



UNIVERSIDADE FEDERAL DO PIAUÍ  
CENTRO DE CIÊNCIAS DA NATUREZA  
PÓS-GRADUAÇÃO EM MATEMÁTICA  
MESTRADO EM MATEMÁTICA

**Método Proximal Incremental em Otimização  
Convexa**

**Valéria de Sousa Silva**

**Teresina - 2018**

**Valéria de Sousa Silva**

**Dissertação de Mestrado:**

**Método Proximal Incremental em Otimização Convexa**

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

Orientador:

Prof. Dr. Jurandir de Oliveira Lopes

**Teresina - 2018**



MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DO PIAUÍ  
CENTRO DE CIÊNCIAS DA NATUREZA  
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

*Método Proximal Incremental em Otimização Convexa*

VALERIA DE SOUSA SILVA

Esta Dissertação foi submetida como parte dos requisitos necessários à obtenção do grau de **Mestre em Matemática**, outorgado pela Universidade Federal do Piauí.

A citação de qualquer trecho deste trabalho é permitida, desde que seja feita em conformidade com as normas da ética científica.

Dissertação aprovada em 20 de Março de 2018.

**Banca Examinadora:**

Jurandir de Oliveira Lopes  
Prof. Dr. Jurandir de Oliveira Lopes - Presidente

João Carlos de Oliveira Souza  
Prof. Dr. João Carlos de Oliveira Souza - Membro interno

Arnaldo Silva Brito  
Prof. Dr. Arnaldo Silva Brito - Membro externo

FICHA CATALOGRÁFICA  
Serviço de Processamento Técnico da Universidade Federal do Piauí  
Biblioteca Setorial do CCN

S586m Silva, Valéria de Sousa.  
Método proximal incremental em otimização convexa /  
Valéria da Sousa Silva. – Teresina, 2018.  
50f.

Dissertação (Mestrado) – Universidade Federal do Piauí,  
Centro de Ciências da Natureza, Pós-Graduação em  
Matemática, 2018.

Orientador: Prof. Dr. Jurandir de Oliveira Lopes.

1. Matemática. 2. Otimização Matemática. 3. Métodos  
Incrementais. I. Título.

CDD 519.3

*Dedico este trabalho à toda minha família, em particular à minha mãe, minha irmã e minha tia Vanusa.*

# Agradecimentos

Agradeço primeiramente à Deus pela força, saúde, corajem, oportunidades... Enfim, por me dar o sustento necessário ao longo das minhas jornadas.

Agradeço à toda minha família, em especial minha mãe, irmã, sobrinho, prima (Vivi) e minha tia Vanusa pelo apoio, carinho e por me motivarem a lutar por meus objetivos.

Agradeço também às minhas amigas Camila (Stoner), que mesmo geograficamente distante se faz bem presente com suas boas “energias” e Élide por me encorajarem, pelo apoio, pelos momentos de descontração. Grata aos meus colegas Felipe e Yldenilson que tanto me ajudaram nessa caminhada.

Agradecimento também à todos os professores que de alguma forma ajudaram e ajudam, desde o ensino médio até o presente momento, seja com palavras de incentivo, na transmissão de conhecimentos... Dentre alguns: Fábio (Barbosinha), Heliailton, Liane, Mário, Newton, em especial, agradeço também aos professores João Carlos e João Xavier e mais ainda ao professor Jurandir por me orientar neste trabalho.

Meus agradecimentos vão também aos colegas e amigos que me ajudaram desde a graduação até agora, seja com incentivo, nos grupos de estudo, esclarecimento de dúvidas...

Dentre alguns: Samara, Deborah, Yara, Victor, Juliana (Ju), Edimilson, Rafaelber, Fernando Lima, Cícero, Arilson, Roniê, Luciano (in memoriam), Marcos, Luan, Fernando Lima, Hércules, Tiago, Ronaldo e Ray.

Agradeço à CAPES pelo apoio financeiro.

*“Vigiai, fiquem firmes na fé, sejam corajosos, sejam fortes. Que tudo o que fizerem seja feito com amor.”.*

1 Coríntios 16:13-14.

# Resumo

Neste trabalho consideramos um método incremental relativamente novo para problemas de otimização convexa em larga escala: método proximal-subgradiente incremental (Bertsekas, 2010). Visando uma melhor compreensão deste método, fazemos uma breve abordagem sobre os métodos subgradiente incremental e proximal incremental (Wajs e Bertsekas, 2003) e provamos uma relação entre suas iterações, conseguindo também uma estimativa útil para a análise de convergência do método em questão. Para esta análise consideramos o método sob a ordem cíclica, usando algumas hipóteses e a noção de quase-Fejér convergência obtemos uma estimativa do número de iterações necessárias para se obter certo nível de otimalidade e em que condições essa convergência se torna exata.

## Palavras-chave

Otimização convexa, Métodos Incrementais, Método Proximal-Subgradiente Incremental.



