MÉTODO DO PONTO PROXIMAL COM DISTÂNCIA DE BREGMAN PARA PROBLEMAS DE MINIMIZAÇÃO QUASE CONVEXOS

Lara Thaise Bezerra Lima Souza

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Sistemas e Computação.

Orientadores: Paulo Roberto Oliveira
João Xavier da Cruz Neto

Rio de Janeiro
Agosto de 2017

MÉTODO DO PONTO PROXIMAL COM DISTÂNCIA DE BREGMAN PARA PROBLEMAS DE MINIMIZAÇÃO QUASE CONVEXOS

Lara Thaise Bezerra Lima Souza

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Examinada por:

Prof. Paulo Roberto Oliveira, D.Sc.

Prof. João Xavier da Cruz Neto, D.Sc.

Prof. Jurandir de Oliveira Lopes, D.Sc.

Prof. Márcia Helena Costa Fampa, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
AGOSTO DE 2017

Souza, Lara Thaise Bezerra Lima

Método do ponto proximal com distância de Bregman para problemas de minimização quase convexos/Lara Thaise Bezerra Lima Souza. – Rio de Janeiro: UFRJ/COPPE, 2017.

VII, 41 p. 29, 7cm.

Orientadores: Paulo Roberto Oliveira
João Xavier da Cruz Neto

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2017.

Referências Bibliográficas: p. 39 – 41.

1. Método do ponto proximal. 2. Distância de Bregman. 3. Minimização quase convexa. I. Oliveira, Paulo Roberto *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

Agradecimentos

Gratidão a Deus, pela saúde, por sempre guiar os meus passos e decisões, por me manter firme e plantar sonhos em meu coração nunca antes sonhamos.

Gratidão ao meu esposo João Carlos por acreditar em mim, mais do que eu mesma acreditei ser capaz. Por ser um amigo para todos as horas e por sempre me incentivar a querer mais e a nunca desistir.

Gratidão aos meus pais, Maria de Lourdes e Osvaldo, por todo carinho, apoio e por me mostrarem que a maior herança que eles podem me deixar é o estudo. À minha irmã, Laíse, por vibrar com cada conquista minha e incentivar a seguir em frente. À família do meu esposo, em particular à minha sogra por seu incentivo. Aos meus amigos, pela presença, mesmo à distância, ao logo desses anos.

Gratidão aos meus orientadores Dr. Paulo Roberto Oliveira e Dr. João Xavier da Cruz Neto pelos ensinamentos e compreensão.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

MÉTODO DO PONTO PROXIMAL COM DISTÂNCIA DE BREGMAN PARA PROBLEMAS DE MINIMIZAÇÃO QUASE CONVEXOS

Lara Thaise Bezerra Lima Souza

Agosto/2017

Orientadores: Paulo Roberto Oliveira
João Xavier da Cruz Neto

Programa: Engenharia de Sistemas e Computação

Neste trabalho estudamos a convergência do método do ponto proximal para resolver um problema de minimização restrito ao octante não negativo para funções quase convexas. Para isso, a distância Euclidiana no termo de regularização do método do ponto proximal clássico é substituído por uma aplicação com propriedades similares à uma distância mas sem necessariamente satisfazer todos os axiomas da distância. Tal aplicação é conhecida como distância de Bregman.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

PROXIMAL POINT METHOD WITH BREGMAN DISTANCE FOR QUASICONVEX MINIMIZATION PROBLEMS

Lara Thaise Bezerra Lima Souza

August/2017

Advisors: Paulo Roberto Oliveira
João Xavier da Cruz Neto

Department: Systems Engineering and Computer Science

In this work, we study the convergence of the proximal point method for solving a constrained minimization problem within the nonnegative orthant for quasiconvex functions. To this end, the Euclidian distance in the regularization term of the classic proximal point method is replaced by a map with nice similar properties such as a distance but not necessarily satisfying all the axioms of a distance. Such a map is the so called Bregman distance.

Sumário

Capítulo 1

Introdução

Vários problemas de otimização se reduzem a encontrar uma solução do seguinte problema

$$\nabla f(x) = 0 \tag{1.1}$$

ou de forma equivalente

$$\min_{x \in \mathbb{R}^n} f(x), \tag{1.2}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$ é uma função convexa. Assuma que f é limitada inferiormente e tome $g : \mathbb{R}^n \to \mathbb{R}$ uma função estritamente convexa e coerciva. O problema (1.1) pode não ter solução ou ter mais de uma solução, mas o problema regularizado

$$\min_{x \in \mathbb{R}^n} f(x) + \lambda g(x) \tag{1.3}$$

tem única solução para cada $\lambda > 0$. Com efeito, a aplicação $f(\cdot) + \lambda g(\cdot)$ é coerciva (usando o fato que f é limitada inferiormente) reduzindo o problema a um conjunto compacto e com isso garantindo a existência de soluções. A convexidade estrita da aplicação $f(\cdot)+\lambda g(\cdot)$ garante a unicidade da solução do problema (1.3) denotada por $x(\lambda)$. Sob algumas hipóteses razoáveis, incluindo a existência de solução de (1.2), pode-se provar que

$$\lim_{\lambda \to 0^+} x(\lambda)$$

existe e resolve (1.2). Conforme mencionado em Iusem [14], mesmo $f(\cdot)+\lambda g(\cdot)$ sendo estritamente convexa e coerciva para cada $\lambda > 0$, para λ suficientemente pequeno essa função pode ter numericamente um mal comportamento tanto quanto f. Em outras palavras sendo o sistema (1.1) mal condicionado, então o sistema

$$(\nabla f + \lambda \nabla g)(x) = 0$$

será mal condicionado quando $\lambda \to 0$, apesar do fato de que (1.3) tem única solução para todo $\lambda > 0$.

Afim de evitar essa dificuldade seria razoável desenvolver uma abordagem de regularização que não requer que o parâmetro de regularização λ se aproxime de 0. O método do ponto proximal pode ser usado para atingir esse objetivo.

O método do ponto proximal para resolver um problema de minimização foi introduzido na literatura de otimização por Martinet [21] e popularizado por Rockafellar [26] que por sua vez estuda versões exatas e inexatas do método para encontrar zeros de operadores monótonos maximal usando alguns conceitos estabelecidos por Moreau [22]. A ideia do método é, a partir de um ponto inicial dado, gerar uma sequência de pontos que são mínimos da função objetivo acrescida de uma regularização (quadrado da função distância), isto é, dado $x^0 \in \mathbb{R}^n$ definimos

$$x^{k+1} = \arg\min_{x \in \mathbb{R}^n} \{f(x) + \lambda_k ||x - x^k||^2\}, \tag{1.4}$$

onde $\{\lambda_k\}$ é uma sequência de parâmetros positivos. Quando a função em consideração é convexa o método está bem definido e pode determinar um ponto de mínimo de f sob hipóteses bastantes razoáveis.

Quando no problema (1.2) a minimização é restrita a um conjunto $C \subset \mathbb{R}^n$ fechado, ou seja,

$$\min_{x \in C} f(x),$$

o grande desafio do método do ponto proximal é manter as iteradas dentro do interior do conjunto C. Para resolver esse tipo de problema, introduz-se uma função penalização g que force as iteradas a permanecerem no interior de C de tal forma que exista uma solução $x(\lambda)$ do problema

$$\min_{x \in \mathbb{R}^n} f(x) + \lambda g(x)$$

e $\lim_{\lambda \to 0} x(\lambda)$ exista e resolva o problema a cima.

Diversos autores tem proposto variações desse método substituindo a função distância na regularização por aplicações que não satisfazem todos os axiomas da função distância, mas preservam algumas de suas boas propriedades tais como continuidade, coercividade, etc; veja [2, 8, 16, 29]. Essas variações tem diversas aplicações em diferentes áreas; veja por exemplo [1, 4, 5, 7, 9, 11, 12, 16, 17, 23, 24, 27]. Neste trabalho iremos considerar uma das mais populares regularizações do tipo "distância"conhecida como distância de Bregman (veja capítulo 2). O conceito de função de Bregman foi originalmente proposto por Bregman [6].

Os resultados deste trabalho foram originalmente obtidos em Rockafellar [26], Chen e Teboulle [9] e Souza et al.[28].

A organização deste trabalho é feita da seguinte forma. No capítulo 2 apresentamos os principais conceitos e resultados de análise em $\mathbb{R}^n$, análise convexa, otimização, distância de Bregman e Fejér convergência. No capítulo 3 definimos o método do ponto proximal para resolver um problema de minimização irrestrito onde a função objetivo é convexa e, mais geral, quase convexa. No capítulo 4 estendemos os resultados anteriormente demonstrados para resolver um problema de minimização restrito ao $\mathbb{R}^n_+$ usando distância de Bregman para os casos onde a função objetivo é convexa ou quase convexa. No último capítulo consideramos as possibilidades de trabalhos futuros.

Capítulo 2

Noções Preliminares

Neste capítulo, trataremos de conceitos básicos necessários para compreensão dos capítulos seguintes. Para isso, nas duas primeiras seções abordaremos alguns conceitos e resultados de análise no $\mathbb{R}^n$ e análise convexa. Nas duas seções finais apresentaremos alguns conceitos e resultados envolvendo Fejér convergência e distância de Bregman.

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

As definições e resultados a seguir podem ser encontrados em qualquer livro de análise no $\mathbb{R}^n$, como por exemplo Elon Lima [19].

Definição 2.1.1 *Um produto interno no espaço euclidiano $\mathbb{R}^n$ é uma regra que faz corresponder a cada par de vetores $x, y \in \mathbb{R}^n$ um número real, denotado por $\langle x, y \rangle$, de modo que, para quaisquer $x, y, z \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$, tenhamos:*

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

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

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

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

Os resultados acima informam que um produto interno é uma forma bilinear, simétrica, positiva definida; veja Elon Lima [20]. Um exemplo bastante importante é o produto interno canônico no espaço euclidiano $\mathbb{R}^n$, o qual é dado por

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

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

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

Teorema 2.1.1 (Desigualdade de Cauchy-Schwarz) *Para quaisquer $x, y \in \mathbb{R}^n$, tem-se $|\langle x, y\rangle| \leq ||x||||y||$. Vale a igualdade se, e somente se, um dos vetores x, y é um múltiplo escalar do outro.*

Demonstração: Veja página 5 de [19]. ■

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

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

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

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

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

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

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

Uma norma no $\mathbb{R}^n$ permite que sejam definidos alguns conceitos. Dentre eles, uma bola de centro $a \in \mathbb{R}^n$ e raio $r > 0$ é o conjunto

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

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

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

Definição 2.1.2 *Sequência no $\mathbb{R}^n$ é uma aplicação $x : \mathbb{N} \to \mathbb{R}^n$. O valor que essa aplicação assume no número k é indicado por x^k e denominado k-ésimo termo da sequência. Usaremos a notação $\{x^k\}$ para indicar a sequência cujo k-ésimo termo é $x^k \in \mathbb{R}^n$.*

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

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

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

Definição 2.1.3 *Uma aplicação $f : X \to \mathbb{R}$ definida no conjunto $X \subseteq \mathbb{R}^n$ é uma função contínua no ponto $a \in X$ quando,*

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

As funções contínuas tem a seguinte propriedade:

Proposição 2.1.1 *Uma aplicação $f : X \to \mathbb{R}$ definida no subconjunto $X \subseteq \mathbb{R}^n$ é contínua no ponto $a \in X$ se, e somente se, para toda sequência de pontos $\{x^k\} \subset X$, com $\lim_{k\to\infty} x^k = a$, tem-se $\lim_{k\to\infty} f(x^k) = f(a)$.*

Demonstração: Veja página 26 de [19]. ∎

Definição 2.1.4 *Dizemos que uma função $f : X \to \mathbb{R}$ definida no subconjunto $X \subseteq \mathbb{R}^n$ é semicontínua inferiormente (resp. superiormente) se*

$$\liminf_{k\to\infty} f(x^k) \geq f(a) \quad (\textit{resp.} \; \limsup_{k\to\infty} f(x^k) \leq f(a)),$$

para toda sequência $x^k \to a$ quando $k \to \infty$.

