [13]), publicado em 2008, reestrutura [7], com todas a hipóteses, teoremas e lemas no contexto de variedades riemanianas.

Referências Bibliográficas

[1] A. N. Iusem, B. F. Svaiter, M. Teboulle, Entropy-like proximal methods in convex programming *(to be published in Mathematics of Operations Research).*

[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] B. Polyak, Introduction to Optimization*, Optimization Software, New York, 1987.*

[4] Dennis, J. E. and Schnabel, R. B., Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, New Jersey, 1983.*

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

[6] Iusem, A. N., Proximal Point Methods in Optimization*, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1995.*

[7] 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.*

[8] Lima,E. L., Análise Real*, Instituto de Matemática Pura e Aplicada, vol. 1, Rio de Janeiro, 2006.*

[9] Lima,E. L., Curso de Análise*, Instituto de Matemática Pura e Aplicada, vol. 1, Rio de Janeiro, 2006.*

[10] Lima,E. L., Curso de Análise*, Instituto de Matemática Pura e Aplicada, vol. 2, Rio de Janeiro, 2005.*

[11] *Lima,E. L.,* Análise no Espaço $\mathbb{R}^n$*, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 2004.*

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

[13] *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.*

[14] *Rey Pastor, J., Pi Calleja, T. e Trejo, C. A.,* Análisis Matemático*, Editorial Kapelusz, vol. 2, Buenos Aires, 1957.*

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

[16] *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.*

[17] *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.*

[18] *Y. M. Ermol'ev,* On the method of generealized stochastic gradients and quasi-Fejér sequences*, Cybernetics, 1969, 5, p. 208-220.*