# Abstract

In this work we consider a relatively new incremental method for large-scale convex optimization problems: incremental proximal-subgradient method (Bertsekas, 2010). Aiming at a better understanding of this method, we make a brief approach to the incremental proximal and incremental subgradient methods (Wajs and Bertsekas, 2003) and prove a relation between their iterations, also obtaining a useful estimate for the convergence analysis of the method in question. For this analysis we consider the method under the cyclic order, using some hypotheses and the notion of quase-Fejér convergence we obtain an estimate of the number of iterations necessary to have a certain level of optimality and what conditions this convergence becomes accurate.

## Keywords

Convex optimization, incremental methods, incremental proximal-subgradient method.

# Sumário

<b>Resumo</b>	<b>iv</b>
<b>Abstract</b>	<b>v</b>
<b>1 Noções Preliminares</b>	<b>4</b>
1.1 Definições e Alguns Resultados de Análise . . . . .	4
1.2 Definições e Alguns Resultados de Otimização . . . . .	6
1.2.1 Operador Projeção . . . . .	13
1.2.2 Operador Proximal . . . . .	15
<b>2 Métodos Incrementais</b>	<b>17</b>
2.1 Método do Subgradiente Incremental . . . . .	18
2.1.1 Método do Subgradiente . . . . .	18
2.1.2 Método do Subgradiente Incremental(SI) . . . . .	18
2.2 Método Proximal Incremental . . . . .	19
2.2.1 Método do Ponto Proximal . . . . .	19
2.2.2 Método Proximal Incremental . . . . .	23
<b>3 Método Proximal-Subgradiente Incremental</b>	<b>26</b>
3.1 Análise de Convergência para o Método com a Ordem Cíclica . . . . .	28
<b>4 Considerações Finais</b>	<b>38</b>
<b>Referências Bibliográficas</b>	<b>39</b>

# Introdução

Neste trabalho, a nossa atenção está concentrada em alguns métodos voltados para o seguinte problema de minimização

$$\min_{\mathbf{x} \in C} \sum_{i=1}^m f_i(\mathbf{x}) \quad (1)$$

onde  $f_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ ,  $i = 1, \dots, m$ , são funções convexas,  $C$  é um conjunto fechado, convexo e não vazio. Quando  $m$ , o número das funções componentes, é grande há um interesse em métodos incrementais, isto é, que usam somente uma função componente a cada iteração. Problemas como em (1) surgem em muitos contextos: Problemas dual, aprendizado de máquina (regularização dos mínimos quadrados), o Problema de Fermat-Weber em teoria da localização, estimativa de máxima verossimilhança, para mais detalhes ver [9], [26], [29]. Abaixo, temos a explanação de alguns desses problemas, que são interessantes à nossa abordagem.

## 1. Regularização dos Mínimos Quadrados

Em muitos problemas de inferência estatística, aprendizado mecânico e sinal de processamento envolvem a minimização da soma de funções componentes  $f_i(\mathbf{x})$ , que corresponde a erros entre dados e a saída de um modelo que é parametrizado por um vetor  $\mathbf{x}$ . Um exemplo seria o problema dos mínimos quadrados, onde  $f_i$  é quadrático. Normalmente uma função de regularização convexa  $R(\mathbf{x})$  é adicionada aos mínimos quadrados objetivo, para induzir algumas propriedades desejáveis na solução. E assim o nosso problema é da seguinte forma

$$\min_{\mathbf{x} \in \mathbb{R}^n} R(\mathbf{x}) + \frac{1}{2} \sum_{i=1}^m (\langle \mathbf{c}_i, \mathbf{x} \rangle - d_i)^2$$

onde  $\mathbf{c}_i$  e  $d_i$  são vetores e escalares, respectivamente.  $R(\mathbf{x})$  pode ter uma forma

diferenciável (quadrática) como no exemplo abaixo:

$$R(x) = \frac{\gamma}{2} \|x\|^2$$

Onde  $\gamma$  é um escalar positivo. Mas, para que possamos usar os métodos incrementais, possuir esta forma não é fundamental. O mais importante é que  $R$  tenha uma forma simples tal qual facilite o uso de algoritmos proximais. Por exemplo,  $R$  poderia ter uma forma separável, de forma que a iteração proximal aplicada em  $R$  seja simplificada através de uma decomposição. Um exemplo é no problema de  $\ell_1$  - regularização, onde  $R$  é da seguinte forma

$$R(x) = \gamma \|x\|_1 = \gamma \sum_{j=1}^n |x^j|.$$

$\gamma$  é um escalar positivo e  $x^j$  é a  $j$ -ésima coordenada de  $x$ . Para mais detalhes veja [5].

## 2. Problema de Weber em Teoria da Localização

Queremos encontrar um ponto  $x$  no plano, cuja a soma das distância ponderadas de um dado conjunto de pontos  $y_1, \dots, y_m$  é minimizada. O nosso problema tem a seguinte forma

$$\min_{x \in \mathbb{R}^n} f(x) = \sum_{i=1}^m w_i \|x - y_i\|,$$

onde  $w_1, \dots, w_m$  são escalares positivos dados. Para mais detalhes a respeito deste problema, ver [29].