2.2 Conceitos e resultados de análise convexa

Por se tratar de um assunto extenso, trataremos apenas dos conceitos e resultados que serão utilizados no decorrer deste trabalho. Definiremos alguns conceitos introdutórios de análise convexa tais como funções convexa e quase convexas, minimizadores locais e globais, dentre outros conceitos e resultados básicos. As definições e resultados apresentados a seguir podem ser encontrados em qualquer livro de análise convexa, como por exemplo em Solodov e Izmailov [15] ou Rockafellar e Wets [25].

Consideremos um conjunto $X \subset \mathbb{R}^n$ e uma função $f : X \to \mathbb{R}$. O problema de minimizar f em X é exposto da seguinte maneira:

$$\min_{x\in X} f(x), \tag{2.1}$$

onde X é chamado conjunto viável e f função objetivo, ou função custo.

Definição 2.2.1 *Seja a função $f : X \to \mathbb{R}$. Dizemos que $\overline{x} \in X$ é minimizador local do problema de (2.1), se existir $\delta > 0$, tal que*

$$f(\overline{x}) \leq f(x), \qquad \forall x \in B(\overline{x}, \delta) \cap X. \tag{2.2}$$

E dizemos que $\overline{x}$ é minimizador global do problema de (2.1) se

$$f(\overline{x}) \leq f(x), \quad \forall x \in X. \tag{2.3}$$

Se as desigualdades (2.2) e (2.3) forem estritas, então $\overline{x}$ será chamado, respectivamente, de minimizador estrito local e minimizador estrito global. As definições envolvendo máximo (local, global e estrito) são similares.

Definição 2.2.2 *Dizemos que $\overline{v} \in [-\infty, +\infty)$ definido por*

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

é o valor ótimo do problema (2.1). O valor ótimo também é denotado por f^*.

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

Existem alguns critérios que garantem a existência de solução global para o problema (2.1). Um critério bastante conhecido é o seguinte teorema.

Teorema 2.2.1 (Teorema de Weierstrass) *Sejam $X \subset \mathbb{R}^n$ um conjunto compacto e não vazio e $f : X \to \mathbb{R}$ uma função contínua. Então, o problema (2.1) tem solução global.*

Demonstração: Veja página 7 de [15]. ■

Definição 2.2.3 *Dizemos que uma função $f : D \to \mathbb{R}$ é coerciva no conjunto D quando para toda sequência $\{x^k\} \subset D$ tal que $||x^k|| \to +\infty$ ou $x^k \to x \in \overline{D} \setminus D$ quando $k \to +\infty$ tem-se*

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

onde $\overline{D}$, chamado de fecho de D, denota o conjunto dos pontos de acumulação de D.

Proposição 2.2.1 *Sejam $D \subset \mathbb{R}^n$ e $f : D \to \mathbb{R}$ uma função semicontínua inferiormente e coerciva em D. Então f tem minimizador global em D.*

Demonstração: Veja página 14 de [15]. ■

Quando, no problema (2.1), consideramos o conjunto X como sendo o próprio $\mathbb{R}^n$, então dizemos que temos um problema irrestrito e apresentamos esse problema da seguinte forma:

$$\min_{x\in\mathbb{R}^n} f(x). \tag{2.4}$$

A seguir apresentamos o conceito de convexidade para conjuntos e funções.

Um conjunto convexo é caracterizado por conter todos os segmento de retas cujos os extremos pertencem ao conjunto, conforme definição abaixo.

Definição 2.2.4 *Um conjunto $X \subset \mathbb{R}^n$ é convexo se, para quaisquer $x, y \in X$ e $\lambda \in [0,1]$, tem-se que*

$$\lambda x + (1-\lambda)y \in X.$$

Exemplo 2.2.1 *Um hiperplano do $\mathbb{R}^n$ é um conjunto da forma*

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

onde $a \in \mathbb{R}^n$ e $c \in \mathbb{R}$. Mostraremos que tal conjunto é convexo em $\mathbb{R}^n$. Dados $x, y \in H(a,c)$ e $\lambda \in [0,1]$, temos

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

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

Definição 2.2.5 *Se $X \subset \mathbb{R}^n$ é um conjunto convexo, dizemos que a função $f : X \to \mathbb{R}$ é convexa em X quando, para quaisquer $x, y \in X$ e $\lambda \in [0,1]$, tem-se*

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

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

O resultado a seguir é uma consequência imediata da definição acima.

Proposição 2.2.2 *Sejam $f, g : \mathbb{R}^n \to \mathbb{R}$ funções convexa e estritamente convexa, respectivamente. Então $f + g$ é uma função estritamente convexa.*

Obviamente, toda função estritamente convexa é também uma função convexa. A seguir, veremos alguns resultados úteis para se verificar se uma função é convexa (ou estritamente convexa).

Teorema 2.2.2 (Caracterização de Funções Convexas) *Sejam $X \subset \mathbb{R}^n$ um conjunto convexo e aberto e $f : X \to \mathbb{R}$ uma função diferenciável em X. Então as afirmações abaixo são equivalentes:*

1) *A função f é convexa em X.*

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

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

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

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

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

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

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

Demonstração: Veja página 146 de [15]. ■

O teorema acima ainda se verifica para o caso de funções estritamente convexas apenas substituindo as desigualdades por desigualdades estritas; veja página 150 de [15].

Exemplo 2.2.2 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ dada por $f(x) = \frac{1}{2}||x||^2$. Mostraremos que f é (estritamente) convexa. Com efeito, sabemos que f é diferenciável e que $\nabla f(x) = x$. Assim, $\nabla^2 f(x) = I$, para todo $x \in \mathbb{R}^n$, onde I é a matriz identidade e então $\nabla^2 f(x)$ é positiva definida. Portanto, pelo teorema anterior, temos que f é (estritamente) convexa.*

Definição 2.2.6 *O epígrafo de uma função $f : X \to \mathbb{R}$ é o conjunto*

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

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

Proposição 2.2.3 *Seja $X \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : X \to \mathbb{R}$ é convexa em X se, e somente se, o epígrafo de f é um conjunto convexo em $\mathbb{R}^n \times \mathbb{R}$.*

Demonstração: Veja página 67 de [15]. ■

Dizemos que um problema de minimização é convexa quando o conjunto de restrições $X \subset \mathbb{R}^n$ é um conjunto convexo e a função objetivo $f : X \to \mathbb{R}$ é uma função convexa em X. A importância da convexidade já pode ser vista no seguinte resultado.

Teorema 2.2.3 (Teorema de minimização convexa) *Sejam $X \subset \mathbb{R}^n$ um conjunto convexo e $f : X \to \mathbb{R}$ é uma função convexa em X. Então, todo minimizador local do problema (2.1) é global. Além disso, o conjunto de minimizadores é convexo. Se f é estritamente convexa, não pode haver mais de um minimizador.*

Demonstração: Veja página 69 de [15]. ■

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

Teorema 2.2.4 *Sejam $X \subset \mathbb{R}^n$ um conjunto convexo, aberto e $f : X \longrightarrow \mathbb{R}$ uma função convexa. Então f é localmente Lipschitz e, em particular, f é contínua.*

Demonstração: Veja página 136 de [15]. ■

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

Exemplo 2.2.3 *Sejam $A = \{x \in \mathbb{R} \ ; \ x \geq -1\}$ e $f : A \to \mathbb{R}$,*

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

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

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

Definição 2.2.7 *O conjunto de nível de uma função $f : X \to \mathbb{R}$ associado a um ponto $c \in \mathbb{R}$ é dado por*

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

Proposição 2.2.4 *Seja f uma função convexa definida no conjunto convexo $X \subset \mathbb{R}^n$. Então $L_{f,X}(c)$ é convexo, para todo $c \in \mathbb{R}$.*

Demonstração: Veja página 133 de [15]. ■

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

Exemplo 2.2.4 *Seja $f : \mathbb{R} \to \mathbb{R}$, com $f(x) = x^3$. A função f tem todos os seus conjuntos de níveis convexos, mas f não é convexa (basta tomar $x = -1$, $y = 0$ e $\lambda = 1/2$ e verificar que $f(\lambda x + (1-\lambda)y) > \lambda f(x) + (1-\lambda)f(y)$.*

Definição 2.2.8 *Seja $f : X \longrightarrow \mathbb{R}$, onde $X \subset \mathbb{R}^n$ é não vazio e convexo. A função f é dita quase convexa se, para todo $x, y \in X$*

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

Quando a desigualdade acima é estrita dizemos que f é estritamente quase convexa.

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

Teorema 2.2.5 *Seja $f : X \longrightarrow \mathbb{R}$, onde $X \subset \mathbb{R}^n$ é não vazio e convexo. A função f é quase-convexa se, e somente se,*

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

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

Demonstração: Veja página 134 de [15]. ∎

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

Exemplo 2.2.5 *Seja*

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

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

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

Definição 2.2.9 *Seja $f : X \to \mathbb{R}$ uma função definida em um subconjunto $X \subseteq \mathbb{R}^n$. Dizemos que $s \in \mathbb{R}^n$ é um subgradiente de f no ponto $x \in X$ se*

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

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

Na definição acima, o subdiferencial de f em um ponto x também pode ser escrito da seguinte forma

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

onde

$$f'(x,d) = \lim_{t\to 0^+} \frac{f(x+td)-f(x)}{t},$$

veja [15, página 164]. O subdiferencial de uma função convexa está bem definido conforme teorema a seguir.

Teorema 2.2.6 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Então, para todo $x \in \mathbb{R}^n$, o conjunto $\partial f(x)$ é convexo, compacto e não vazio.*

Demonstração: Veja página 164 de [15]. ■

O resultado abaixo mostra que o subdiferencial de uma função generaliza o conceito de diferenciabilidade.

Proposição 2.2.5 *Uma função convexa $f : \mathbb{R}^n \to \mathbb{R}$ é diferenciável no ponto $x \in \mathbb{R}^n$ se, e somente se, $\partial f(x) = \{\nabla f(x)\}$.*

Demonstração: Veja página 167 de [15]. ■

Proposição 2.2.6 *Sejam $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, p$, funções convexas. Então*

$$\partial\left(\sum_{i=1}^{p} f_i(x)\right) = \sum_{i=1}^{p} \partial f_i(x).$$

Demonstração: Veja página 172 de [15]. ■

A importância do conceito de subdiferencial em otimização pode ser visto do resultado a seguir.

Teorema 2.2.7 *Sejam $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa e $D \subset \mathbb{R}^n$ um conjunto convexo. Então, $\overline{x} \in \mathbb{R}^n$ é um minimizador de f em D se, e somente se, existe $y \in \partial f(\overline{x})$ tal que*

$$\langle y, x - \overline{x}\rangle \geq 0, \quad \forall x \in D.$$

Em particular, $\overline{x} \in \mathbb{R}^n$ é um minimizador de f no $\mathbb{R}^n$ se, e somente se

$$0 \in \partial f(\overline{x}).$$

Demonstração: Veja página 168 de [15]. ■

Exemplo 2.2.6 *Seja $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = |x|$. O ponto $\overline{x} = 0$ é único minimizador (global) irrestrito de f. Pela definição de subdiferencial, obtemos*

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

Em particular, temos que $0 \in \partial f(\overline{x})$ e $0 \notin \partial f(x)$ para todo $x \neq \overline{x}$.

O exemplo abaixo mostra que, para o caso que f não é uma função convexa, a condição $0 \in \partial f(x)$ não é suficiente para que o ponto x seja mínimo irrestrito de f.

Exemplo 2.2.7 *Seja $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = x^3$. Pela proposição 2.2.5, temos que $\partial f(0) = \{0\}$. Por outro lado, sabemos que $x = 0$ não é ponto de mínimo irrestrito de f.*

Definição 2.2.10 *Dada uma função $f : \mathbb{R}^n \to \mathbb{R}$ dizemos que um ponto $x \in \mathbb{R}^n$ é ponto crítico de f quando $0 \in \partial f(x)$.*

Definição 2.2.11 *O subdiferencial de Fréchet de f em x, denotado por $\partial_F f(x)$, é definido da seguinte forma*

$$\partial_F f(x) = \{v \in \mathbb{R}^n \ : \ \liminf_{\substack{y \to x \\ y \neq x}} \frac{f(y) - f(x) - \langle v, y - x \rangle}{||x - y||} \geq 0\}.$$

Se f é uma função convexa e semicontínua inferiormente em x, então $\partial f(x) = \partial_F f(x) \neq \emptyset$.

Definição 2.2.12 *O subdiferencial de Mordukhovich (ou subdiferencial limite) de f em x, denotado por $\partial_L f(x)$, é definido da seguinte forma*

$$\partial_L f(x) = \{v \in \mathbb{R}^n \ : \ \exists x^k \to x, \ f(x^k) \to f(x), \ v^k \in \partial_F f(x^k) \to v\}.$$

Proposição 2.2.7 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ e $x \in \mathbb{R}^n$. Então os subdiferenciais $\partial_L f(x)$ e $\partial_F f(x)$ são fechados, com $\partial_F f(x)$ convexo e*

$$\partial_F f(x) \subset \partial_L f(x).$$

Demonstração: Veja [25, Teorema 8.6]. ∎

Proposição 2.2.8 *Se uma função $f : \mathbb{R}^n \to \mathbb{R}$ tem um ponto de mínimo local em x, então $0 \in \partial_F f(x)$ e $0 \in \partial_L f(x)$. Se f é convexa, essas condições são necessárias e suficientes para um ponto ser mínimo global. Além disso, se $f = f_1 + f_2$ com f_2 continuamente diferenciável, a condição*

$$0 \in \partial_F f(x) \quad (\text{resp. } 0 \in \partial_L f(x))$$

assume a forma

$$-\nabla f_2(x) \in \partial_F f_1(x) \quad (\text{resp. } -\nabla f_2(x) \in \partial_L f_1(x)).$$

Demonstração: Veja [25, Teorema 10.1]. ∎

Proposição 2.2.9 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função semicontínua inferiormente e quase convexa. Se $v \in \partial_F f(x)$ e $f(y) \leq f(x)$, então*

$$\langle v, y - x \rangle \leq 0.$$

Demonstração: Veja [3, Theorem 3.5.4]. ∎

Definição 2.2.13 *Dizemos que $f : \mathbb{R}^n \to \mathbb{R}$ é uma função Lipschitz contínua com constante L, se existe $L > 0$ tal que para todo $x, y \in \mathbb{R}^n$, vale*

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

Quando a propriedade acima é válida em uma vizinhança U de um ponto $\hat{x} \in \mathbb{R}^n$ dizemos que f é localmente Lipschitz em $\hat{x}$. Quando f é localmente Lipschitz em todos os pontos dizemos apenas que f é localmente Lipschitz.

Definição 2.2.14 *Dizemos que uma aplicação ponto-conjunto $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ é localmente limitada em um ponto $x \in \mathbb{R}^n$, se para alguma vizinhança $V \in \mathcal{N}(x)$ o conjunto $S(V) \subset \mathbb{R}^m$ é limitado, onde $\mathcal{N}(x)$ é o conjunto de todas as vizinhanças de x. Uma aplicação é dita localmente limitada (em $\mathbb{R}^n$) se essa condição for verificada para todo $x \in \mathbb{R}^n$.*

Proposição 2.2.10 *Suponha que $f : \mathbb{R}^n \to \mathbb{R}$ localmente semicontínua inferior em x. Então as seguintes afirmações são equivalentes:*

1. *f é localmente Lipschitz em x;*

2. *a aplicação $\partial_F f : x \mapsto \partial_F f(x)$ é localmente limitada em x;*

3. *a aplicação $\partial_L f : x \mapsto \partial_L f(x)$ é localmente limitada em x;*

Além disso, quando uma dessas condições se verificam, temos que $\partial_L f(x)$ é não vazio e compacto.

Demonstração: Veja [25, Teorema 9.13]. ∎

2.3 Distância de Bregman

Seja S um subconjunto aberto e convexo do $\mathbb{R}^n$ e $\bar{S}$ seu fecho. Considere h uma função real convexa definida em S e seja $D_h : \bar{S} \times S \to \mathbb{R}$

$$D_h(x, y) = h(x) - h(y) - \langle \nabla h(y), (x - y) \rangle. \tag{2.5}$$

h é dita ser uma função de Bregman (e D_h a distância de Bregman induzida por h) se as seguintes condições são verificadas:

B1) h é continuamente diferenciável em S.

B2) h é estritamente convexa e contínua em $\bar{S}$.

B3) Para todo $\delta \in \mathbb{R}$ os conjuntos de níveis

$$\Gamma_1(y,\delta) = \{x \in \bar{S} : D_h(x,y) \leq \delta\} \quad \text{e} \quad \Gamma_2(x,\delta) = \{y \in S : D_h(x,y) \leq \delta\}$$

são limitados para todo $y \in S$ e todo $x \in \bar{S}$ respectivamente.

B4) Se $\{y^k\} \subset S$ converge para y^* então $D_h(y^*, y^k)$ converge para 0.

B5) Se $\{z^k\} \subset \bar{S}$ e $\{y^k\} \subset S$ são sequências tal que $\{z^k\}$ é limitada,

$$\lim_{k\to\infty} y^k = y^* \quad \text{e} \quad \lim_{k\to\infty} D_h(z^k, y^k) = 0,$$

então

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

S é chamado a zona de h. Note que $D_h(x,y) \geq 0$ para todo $x \in \bar{S}$, $y \in S$ e $D_h(x,y) = 0$ se, e somente se $x = y$. Como uma consequência de B1, B2 e B3 observamos que B4 e B5 se verificam automaticamente quando x^k, y^* estão em S. Assim precisam ser verificadas somente em pontos da fronteira, $\operatorname{fr} S$, de S. Em [13] está provado que quando $S = \mathbb{R}^n$ uma condição para h, uma função convexa e diferenciável, ser uma função de Bregman é tal que

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

Antes de apresentar exemplos de funções de Bregman, vamos introduzir duas subclasses para ser usada na continuação.

Uma função de Bregman h é dita ser coerciva limitada se:

B6) Se $\{y^k\} \subset S$ é tal que $\lim_{k\to\infty} y^k = y \in \operatorname{fr} S$, então

$$\lim_{k\to\infty} \nabla h(y^k)^t (x - y^k) = -\infty,$$

para todo $x \in S$.

Uma função de Bregman h é dita ser zona coerciva se:

B7) Para cada $y \in \mathbb{R}^n$ existe $x \in S$ tal que $\nabla h(x) = y$.

B6 seria o conceito chave em conexão com o método do ponto proximal pela seguinte razão. É claro de B1 até B5 que se h é uma função de Bregman com zona

S e S' é um conjunto aberto de S então h também é uma função de Bregman com zona S', isto é, não podemos recuperar S de h. Por outro lado, queremos usar D_h para fins de penalização, para minimizar funções em um conjunto fechado convexo C. A informação sobre o conjunto C no algoritmo considerado na continuação será encapsulada em D_h, de modo que C terá que ser recuperado a partir de h. B6 se encaixa nessa situação, porque a divergência de ∇h na $\operatorname{fr} S$ torna S univocamente determinado por h. Em tudo nosso algoritmo tornará C igual no fecho $\bar{S}$ da zona S na função de Bregman h.

Definição 2.3.1 *Dizemos que uma função de Bregman h é separável quando h pode ser escrita da forma*

$$h(x) = \sum_{i=1}^{n} h_i(x_i),$$

com h_i funções escalares. Nesse caso, dizemos que D_h associada a h é separável.

Observação 2.3.1 *Quando h é uma função de Bregman com zona coerciva e separável em $\mathbb{R}^n_{++}$, temos que*

$$h_i'(0, +\infty) = (-\infty, +\infty),$$

e consequentemente por B6, temos que

$$\lim_{t \to 0^+} h_i'(t) = -\infty, \quad \forall i = 1, \ldots, n.$$

No decorrer deste trabalho iremos denotar por $\nabla_1 D(x, y)$ e $\nabla_2 D(x, y)$ como o gradiente da função $D(x, y)$ com respeito à primeira e segunda variável, respectivamente.

Proposição 2.3.1 *Se h é uma função de Bregman com zona S então*

i) $D_h(x, y) - D_h(x, z) - D_h(z, y) = \langle \nabla h(y) - \nabla h(z), z - x \rangle$ *para todo* $x \in \bar{S}$*;* $y, z \in S$,

ii) $\nabla_1 D_h(x, y) = \nabla h(x) - \nabla h(y)$ *para todo* $x, y \in S$,

iii) $D_h(\cdot, y)$ *é estritamente convexa para todo* $y \in S$.

Demonstração: Segue diretamente da definição (2.5) e das propriedades de função de Bregman. ∎

2.4 Fejér convergência

Definição 2.4.1 *Uma sequência* $\{y^k\} \subset \mathbb{R}^n$ *é dita Fejér convergente ao conjunto não vazio* $C \subset \mathbb{R}^n$*, se*

$$||y^{k+1} - y|| \leq ||y^k - y||, \quad \forall k \geq 0, \; \forall y \in C. \tag{2.6}$$

Lema 2.4.1 *Se* $\{y^k\}$ *é Fejér convergente ao conjunto não vazio* $C \subset \mathbb{R}^n$*, então* $\{y^k\}$ *é uma sequência limitada. Se, além disso, um ponto de acumulação* $\bar{y}$ *da sequência* $\{y^k\}$ *pertence à* C*, então* $\lim\limits_{k \to +\infty} y^k = \bar{y}$*.*

Demonstração: Dado $y \in C$, a desigualdade (2.6) implica que

$$||y^k - y|| \leq ||y^{k-1} - y|| \leq \ldots \leq ||y^0 - y||, \quad \forall k \in \mathbb{N}.$$

Desse modo a sequência $\{y^k\}$ está contida em uma bola de centro y e raio $||y^0 - y||$, ou seja, $\{y^k\}$ é limitada. Agora sejam $\bar{y} \in U$ um ponto de acumulação de $\{y^k\}$ e $\{y^{k_j}\}$ uma subsequência de $\{y^k\}$ tal que $\lim\limits_{j \to +\infty} y^{k_j} = \bar{y}$. Como $\bar{y} \in U$, segue de (2.6) que a sequência de números reais positivos $\{||y^k - \bar{y}||\}$ é monótona (não crescente) e possui uma subsequência convergente, a saber $\{||x^{k_j} - \bar{x}||\}$ converge para zero. Então a sequência $\{||y^k - \bar{y}||\}$ converge para zero, que implica em $\lim\limits_{k \to +\infty} y^k = \bar{y}$. ■

Capítulo 3

Método do ponto proximal

Neste capítulo, apresentamos o método do ponto proximal para resolver um problema de minimização irrestrito

$$\min_{x \in \mathbb{R}^n} f(x),$$

onde $f : \mathbb{R}^n \to \mathbb{R}$ é uma função semicontínua inferiormente não necessariamente diferenciável. O método do ponto proximal gera uma sequência $\{x^k\} \subset \mathbb{R}^n$ definida da seguinte forma.

Algoritmo 3.1 - Método do Ponto Proximal
Passo 1: Tome um ponto inicial $x^0 \in \mathbb{R}^n$ *e* $\{\lambda_k\}$ *uma sequência auxiliar de parâmetros tal que* $0 < \lambda_k \leq \tilde{\lambda}$.
Passo 2: Calcule

$$x^{k+1} \in \arg\min_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{\lambda_k}{2} ||x - x^k||^2 \right\}. \tag{3.1}$$

Passo 3: Se $x^{k+1} = x^k$*, pare. Caso contrário, faça* $k = k + 1$ *e retorne ao Passo 2.*

A seguir, estudaremos a convergência do algoritmo 3.1 para os casos em que f é convexa (nesse caso a inclusão em (3.1) vira uma igualdade) e, mais geral, o caso em que f é quase convexa. O primeiro caso será estudado apenas por questão de completude do trabalho, mesmo sabendo que toda função convexa é quase convexa. Assim, iremos supor que S, o conjunto dos minimizadores de f, é não vazio. A partir de agora, $\{x^k\}$ denota a sequência gerada pelo Algoritmo 3.1.

Para o caso em que a função objetivo é convexa mostraremos que o Algoritmo 3.1 converge para um minimizador de f independente da escolha do ponto inicial x^0. Essa propriedade é conhecida como convergência global. Para isso, a convexidade da função desempenha um papel muito importante conforme vemos no exemplo a seguir.

Exemplo 3.0.1 *Seja $f : \mathbb{R} \to \mathbb{R}$ dada por*

$$f(x) = \begin{cases} (1-x)^{-1}, & se\ x \in (-\infty, 0], \\ 1, & se\ x \in (0, 2), \\ (x-3)^2, & se\ x \in [2, +\infty). \end{cases}$$

Temos que $f(x) \geq 0$ para todo $x \in \mathbb{R}$ e $x^* = 3$ é único minimizador de f em $\mathbb{R}$. Vamos analizar o Algoritmo 3.1 para o caso em que $\lambda_k = 2$ para todo $k \in \mathbb{N}$. Segue de Lagenberg e Tichatschke [18] que para qualquer $x^0 > 2$ a sequência $\{x^k\} \subset [2, +\infty)$ e converge para o único minimizador de f em $[2, +\infty)$. Se $x^0 < 0$, nesse caso a sequência $\{x^k\}$ não converge mesmo f sendo convexa em $(-\infty, 0]$ (f não tem minimizador em $(-\infty, 0]$). Note que f não é quase convexa, uma vez que $N_f(\alpha)$ não é convexo para $\alpha \in (0, 1)$.

O exemplo anterior é uma motivação para considerarmos o método do ponto proximal para funções convexas. Primeiramente, vamos analisar a convergência do método para o caso clássico em que f é convexa.

3.1 Caso convexo

Para esta seção iremos assumir adicionalmente que $f : \mathbb{R}^n \to \mathbb{R}$ é uma função convexa.

O resultado a seguir mostra a boa definição do Algoritmo 3.1. Esse resultado foi demonstrado de uma forma mais geral, para operadores monótonos maximais, por Rockafellar [26].

Teorema 3.1.1 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa. A sequência $\{x^k\} \subset \mathbb{R}^n$ gerada por* (3.1) *está bem definida e caracterizada pela relação*

$$\lambda_k(x^k - x^{k+1}) \in \partial f(x^{k+1}), \quad \forall k \in \mathbb{N}. \tag{3.2}$$

Demonstração: A prova será feita por indução sobre k. Para $k = 0$, o ponto x^0 é escolhido pela inicialização do algoritmo no passo 1. Agora, suponha que o método já tenha atingido a iterada x^k provaremos que existe a iterada $x^{k+1} \in \mathbb{R}^n$. Para isso, defina

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

Como S, o conjunto dos minimizadores de f, é não-vazio, temos que existe $x^* \in S$ tal que $f(x^*) \leq f(x)$, para todo $x \in \mathbb{R}^n$. Assim,

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

Logo, quando $||x|| \to +\infty$, temos que $f_k(x) \to +\infty$, ou seja, f_k é coerciva. Sendo f semicontínua inferiormente, temos que f_k também é, e pela Proposição 2.2.1 temos que f_k possui minimizador. Afirmamos que tal minimizador é único. Com efeito, sendo a função $h(x) = \lambda_k||x-x^k||^2$ estritamente convexa (conforme Exemplo 2.2.2), temos Proposição 2.2.2 que f_k é estritamente convexa e, portanto, pelo Teorema 2.2.3 f_k possui um único minimizador $\hat{x}$. Assim, a sequência $\{x^k\}$ dada por (3.1) está bem definida tomando $x^{k+1} = \hat{x}$.

Agora, seja x^{k+1} o único minimizador de f_k. Pelo Teorema 2.2.7, temos que

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

Por outro lado, combinando as Proposições 2.2.5 e 2.2.6, temos que

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

o que equivale dizer que

$$\lambda_k(x^k - x^{k+1}) \in \partial f(x^{k+1}).$$

■

Vejamos a seguir um resultado técnico usado na análise de convergência do método.

Proposição 3.1.1 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Se a sequência $\{x^k\}$ é gerada pelo algoritmo 3.1, então vale a desigualdade*

$$||x - x^{k+1}||^2 \leq ||x - x^k||^2 - ||x^{k+1} - x^k||^2 + \frac{1}{\lambda_k}(f(x) - f(x^{k+1})), \quad \forall k \in \mathbb{N} \quad (3.3)$$

para todo $x \in \mathbb{R}^n$. Em particular, vale

$$||x - x^{k+1}||^2 \leq ||x - x^k||^2 + \frac{1}{\lambda_k}(f(x) - f(x^{k+1})), \quad \forall k \in \mathbb{N}. \quad (3.4)$$

Demonstração: Seja $x \in \mathbb{R}^n$. Considere as igualdades

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

Isso implica que

$$||x - x^{k+1}||^2 = ||x - x^k||^2 - ||x^{k+1} - x^k||^2 + 2\langle x^k - x^{k+1}, x - x^{k+1} \rangle. \quad (3.5)$$

Pelo Teorema 3.1.1,

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

Da definição de $\partial f(x^{k+1})$ temos

$$\begin{aligned} f(x) &\geq f(x^{k+1}) + \langle 2\lambda_k(x^k - x^{k+1}), x - x^{k+1}\rangle \\ &= f(x^{k+1}) + 2\lambda_k\langle x^k - x^{k+1}, x - x^{k+1}\rangle, \end{aligned}$$

e isso implica que

$$\frac{1}{\lambda_k}(f(x) - f(x^{k+1})) \geq 2\langle x^k - x^{k+1}, x - x^{k+1}\rangle. \tag{3.6}$$

Substituindo (3.6) em (3.5) temos

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

que é a desigualdade (3.3). Para provar (3.4) basta usar (3.3) e o fato que $||x^{k+1} - x^k||^2 \geq 0$, para todo $k \in \mathbb{N}$. ■

Demonstraremos a seguir o principal resultado deste capítulo: a convergência do algoritmo ponto proximal.

Teorema 3.1.2 *Suponha que a sequência* $\{\lambda_k\}$ *é tal que* $\displaystyle\sum_{k=0}^{\infty} \frac{1}{\lambda_k} = +\infty$, *então*

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

onde $f^* = \inf_{x\in\mathbb{R}^n} f(x)$. *Além disso,* $\lim_{k\to+\infty} x^k = x^*$, *com* $x^* \in S$.

Demonstração: Seja S o conjunto do minimizadores da função convexa f. Se $x^{k+1} = x^k$, para algum $k \geq 0$ então pela definição do algoritmo obtemos $x^k \in S$ e o algoritmo para, ou seja, $x^k = x^{k+1} = x^{k+2} = \ldots$ e daí segue que $f(x^k) = f^*$ e não há mais nada a fazer. Agora suponhamos que $x^{k+1} \neq x^k$, para todo $k \geq 0$. Substituindo x por x^k em (3.4) obtemos que

$$0 < \lambda_k||x^{k+1} - x^k||^2 \leq f(x^k) - f(x^{k+1}), \quad \forall k \in \mathbb{N},$$

ou seja, concluímos que a sequência $\{f(x^k)\}$ é estritamente decrescente. Devemos provar que $\lim_{k\to+\infty} f(x^k) = f^*$. Suponhamos então, por absurdo, que $\lim_{k\to+\infty} f(x^k) > f^*$. Então existe $x \in \mathbb{R}^n$ e $\delta > 0$ tal que

$$f(x) < f(x^k) - \delta, \tag{3.7}$$

para todo $k > k_0$. Substituindo a desigualdade (3.7) em (3.4), temos

$$||x - x^{k+1}||^2 \leq ||x - x^k||^2 - \frac{\delta}{\lambda_k},$$

para todo $k > k_0$. Mas isso implica que

$$\frac{\delta}{\lambda_k} \leq ||x - x^k||^2 - ||x - x^{k+1}||^2,$$

para todo $k > k_0$. Somando termo a termo, obtemos

$$\sum_{k=k_0}^{j} \frac{1}{\lambda_k} \leq \frac{1}{\delta}(||x - x^0||^2 - ||x - x^{j+1}||^2) \leq \frac{1}{\delta}||x - x^0||^2.$$

Fazendo $j \to +\infty$ na desigualdade acima contrariamos o fato de que

$$\sum_{k=0}^{\infty} \frac{1}{\lambda_k} = +\infty.$$

Logo

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

Por outro lado, sabemos que $S \neq \emptyset$. Assim, tome $\bar{x} \in S$ qualquer. Desse modo $f(\bar{x}) \leq f(x^k)$, para todo $k \geq 0$. Substituindo x por $\bar{x}$ em (3.4) obtemos

$$||\bar{x} - x^{k+1}||^2 \leq ||\bar{x} - x^k||^2 + \frac{1}{\lambda_k}(f(\bar{x}) - f(x^{k+1})) \leq ||\bar{x} - x^k||^2, \quad \forall k \geq 0,$$

pois $f(\bar{x}) - f(x^k) \leq 0$. Logo,

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

para todo $k \geq 0$, e assim a sequência $\{x^k\}$ é Féjer convergente ao conjunto S. Logo, pelo Lema 2.4.1, temos que $\{x^k\}$ é limitada. Seja $\{x^{k_j}\}$ uma subsequência convergente de $\{x^k\}$, digamos $\lim_{j \to +\infty} x^{k_j} = x^*$. Da primeira parte $\lim_{j \to +\infty} f(x^{k_j}) = f^*$ e, sendo f contínua, $f(x^*) = f^*$, o que implica $x^* \in S$. Portanto, pela segunda parte do Lema 2.4.1, concluímos que a sequência $\{x^k\}$ converge para x^*, isto é, $\lim_{k \to +\infty} x^k = x^*$. ■

3.2 Caso quase convexo

Agora, iremos estudar a convergência do algoritmo 3.1 para o caso em que f é uma função quase convexa.

Teorema 3.2.1 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função quase convexa e semicontínua inferiormente. A sequência $\{x^k\} \subset \mathbb{R}^n$ gerada por (3.1) está bem definida e caracterizada pela relação*

$$\lambda_k(x^k - x^{k+1}) \in \partial_F f(x^{k+1}), \quad \forall k \in \mathbb{N}. \tag{3.8}$$

Demonstração: A prova será feita por indução sobre k. Para $k = 0$, o ponto x^0 é escolhido pela inicialização do algoritmo no passo 1. Agora, suponha que o método já tenha atingido a iterada x^k provaremos que existe a iterada $x^{k+1} \in \mathbb{R}^n$. Para isso, defina

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

Como S, o conjunto dos minimizadores de f, é não-vazio, temos que existe $x^* \in S$ tal que $f(x^*) \leq f(x)$, para todo $x \in \mathbb{R}^n$. Assim,

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

Logo, quando $||x|| \to +\infty$, temos que $f_k(x) \to +\infty$, ou seja, f_k é coerciva. Sendo f semicontínua inferiormente, temos que f_k também é, e pela Proposição 2.2.1 temos que f_k possui minimizador.

Agora, seja x^{k+1} um minimizador de f_k. Pela Proposição 2.2.8, temos que

$$0 \in \partial_F f_k(x^{k+1}), \quad \forall k \geq 0.$$

Ainda pela Proposição 2.2.8, temos que

$$0 \in \partial_F f(x^{k+1}) + \lambda_k(x^{k+1} - x^k),$$

o que equivale dizer que

$$\lambda_k(x^k - x^{k+1}) \in \partial_F f(x^{k+1}).$$

■

Observação 3.2.1 *Note que de (3.8) segue que se $x^k = x^{k+1}$, temos que o algoritmo para em um ponto crítico, ou seja, $0 \in \partial_F f(x^{k+1})$. Dessa forma, iremos supor de agora em diante que $x^k \neq x^{k+1}$, para todo $k \in \mathbb{N}$.*

Proposição 3.2.1 *A sequência $\{f(x^k)\}$ é estritamente decrescente e convergente.*

Demonstração: Como $x^k \neq x^{k+1}$, então

$$\langle x^{k-1} - x^k, x^{k-1} - x^k \rangle > 0.$$

Segue do Teorema 3.2.1 que existe $v^k \in \partial_F f(x^k)$ tal que $v^k = x^{k-1} - x^k$. Assim,

$$\langle v^k, x^{k-1} - x^k \rangle > 0.$$

Dessa forma, pela Proposição 2.2.9 usando a quase convexidade de f temos que

$$f(x^k) < f(x^{k-1}), \quad \forall k \in \mathbb{N}.$$

A convergência de $\{f(x^k)\}$ segue da limitação inferior de f. ■

Observação 3.2.2 *A sequência $\{x^k\}$ não cicla. Com efeito, suponha que exista $l > j$ tal que $x^l = x^j$. Da Proposição 3.2.1 temos que*

$$f(x^j) = f(x^l) < \ldots < f(x^{j+1}) < f(x^j)$$

que é uma contradição, então $x^l \neq x^j$.

Observação 3.2.3 *Considere o seguinte conjunto*

$$U := \{x \in \mathbb{R}^n : f(x) \leq \inf_{j \in \mathbb{N}} f(x^j)\}.$$

Observe que esse conjunto depende da escolha do ponto inicial x^0 e da sequência $\{\lambda_k\}$. Se $U = \emptyset$, então podemos concluir que

i) $\lim_{k \to +\infty} f(x^k) = \inf_{x \in \mathbb{R}^n} f(x)$,

ii) $\{x^k\}$ *é ilimitada.*

Dessa forma, assumiremos que $U \neq \emptyset$.

Proposição 3.2.2 *A sequência $\{x^k\}$ gerada pelo Algoritmo 3.1 é Fejér convergente a U. Além, disso, para todo $x \in U$, a sequência $\{||x - x^k||\}$ é convergente e $\lim_{k \to +\infty} ||x^k - x^{k-1}|| = 0$.*

Demonstração: Seja $x \in U$, então $f(x) \leq f(x^k)$. Segue de (3.8) que

$$g^k = \lambda_k (x^{k-1} - x^k) \in \partial_F f(x^k).$$

Sendo f quase convexa, da Proposição 2.2.9, obtemos

$$\langle x - x^k, x^{k-1} - x^k \rangle \leq 0. \tag{3.9}$$

Por outro lado, para todo $x \in \mathbb{R}^n$, temos

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

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

Agora, a última igualdade e (3.9) implica, em particular, para $x \in U$,

$$0 \leq ||x^k - x^{k-1}||^2 \leq ||x - x^{k-1}||^2 - ||x - x^k||^2. \tag{3.10}$$

Assim

$$||x - x^k||^2 \leq ||x - x^{k-1}||^2, \quad \forall x \in U. \tag{3.11}$$

Donde segue que $\{x^k\}$ é Fejér convergente a U. De (3.11), $\{||x - x^k||\}$ é uma sequência não-crescente limitada inferiormente e com isso convergente. Assim, tomando o limite quando $k \to +\infty$ em (3.10) e usando o resultado anterior, obtemos

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

■

Teorema 3.2.2 *A sequência $\{x^k\}$ gerada pelo Algoritmo 3.1 converge para um ponto de U e*

$$\lim_{k \to \infty} g^k = 0,$$

para algum $g^k \in \partial_F f(x^k)$. Além disso, se f é contínua, então $\{x^k\}$ converge para um ponto crítico de f.

Demonstração: Da Proposição 3.2.2, $\{x^k\}$ é Fejér convergente a U. Logo, segue do Lema 2.4.1 que $\{x^k\}$ é limitada. Assim, existe uma subsequência $\{x^{k_j}\}$ de $\{x^k\}$ convergindo para $\bar{x}$. Da f ser semicontínua inferiormente obtemos

$$\liminf_{j \to +\infty} f(x^{k_j}) \geq f(\bar{x}).$$

Como $\{f(x^k)\}$ é decrescente e converge então

$$f(\bar{x}) \leq \lim_{k \to +\infty} f(x^k) \leq f(x^k), \quad \forall k \in \mathbb{N}.$$

Isso implica que $\bar{x} \in U$. Logo, usando novamente o Lema 2.4.1, obtemos que $\{x^k\}$ converge para $\bar{x}$. Por outro lado, segue de (3.8) que

$$g^k = \lambda_k (x^{k-1} - x^k) \in \partial_F f(x^k).$$

Sendo $\{\lambda_k\}$ uma sequência limitada, segue da Proposição 3.2.2 que $\lim_{k \to \infty} g^k = 0$.

Finalmente, sendo $\{x^k\}$, $\{f(x^k)\}$ e $\{g^k\}$ tais que $g^k \in \partial_F f(x^k)$, $\lim_{k \to \infty} x^k = \bar{x}$, $\lim_{k \to \infty} f(x^k) = f(\bar{x})$ (da continuidade de f) e $\lim_{k \to \infty} g^k = 0$. Portanto, segue da definição de subdiferencial que $0 \in \partial_F f(\bar{x})$. ■

Capítulo 4

Método ponto proximal com distância de Bregman

Neste capítulo estudaremos um método ponto proximal interior para encontrar soluções de problemas de minimização com restrições não negativas denotado da seguinte forma:

$$\min\{f(x) \ : \ x \geq 0\}, \tag{4.1}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$ é semicontínua inferiormente (não necessariamente diferenciável). Por questões de completeza deste trabalho analisaremos o caso em que f é uma função convexa e quase convexa, mesmo o primeiro sendo um caso particular do segundo.

Para resolver (4.1) iremos considerar um método interior do tipo ponto proximal onde a regularização dada pela distância Euclidiana é substituída por uma distância de Bregman. Em problemas de otimização com restrições é crucial para a boa definição e eficiência do método que as iteradas permaneçam no interior do conjunto de restrições. Esse é o caso do método do ponto proximal com distância de Bregman considerado a seguir: De agora em diante, neste capítulo iremos considerar $\{x^k\}$ a

---

Algoritmo 4.1 Método do Ponto Proximal com distância de Bregman

Passo 1: Escolha $x^0 > 0$ *e* $\{\lambda_k\}$ *uma sequência de números reais tal que* $0 < \lambda_k \leq \lambda$, *para todo* $k \in \mathbb{N}$;

Passo 2: Dado x^k *calcule*

$$x^{k+1} \in \arg\min_{x \in \mathbb{R}^n_+} \{f(x) + \lambda_k D_h(x, x^k)\}, \tag{4.2}$$

onde h *é uma função de Bregman associada à* D_h.

Passo 3: Se $x^{k+1} = x^k$, *pare. Contrário, faça* $k = k + 1$ *e retorne ao Passe 2.*

---

sequência gerada pelo algoritmo 4.1 acima.

4.1 Caso convexo

Nesta seção iremos considerar o caso em que f é uma função convexa. Além disso, iremos supor que o problema (4.1) tem solução, e com isso, f é limitada inferiormente, e que h é fronteira coerciva em relação ao conjunto de restrições $\mathbb{R}^n_+$.

A seguir, provaremos a boa definição do algoritmo MPPB.

Teorema 4.1.1 *A sequência $\{x^k\}$ está bem definida e contida em $\mathbb{R}^n_{++}$.*

Demonstração: Seja $\beta \in \mathbb{R}$ um limite inferior de f, ou seja, $f(x) \geq \beta$ para todo $x \in \mathbb{R}^n$. Definida

$$f_k(x) := f(x) + \lambda_k D_h(x, x^k).$$

Então $f_k \geq \beta + \lambda_k D_h(x, x^k)$ e segue de B3 que o conjunto de nível de f_k é limitado. Pela continuidade de f_k segue que o conjunto de nível de f_k é fechado, logo compacto e o mínimo é atingido. Por B2, a função h é estritamente convexa, donde segue que a função f_k é estritamente convexa. Logo, pelo Teorema 2.2.3 f_k possui único minimizador $x^{k+1} \in \mathbb{R}^n_+$. Provaremos agora que $x^{k+1} > 0$. Com efeito, segue de (4.2) combinado com o Teorema 2.2.7 que

$$0 \in \partial(f(\cdot) + \lambda_k D_h(\cdot, x^k))(x^{k+1}).$$

Assim, segue da combinação das Proposições 2.2.5, 2.2.6 e 2.3.1(ii) com a inclusão acima que

$$\lambda_k \nabla h(x^k) \in \partial(f + \lambda_k h)(x^{k+1}). \tag{4.3}$$

Agora, basta mostrar que, sob a hipótese B6, $\partial(f + \lambda_k h)(x) = \emptyset$, para todo x na fronteira de $\mathbb{R}^n_{++}$, e isso implicará, tendo em vista (4.3), que $x^{k+1} > 0$.

Seja $x \in \operatorname{fr} \mathbb{R}^n_{++}$ e assuma que exista $\xi \in \partial(f + \lambda_k h)(x)$. Tome $z \in \mathbb{R}^n_{++}$ e defina

$$y^l = (1 - \varepsilon_l)x + \varepsilon_l z \tag{4.4}$$

com $\varepsilon_l > 0$ e $\lim\limits_{l\to\infty} \varepsilon_l = 0$. Dessa forma, $y^l \in \mathbb{R}^n_{++}$ e $\lim\limits_{l\to\infty} y^l = x$. Como $\xi \in \partial(f + \lambda_k h)(x)$, temos que

$$f(y^l) - f(x) + \lambda_k(h(y^l) - h(x)) \geq \langle \xi, y^l - x \rangle = \varepsilon_l \langle \xi, z - x \rangle. \tag{4.5}$$

Como h é diferenciável e convexa, segue do Teorema 2.2.2 (2) que

$$h(x) - h(y^l) \geq \langle \nabla h(y^l), x - y^l \rangle.$$

Aplicando essa última desigualdade em (4.5), obtemos

$$f(y^l) - f(x) + \lambda_k \langle \nabla h(y^l), y^l - x \rangle \geq \varepsilon_l \langle \xi, z - x \rangle. \tag{4.6}$$

Por outro lado, segue de (4.4) que $y^l - x = \frac{\varepsilon_l}{1-\varepsilon_l}(z - y^l)$ e $f(y^l) \leq (1-\varepsilon_l) f(x) + \varepsilon_l f(z)$, tendo em vista a convexidade de f. Usando esses fatos em (4.6), temos que

$$\varepsilon_l (f(z) - f(x)) + \lambda_k \frac{\varepsilon_l}{1-\varepsilon_l} \langle \nabla h(y^l), z - y^l \rangle \geq \varepsilon_l \langle \xi, z - x \rangle$$

que implica

$$\frac{1-\varepsilon_l}{\lambda_k} (f(x) - f(z) + \langle \xi, z - x \rangle) \leq \langle \nabla h(y^l), z - y^l \rangle. \tag{4.7}$$

Como $\lim_{l \to \infty} y^l = x \in \operatorname{fr} \mathbb{R}^n_{++}$, B6 implica que o lado direito de (4.7) tende para $-\infty$ quando l tende para ∞, enquanto o lado esquerdo é um limite finito. Essa contradição implica que $\partial(f + \lambda_k h) = \emptyset$ para todo $x \in \operatorname{fr} \mathbb{R}^n_{++}$. Com isso $x^{k+1} \in \mathbb{R}^n_{++}$ e a prova está concluida. ■

Teorema 4.1.2 *Para cada solução $\bar{x}$ de* (4.1)*, temos que*

$$D_h(\bar{x}, x^{k+1}) \leq D_h(\bar{x}, x^k) - D_h(x^{k+1}, x^k), \quad \forall k \in \mathbb{N}.$$

Demonstração: Aplicando a Proposição 2.3.1(i) com $x = \bar{x}$, $y = x^k$ e $z = x^{k+1}$, obtemos

$$D_h(\bar{x}, x^k) - D_h(\bar{x}, x^{k+1}) - D_h(x^{k+1}, x^k) = \langle \nabla h(x^k) - \nabla h(x^{k+1}), x^{k+1} - \bar{x} \rangle. \tag{4.8}$$

Segue de (4.2) e do Teorema 2.2.7 que

$$0 \in \partial(f(\cdot) + \lambda_k D_h(\cdot, x^k))(x^{k+1}), \tag{4.9}$$

onde usamos o fato que $x^{k+1} \in \mathbb{R}^n_{++}$ para garantir que $N_{\mathbb{R}^n_+}(x^{k+1}) = \{0\}$. Com isso, combinando (4.9) com a Proposição 2.3.1(ii), temos que

$$\lambda_k [\nabla h(x^k) - \nabla h(x^{k+1})] \in \partial f(x^{k+1}). \tag{4.10}$$

Combinando (4.8) e (4.10) com a definição de subgradiente, temos que

$$\begin{aligned} D_h(\bar{x}, x^k) - D_h(\bar{x}, x^{k+1}) - D_h(x^{k+1}, x^k) &= \frac{1}{\lambda_k} \langle \lambda_k (\nabla h(x^k) - \nabla h(x^{k+1})), x^{k+1} - \bar{x} \rangle \\ &\geq \frac{1}{\lambda_k} (f(x^{k+1}) - f(\bar{x})) \geq 0, \end{aligned}$$

usando o fato que $\bar{x}$ minimiza f em $\mathbb{R}^n_+$ e que $\lambda_k > 0$, para todo $k \in \mathbb{N}$. Isso finaliza a prova. $\blacksquare$

Corolário 4.1.1 *A sequência $\{x^k\}$ é limitada. Além disso, se para alguma subsequência $\{x^{k_j}\}$ de $\{x^k\}$ tivermos que $\lim\limits_{j\to\infty} x^{k_j} = \hat{x}$ então $\lim\limits_{j\to\infty} x^{k_j+1} = \hat{x}$.*

Demonstração: Segue do Teorema 4.1.2 que $D_h(\bar{x}, x^{k+1}) \leq D_h(\bar{x}, x^k)$, para todo $k \in \mathbb{N}$, ou seja, $\{D_h(\bar{x}, x^k)\}$ é decrescente e não negativa, portanto convergente, e $D_h(x^{k+1}, x^k) \leq D_h(\bar{x}, x^k) - D_h(\bar{x}, x^{k+1})$, para todo $k \in \mathbb{N}$, implicando que

$$\lim_{k\to\infty} D_h(x^{k+1}, x^k) = 0. \tag{4.11}$$

Como $\{D_h(\bar{x}, x^k)\}$ é decrescente, temos $D_h(\bar{x}, x^k) \leq D_h(\bar{x}, x^0)$ e assim $\{x^k\}$ é limitada por B3. Se $\lim\limits_{j\to\infty} x^{k_j} = \hat{x}$ para uma subsequência $\{x^{k_j}\}$ de $\{x^k\}$ então $\lim\limits_{j\to\infty} x^{k_j+1} = \hat{x}$ por B5. $\blacksquare$

Teorema 4.1.3 *A sequência $\{x^k\}$ converge para uma solução $\hat{x}$ do problema* (4.1).

Demonstração: Primeiramente, provaremos que todos os pontos de acumulação da sequência $\{x^k\}$ são soluções de (4.1). Com efeito, tome uma solução $\bar{x}$ de (4.1). Seja $\hat{x}$ um ponto de acumulação de $\{x^k\}$ e $\{x^{k_j}\}$ uma subsequência de $\{x^k\}$ tal que $\lim\limits_{j\to\infty} x^{k_j} = \hat{x}$. A existência de $\hat{x}$ segue do Corolário 4.1.1 que também garante que $\lim\limits_{j\to\infty} x^{k_j+1} = \hat{x}$. Como $\lambda_k \leq \lambda$, para todo $k \in \mathbb{N}$, temos que

$$\begin{aligned}
0 &\leq \frac{1}{\lambda}(f(x^{k_j+1}) - f(\bar{x})) \leq \frac{1}{\lambda_k}(f(x^{k_j+1}) - f(\bar{x})) \\
&\leq D_h(\bar{x}, x^{k_j}) - D_h(\bar{x}, x^{k_j+1}) - D_h(x^{k_j+1}, x^{k_j}) \to 0, \quad j \to \infty
\end{aligned}$$

usando a convergência de $\{D_h(\bar{x}, x^k)\}$ e B5. Logo, $f(\hat{x}) = f(\bar{x})$. Como $\mathbb{R}^n_+$ é fechado e $\{x^k\} \subset \mathbb{R}^n_+$, temos $\hat{x} \in \mathbb{R}^n_+$ e então $\hat{x}$ revolve (4.1).

A fim de completar a prova precisamos de um teorema de Féjer convergência para distâncias de Bregman, que de fato se verifica, mas podemos proceder diretamente: seja $\hat{x}$ um ponto de acumulação de $\{x^k\}$ e tome uma subsequência $\{x^{k_j}\}$ de $\{x^k\}$ tal que $\lim\limits_{j\to\infty} x^{k_j} = \hat{x}$. Então por B4, temos que $\lim\limits_{j\to\infty} D_h(\hat{x}, x^{k_j}) = 0$ e $\hat{x}$ resolve (4.1). Pelo Teorema 4.1.2 $\{D_h(\hat{x}, x^k)\}$ é uma sequência não negativa decrescente, com isso convergente. Tendo uma subsequência convergindo para 0 segue que a sequência inteira converge para 0 e por B4 mais uma vez obtemos $\lim\limits_{k\to\infty} x^k = \hat{x}$ e a prova está concluída. $\blacksquare$

4.2 Caso quase convexo

Nesta seção iremos considerar o Algoritmo 4.1 para o caso em que f é uma função continuamente diferenciável e quase convexa. Além disso, iremos supor que o problema 4.1 tem solução e que a função de Bregman h é separável em $\mathbb{R}^n_{++}$.

Proposição 4.2.1 *A sequência* $\{x^k\}$ *está bem definida e contida em* $\mathbb{R}^n_{++}$.

Demonstração: A prova será feita por indução. O ponto inicial $x^0 > 0$ é dado pela inicialização do algoritmo. Agora, supondo que existe $x^k > 0$. Mostraremos a existência da iterada $k+1$. Como f é limitada inferiormente e por B3 temos que $D_h(x, x^k)$ tem nível limitado para todo x, obtemos que f_k é também de nível limitado. Agora, pela continuidade de f_k segue que o conjunto de nível de f_k é fechado, logo compacto. Com isso, temos que f_k tem minimizador $z \geq 0$ (que pode não ser único devido a não convexidade de f). Iremos mostrar que $z > 0$. Primeiro, note que z satisfaz as condições de otimalidade de f_k, isto é,

$$z_i \geq 0, \quad (\nabla f_k(z))_i \geq 0, \quad z_i(\nabla f_k(z))_i = 0, \quad \forall i = 1, \ldots, n, \tag{4.12}$$

onde $\nabla f_k(z) = \nabla f(z) + \lambda_k \nabla_1 D_h(z, x^k)$.

Pela Proposição 2.3.1 item ii), sabemos que $\nabla_1 D_h(z, x^k) = \nabla h(z) - \nabla h(x^k)$. Assim, como h é separável

$$(\nabla_1 D_h(z, x^k))_i = h_i'(z_i) - h_i'(x_i^k), \quad \forall i = 1, \ldots, n.$$

Sendo $\nabla f(\cdot)$ uma aplicação contínua e sendo $L_{f_k}(\alpha)$ compacto, temos que $\nabla f(L_{f_k}(\alpha))$ é limitado. Logo, como $z \in L_{f_k}(\alpha)$ segue que existe $c > 0$ tal que $|\nabla f(z)| \leq c$ e assim $(\nabla f(z))_i \leq c$, para todo $i = 1, \ldots, n$. Então,

$$\begin{aligned}
(\nabla f_k(z))_i &= (\nabla f(z))_i + \lambda_k (\nabla_1 D_h(z, x_i^k))_i \\
&= (\nabla f(z))_i + \lambda_k (h_i'(z_i) - h_i'(x_i^k) \\
&\leq c + \lambda_k (h_i'(z_i) - h_i'(x_i^k)).
\end{aligned} \tag{4.13}$$

Agora, como h tem zona coerciva em $\mathbb{R}^n_{++}$ e separável, então pela Observação 2.3.1 segue que $\lim\limits_{t \to 0^+} h_i'(t) = -\infty$. Dessa forma, concluímos que não se pode ter $z_i = 0$, para algum $i = 1, \ldots, n$, pois nesse caso temos que $z \in \operatorname{fr} \mathbb{R}^n_+$ donde segue que existe uma sequência $\{y^l\} \subset \mathbb{R}^n_+$ tal que $\lim_{l \to +\infty} y^l = z$, ou seja, $\lim_{l \to +\infty} y_i^l = 0$, para pelo menos um $i = 1, \ldots, n$. Logo, segue de (4.13) que

$$(\nabla f_k(z))_i = \lim_{l \to +\infty} (\nabla f_k(y^l))_i = -\infty,$$

que contradiz a otimalidade de z (na segunda inequação de (4.12)), então não podemos ter $z_i = 0$, ou seja, $z > 0$. Portanto, basta tomar $x^{k+1} = z > 0$ que conclui a prova. ∎

Observação 4.2.1 *Da última proposição,* $z \in \arg\min\{f_k(x), x \geq 0\}$ *é tal que* $z > 0$*. Assim, da condição de otimalidade do problema, temos* $\nabla f_k(z) = 0$*, e então*

$$\nabla f(z) = -\lambda_k \nabla_1 D_h(z, x^k),$$

para todo $k \in \mathbb{N}$*. Em particular,*

$$\nabla f(x^{k+1}) = -\lambda_k \nabla_1 D_h(x^{k+1}, x^k),$$

para todo $k \in \mathbb{N}$*.*

Proposição 4.2.2 *Seja* $\{x^k\}$ *uma sequência gerada pelo Algoritmo 4.1. Então,*

(i) $0 \leq \lambda_k D_h(x^{k+1}, x^k) \leq f(x^k) - f(x^{k+1})$*, para todo* $k \in \mathbb{N}$*;*

(ii) $\sum_{k=0}^{+\infty} \lambda_k D_h(x^{k+1}, x^k) < +\infty$ *e, em particular,*

$$\lim_{k \to +\infty} \lambda_k D_h(x^{k+1}, x^k) = 0;$$

(iii) $f(x^k)$ *é decrescente e convergente.*

Demonstração: (i) Pela definição de x^{k+1} dada no Algoritmo 4.1, temos que

$$f(x^{k+1}) + \lambda_k D_h(x^{k+1}, x^k) \leq f(x^k) + \lambda_k D_h(x^k, x^k), \quad \forall k \in \mathbb{N}.$$

Assim, uma vez que $D_h(x^k, x^k) = 0$, temos que

$$\lambda_k D_h(x^{k+1}, x^k) \leq f(x^k) - f(x^{k+1}), \quad \forall k \in \mathbb{N}. \tag{4.14}$$

Por outro lado, como $\lambda_k > 0$ segue que $\lambda_k D_h(x^{k+1}, x^k) \geq 0$. Que junto com (4.14) prova-se o item (i).

(ii) Somando (4.14) com $k = 0, \ldots, l$, obtemos

$$\sum_{k=0}^{l} \lambda_k D_h(x^{k+1}, x^k) \leq \sum_{k=0}^{l} [f(x^k) - f(x^{k+1})] = f(x^0) - f(x^{l+1}) \leq f(x^0) - f^*,$$

onde na última desigualdade usamos o fato de que f é limitada inferiormente. Tomando o limite com $l \to +\infty$ na desigualdade acima, obtemos que

$$\sum_{k=0}^{+\infty} \lambda_k D_h(x^{k+1}, x^k) < +\infty$$

e, em particular,

$$\lim_{k \to +\infty} \lambda_k D_h(x^{k+1}, x^k) = 0$$

provando o item (ii).

(iii) De (i), encontramos $f(x^{k+1}) \leq f(x^k)$ para todo $k \in \mathbb{N}$. Então $f(x^k)$ é decrescente e, além disso, temos que f é limitada inferiormente. Portanto, $f(x^k)$ é convergente concluindo a prova. ∎

A seguir, provaremos dois resultados importantes de convergência. Primeiro, estabelecemos convergência para um ponto KKT quando os parâmetros de regularização λ_k são limitados. E no segundo resultado, mostraremos que a sequência de iteradas converge para um ponto solução do problema (4.1) sob a hipótese de que λ_k converge a zero. Antes, provaremos alguns resultados semelhantes à Fejér convergência, mas para distância de Bregman.

Lema 4.2.1 *$D_h(x^*, x^{k+1}) \leq D_h(x^*, x^k) - D_h(x^{k+1}, x^k)$ para todo $k \in \mathbb{N}$ e algum x^* solução de* (4.1).

Demonstração: Seja x^* solução de (4.1). Tomando $x = x^*$, $y = x^k$ e $z = x^{k+1}$ na Proposição 2.3.1(i), obtemos

$$D_h(x^*, x^k) - D_h(x^*, x^{k+1}) - D_h(x^{k+1}, x^k) = \langle \nabla h(x^k) - \nabla h(x^{k+1}), x^{k+1} - x^* \rangle. \quad (4.15)$$

Pela Observação 4.2.1 e Proposição 2.3.1(ii), segue que

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

que substituindo em (4.15) obtemos

$$D_h(x^*, x^k) - D_h(x^*, x^{k+1}) - D_h(x^{k+1}, x^k) = \frac{1}{\lambda_k} \langle \nabla f(x^{k+1}), x^{k+1} - x^* \rangle. \quad (4.16)$$

Como $f(x^*) \leq f(x^k)$, para todo $k \in \mathbb{N}$, a quase convexidade da f pela Proposição (2.2.9), implica que

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

Assim, usando esse fato em (4.16) e sabendo que $\lambda_k > 0$, para todo $k \in \mathbb{N}$, obtemos

$$D_h(x^*, x^k) - D_h(x^*, x^{k+1}) - D_h(x^{k+1}, x^k) \geq 0,$$

que finaliza a prova. ∎

O próximo resultado estabelece a existência de um ponto limite para sequência $\{x^k\}$.

Proposição 4.2.3 *A sequência $\{x^k\}$ gerada pelo Algoritmo 4.1 é limitada.*

Demonstração: Pelo Lema 4.2.1, $D_h(x^*, x^{k+1}) \leq D_h(x^*, x^k)$. Assim

$$D_h(x^*, x^k) \leq D_h(x^*, x^0), \quad \forall k \in \mathbb{N}. \tag{4.17}$$

Agora, segue de (B3) que $D_h(x^*, x)$ é um conjunto limitado para todo $x \in \mathbb{R}^n_{++}$. Como $x^0 \in \mathbb{R}^n_{++}$, temos assim a limitação de $\{x^k\}$. ■

No resultado anterior, asseguramos que há pontos de acumulação para a sequência $\{x^k\}$. Na proposição abaixo mostramos que a $\{x^k\}$ sequência inteira converge.

Proposição 4.2.4 *A sequência $\{x^k\}$ gerada pelo Algoritmo 4.1 é convergente.*

Demonstração: Seja $\overline{x}$ um ponto de acumulação de $\{x^k\}$, cuja a existência é assegurada pela Proposição 4.2.3. Da Proposição 4.2.2 item (iii), temos

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

para todo $k \in \mathbb{N}$, e pela continuidade de f, segue que

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

para todo $k \in \mathbb{N}$. Dessa forma, usando o mesmo procedimento apresentado no Lema 4.2.1, encontramos que $\{D_h(\overline{x}, x^k)\}$ é decrescente. Agora, seja $\{x^{k_j}\}$ uma subsequência de $\{x^k\}$ tal que $\lim\limits_{j \to +\infty} x^{k_j} = \overline{x}$. Assim, por (B4), temos que

$$\lim_{j \to +\infty} D_h(\overline{x}, x^{k_j}) = 0.$$

Logo, $\{D_h(\overline{x}, x^k)\}$ é uma sequência decrescente não negativa que tem uma subsequência convergindo para 0. Então, toda a sequência $\{D_h(\overline{x}, x^k)\}$ converge para 0, isto é,

$$\lim_{k \to +\infty} D_h(\overline{x}, x^k) = 0. \tag{4.18}$$

Consideremos $\tilde{x}$ um outro ponto de acumulação de $\{x^k\}$, ou seja, existe $\{x^{k_l}\}$ tal que

$$\lim_{l \to +\infty} x^{k_l} = \tilde{x}.$$

Portanto, por (4.18) temos $\lim\limits_{l \to +\infty} D_h(\overline{x}, x^{k_l}) = 0$. Tomando em (B5) $y^k = x^{k_l}$, $y^* = \tilde{x}$ e $z^k = \overline{x}$, obtemos que

$$\tilde{x} = \overline{x}.$$

Então, $\{x^k\}$ possui único ponto de acumulação e com isso é convergente, concluindo a prova. ■

A seguir, apresentamos os dois principais resultados de convergência desta seção.

Teorema 4.2.1 *A sequência* $\{x^k\}$ *gerada pelo Algoritmo 4.1 converge para um ponto KKT do problema* (4.1).

Demonstração: Pela Proposição 4.2.4, temos que $\{x^k\}$ é convergente. Seja $\overline{x} = \lim\limits_{k\to+\infty} x^k$. Iremos mostrar que $\overline{x}$ é um ponto KKT do problema (4.1), isto é,

$$\overline{x_i} \geq 0, \quad (\nabla f(\overline{x}))_i \geq 0 \quad \text{e} \quad \overline{x_i}(\nabla f(\overline{x}))_i = 0, \quad \forall i = 1, \ldots, n. \tag{4.19}$$

Pela Proposição 4.2.1 a primeira condição em (4.19) é verificada. Para provar as outras duas condições em (4.19) consideramos os conjuntos

$$I(\overline{x}) = \{i \in \{1, \ldots, n\} : \overline{x_i} = 0\} \quad \text{e} \quad J(\overline{x}) = \{i \in \{1, \ldots, n\} : \overline{x_i} > 0\}.$$

Analisaremos os casos quando $i \in I(\overline{x})$ e $i \in J(\overline{x})$.

1º Caso: Se $i \in I(\overline{x})$. Suponha por contradição que $(\nabla f(\overline{x}))_i < 0$. Como ∇f é uma aplicação contínua, segue que

$$\lim_{k\to+\infty} (\nabla f(x^{k+1}))_i = (\nabla f(\overline{x}))_i < 0.$$

Assim, existe $k_0 \in \mathbb{N}$ tal que

$$(\nabla f(x^{k+1}))_i < 0, \quad \forall k \geq k_0. \tag{4.20}$$

Agora, pela Observação 4.2.1, Proposição 2.3.1 item (ii) e devido ao fato de h ser separável, obtemos

$$(\nabla f(x^{k+1}))_i = -\lambda_k(\nabla_1 D_h(x^{k+1}, x^k))_i = \lambda_k(h_i'(x_i^k)) - (h_i'(x_i^{k+1})), \quad \forall k \geq k_0. \tag{4.21}$$

Como $\lambda_k > 0$, para todo $k \in \mathbb{N}$, por (4.20) e (4.21) segue que

$$(h_i'(x_i^k) - h_i'(x_i^{k+1})) < 0.$$

Por (B1) e Teorema 2.2.2, temos que

$$[h_i'(x_i^k) - h_i'(x_i^{k+1})](x_i^k - x_i^{k+1}) > 0.$$

Logo, devemos ter $x_i^k - x_i^{k+1} < 0$, para todo $k \geq k_0$. Portanto,

$$0 = \overline{x_i} = \lim_{k\to+\infty} x_i^k \geq x_i^{k^0} > 0.$$

Isso é uma contradição. Então, $(\nabla f(\overline{x}))_i \geq 0$ para todo $i \in I(\overline{x})$.

A última igualdade em (4.19) é fácil de ser verificada, pois se $i \in I(\overline{x})$ temos que $\overline{x_i} = 0$, donde segue que

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

2º Caso: Se $i \in J(\overline{x})$. Como $\lim\limits_{k\to+\infty} x_i^k = \overline{x_i} > 0$ temos

$$\lim_{k\to+\infty} \frac{x_i^{k+1}}{x_i^k} = 1,$$

isto é,

$$\lim_{k\to+\infty} x_i^{k+1} = \lim_{k\to+\infty} x_i^k = \overline{x_i}.$$

Essa última afirmação e a continuidade de h_i' (assegurada por (B1)) implica que $\lim\limits_{k\to+\infty}(h_i'(x_i^k) - h_i'(x_i^{k+1})) = 0$. Portanto, como a sequência λ_k é limitada, tomando o limite com $k \to +\infty$ em (4.21), temos

$$(\nabla f(\overline{x}))_i = \lim_{k\to+\infty}(\nabla f(x^{k+1}))_i = \lim_{k\to+\infty} \lambda_k(h_i'(x_i^k) - h_i'(x_i^{k+1})) = 0.$$

Então, $(\nabla f(\overline{x}))_i = 0$ para todo $i \in J(\overline{x})$, donde segue que

$$(\nabla f(\overline{x}))_i \geq 0 \quad \text{e} \quad \overline{x_i}(\nabla f(\overline{x}))_i = 0.$$

E a prova está finalizada. ∎

No próximo resultado de convergência assumimos que a sequência de parâmetro proximal satisfaz

$$\lim_{k\to+\infty} \lambda_k = 0.$$

Observação 4.2.2 *A hipótese acima é comum em métodos do tipo proximal; veja por exemplo [10, 13]. Além disso, em [26] uma convergência super linear é obtida para o método do ponto proximal clássico com essa hipótese.*

Teorema 4.2.2 *Se $\lambda_k \to 0$, então a sequência $\{x^k\}$ gerada pelo Algoritmo 4.1 converge para uma solução do problema* (4.1).

Demonstração: *Por definição de x^{k+1} como dado pelo Algoritmo 4.1, temos*

$$f(x^{k+1}) + \lambda_k D_h(x^{k+1}, x^k) \leq f(x) + \lambda_k D_h(x, x^k), \quad \forall x \in \mathbb{R}_+.$$

Com isso, seja x^* uma solução do problema (4.1) *e tomemos $x = x^*$ na inequação acima. Logo,*

$$f(x^{k+1}) + \lambda_k D_h(x^{k+1}, x^k) \leq f(x^*) + \lambda_k D_h(x^*, x^k). \tag{4.22}$$

Pela Proposição 4.2.4 sabemos que $\{x^k\}$ *é convergente. Seja* $\overline{x} = \lim\limits_{k\to+\infty} x^k$. *Como* $\lim\limits_{k\to+\infty} \lambda_k D_h(x^{k+1}, x^k) = 0$ *assegurado pela Proposição 4.2.2 (ii),* f *é contínua, por* (4.17) *a sequência* $\{D_h(x^*, x^k)\}$ *é limitada e* $\lim\limits_{k\to+\infty} \lambda_k = 0$, *tomando o limite com* $k \to +\infty$ *em* (4.22), *obtemos*

$$f(\overline{x}) \leq f(x^*),$$

que implica imediatamente que $\overline{x}$ *é uma solução do problema* (4.1) *e prova está concluída.* ∎

Capítulo 5

Conclusão

Neste trabalho estudamos a convergência do método do ponto proximal para resolver um problema de minimização restrita ou irrestrita para os casos em que a função objetivo é convexa e, mais geral, quase convexa. No caso irrestrito e convexo, supondo que a função objetivo é limitada inferiormente, mostramos a convergência da sequência gerada pelo método do ponto proximal para um minimizador de f *a partir de um ponto inicial qualquer supondo que a sequência de parâmetros regularizadores* $\{\lambda_k\}$ *é limitada e* $\sum_{k=0}^{+\infty} \frac{1}{\lambda_k} = +\infty$. *Para o caso quase convexo, apenas supondo que* $\{\lambda_k\}$ *é limitada provamos a convergência para um ponto crítico de* f.

No caso restrito, consideramos o problema de minimizar uma função f *restrita a* $\mathbb{R}^n_+$ *para os casos em que* f *é uma função convexa ou quase convexa. Para isso, substituímos a distância Euclidiana no método do ponto proximal por uma distância de Bregman. Isso permite que as iterações do método permaneçam em* $\mathbb{R}^n_{++}$. *Para o caso convexo, supondo que* $\{\lambda_k\}$ *é limitada obtemos convergência para um minimizador de* f *em* $\mathbb{R}^n_+$. *Para o caso quase convexo e continuamente diferenciável, sob a mesma hipótese em* $\{\lambda_k\}$, *obtemos convergência do método para um ponto KKT. Porém, se supormos que a sequência* $\{\lambda_k\}$ *converge para zero garantimos a convergência para um minimizador de* f *em* $\mathbb{R}^n_+$.

Trabalhos futuros

- *Nos resultados da seção 4.2 a função objetivo* f *considerada é quase convexa e continuamente diferenciável. Além disso, a função de Bregman* h *usada na regularização do método do ponto proximal é separável. Dessa forma, seria natural tentar obter os mesmos resultados de convergência provados na seção 4.2 para o caso em que* f *é não diferenciável e* h *não necessariamente separável;*

- *Sabe-se que toda função convexa definida no* $\mathbb{R}^n$ *é localmente Lipschitz e também quase convexa. Até onde sabemos, o método do ponto proximal para funções localmente Lipschitz com distância de Bregman ainda não foi conside-*

rado. Isso seria um resultado interessante tendo em vista que existem funções localmente Lipschitz que não necessariamente são quase convexas;

- *Extensões do métodos do ponto proximal para otimização multi objetivo ou vetorial é um amplo campo de pesquisa. Nesse contexto, as distâncias de Bregman ainda podem ser exploradas.*

Referências Bibliográficas

[1] *AUSLENDER, A., TEBOULLE, M., "Interior gradient and proximal methods for convex and conic optimization",* SIAM J Optim.*, v. 16, pp. 697–725, 2006.*

[2] *AUSLENDER, A., TEBOULLE, M., BEN-TIBA, S., "Interior Proximal and Multiplier Methods Based on Second Order Homogeneous Functionals",* Math. Oper. Research*, v. 24, pp. 645–668, 1999.*

[3] *BAZAARA, M.S., SHERALI, H.D., SHETTY, C.M., "Nonlinear Programming: Theory and Algorithms,",* 2nd Edition, John Wiley and Sons, New York*, 1993.*

[4] *BENTO, G.C., SOUBEYRAN, A., "Generalized inexact proximal algorithms: Routine's formation with resistance to change, following worthwhile changes",* J. Optim. Theory Appl.*, v. 166, pp. 172–187, 2016.*

[5] *BENTO, G. C., SOUBEYRAN, A., "A generalized inexact proximal point method for nonsmooth functions that satisfies Kurdyka-Lojasiewicz inequality",* Set-Valued Var. Anal.*, v. 23, pp. 501–517, 2015.*

[6] *BREGMAN, L., "The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming",* USSR Computational Mathematics and Mathematical Physics *, v. 7, pp. 200–217, 1967.*

[7] *BURACHIK, R.S., SVAITER, B.F., "A relative error tolerance for a family of generalized proximal point methods",* Math. Oper. Res.*, v. 26, pp. 816–831, 2001.*

[8] *CENSOR, Y., ZENIOS, S. A., "Proximal minimization algorithm with D-functions",* J. Optim. Theory Appl.*, v. 73, pp. 451–464, 1992.*

[9] *CHEN, G., TEBOULLE, M., "Convergence analysis of proximal-like optimization algorithm using Bregman functions",* SIAM J. Optim.*, v. 3, pp. 538–543, 1993.*

[10] *CUNHA, G.F.M., DA CRUZ NETO, J.X., OLIVEIRA, P.R., "A proximal point algorithm with ϕ-divergence to quasiconvex programming",* Optimization*, v. 59, pp. 777–792, 2010.*

[11] *ECKSTEIN, J., "Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming",* Math. Oper. Res.*, v. 18, pp. 202–226, 1993.*

[12] *ECKSTEIN, J., "Approximate iterations in Bregman-function-based proximal algorithms",* Math Program.*, v. 83, pp. 113–123, 1988.*

[13] *GÜLER, O., "New proximal point proximal algorithms for convex minimization",* SIAM J. Control. Optim.*, v. 2(4), pp. 649–664, 1992.*

[14] *IUSEM, A.N., "Métodos de Ponto Proximal em Otimização",* XX Colóquio Brasileiro de Matemática, IMPA, Rio de Janeiro*, 1995.*

[15] *IZMAILOV, A., SOLODOV, M., "Otimização - Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade",* IMPA, Rio de Janeiro*, v. 01, 2005.*

[16] *KIWIEL, K. C., "Proximal minimization methods with generalized Bregman functions",* SIAM J. Control Optim.*, v. 35, pp. 1142–1168, 1997.*

[17] *LANGENBERG, N., "Interior proximal extragradient method for equilibrium problems",* Optim: J Math Program Oper Res.*, v. 64, pp. 2145–2161, 2015.*

[18] *LANGENGERG, N., TICHATSCHKE, R., "Interior proximal methods for quasiconvex optimization",* J. Glob. Optim.*, v. 52(3), pp. 641–661, 2012.*

[19] *LIMA, E.L., "Curso de análise",* Projeto Euclides, IMPA, Rio de Janeiro, Nona edição*, v. 02, 2006.*

[20] *LIMA, E.L., "Álgebra linear",* Coleção Matemática Universitária, IMPA, Rio de Janeiro, Sétima edição*, 2004.*

[21] *MARTINET, B., "Regularisation d'inéquations variationelles par approximations succesives",* Rev. Française d'Inform. Recherche Oper.*, v. 4, pp. 154–159, 1970.*

[22] *MOREAU, J., "Proximité et dualité dans un espace hilbertien",* Bulletin de la Societé Mathématique de France*, v. 93, pp. 273–299, 1965.*

[23] *MORENO, F.G., OLIVEIRA, P. R., SOUBEYRAN, A., "A proximal point algorithm with quasi distance. Application to habit's formation"*, Optimization*, v. 61, pp. 1383–1403, 2012.*

[24] *PAN, S., CHEN, J.S., "Entropy-like proximal algorithms based on a second-order homogeneous distance function for quasi-convex programming",* J. Glob. Optim.*, v. 39, pp. 555–575, 2007.*

[25] *ROCKAFELLAR, R.T., WETS, R.J-B., "Variational Analysis",* Springer, Berlin*, 1998.*

[26] *ROCKAFELLAR, R. T., "Monotone operators and the proximal point algorithm",* SIAM J. Control. Optim.*, v. 14, pp. 877–898, 1976.*

[27] *SOLODOV, M.V., SVAITER, B.F., "An inexact hybrid generalized proximal point algorithm and some new results on the theory of bregman functions",* Math Oper Res.*, v. 25, pp. 214–230, 2000.*

[28] *SOUZA, S.S., OLIVEIRA, P.R., CRUZ NETO, J.X., SOUBEYRAN, A., "A proximal method with separable Bregman distances for quasiconvex minimization over the nonnegative orthant",* Eur. J. Oper. Res.*, v. 201, pp. 365–376, 2010.*

[29] *TEBOULLE, M., "Entropic proximal mappings with applications to nonlinear programming",* Math. Oper. Res.*, v. 17, pp. 670–690, 1992.*