## 3. Problema de Viabilidade

Considere a seguinte generalização do problema de viabilidade

$$\min_{x \in \bigcap_{i=1}^m C_i} f(x),$$

onde  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$  é uma função convexa e  $C_i$ , onde  $i = 1, \dots, m$ , são conjuntos convexos fechados. O problema acima pode ser convertido em um outro problema, tal que envolva uma soma de funções componentes. E dessa maneira, se temos  $f$  Lipschitz contínua e  $\gamma$  suficientemente grande, o problema acima pode ser dado da seguinte forma

$$\min_{x \in \mathbb{R}^n} f(x) + \gamma \sum_{i=1}^m \text{dist}(x; C_i).$$

onde  $\gamma > 0$  é uma parametro de penalidade, veja [5].

Podemos ver na literatura, a utilização do método do subgradiente e do método proximal, em suas versões incrementais, para problemas como (1). Neste trabalho, abordaremos um método que consiste na combinação dos métodos incrementais citados acima, conhecido como método proximal-subgradiente incremental e visando a compreensão deste método, organizaremos esta dissertação da seguinte forma, no Capítulo 1, apresentamos algumas definições e resultados de Análise no  $\mathbb{R}^n$  e Otimização que serão utilizadas no decorrer deste trabalho.

No Capítulo 2, faremos uma breve abordagem sobre os métodos do subgradiente (gradiente) e proximal incremental, vendo como operam seus respectivos algoritmos e usando algumas propriedades conseguiremos uma relação entre suas iterações e um resultado que será usado no capítulo seguinte.

No Capítulo 3, usando alguns resultados do capítulo anterior, faremos uma análise de convergência do método sob a ordem cíclica. Dessa forma, sob certas condições, obtaremos uma estimativa do número de iterações necessárias para obter um certo grau de otimalidade, e que em que condições essa convergência se torna exata.

Finalizando, apresentaremos as conclusões finais e as referências bibliográficas

# Capítulo 1

## Noções Preliminares

Neste capítulo, apresentamos algumas definições e resultados necessários para o desenvolvimento dos demais capítulos, que podem ser encontrados, por exemplo, em [1], [2], [7], [8], [10], [11], [12], [13], [14], [16], [17], [18], [21], [27], [28].

### 1.1 Definições e Alguns Resultados de Análise

**Definição 1.** A norma (euclidiana) de um vetor  $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$  é, por definição, o número real  $\|\mathbf{x}\|$ , dado por

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} .$$

A distância (euclidiana) entre  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ,  $d(\mathbf{x}, \mathbf{y})$ , é definida por

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|.$$

Considerando-se em  $\mathbb{R}^n$  um sistema cartesiano de coordenadas, tem-se que geometricamente, a distância entre dois pontos corresponde ao comprimento do segmento de reta que os une. Em particular, para todo  $\mathbf{x} \in \mathbb{R}^n$ ,  $\|\mathbf{x}\| = d(\mathbf{x}, 0)$  é o comprimento do segmento cujos extremos são a *origem* e o ponto  $\mathbf{x}$ . Dados  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  e  $\lambda \in \mathbb{R}$ , verificam-se as seguintes propriedades:

1.  $\|\mathbf{x}\| \geq 0$  e  $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = 0$ ;
2.  $\|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\|$ ;
3.  $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ .

**Teorema 1.** (*Desigualdade de Cauchy-Schwarz*). Dados  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ , tem-se

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|,$$

valendo a igualdade se, e somente se, um dos vetores  $\mathbf{x}, \mathbf{y}$  é múltiplo do outro.

*Demonstração:* O resultado é imediato para  $\mathbf{x} = 0$  ou  $\mathbf{y} = 0$ . Supondo-se, então,  $\mathbf{x}, \mathbf{y} \neq 0$ , definimos a função  $f(t) = \langle \mathbf{x} - t\mathbf{y}, \mathbf{x} - t\mathbf{y} \rangle$ ,  $t \in \mathbb{R}$ . Note que  $f$  é não negativa e que  $f$  possui um zero se, e somente se,  $\mathbf{x}$  é múltiplo de  $\mathbf{y}$ . Além disso,  $f(t) = \langle \mathbf{x} - t\mathbf{y}, \mathbf{x} - t\mathbf{y} \rangle = \|\mathbf{y}\|^2 t^2 - 2\langle \mathbf{x}, \mathbf{y} \rangle t + \|\mathbf{x}\|^2$ , isto é,  $f$  é uma função quadrática não negativa de  $t$ . Logo, seu discriminante deve ser não positivo e portanto,

$$4\langle \mathbf{x}, \mathbf{y} \rangle^2 - 4\|\mathbf{x}\|^2 \|\mathbf{y}\|^2 \leq 0,$$

em que vale a igualdade se, e somente se,  $f$  possui um único zero.  $\diamond$

Agora veremos algumas propriedades e definições a respeito de supremo e ínfimo que serão úteis nos próximos capítulos.

**Definição 2.** Seja  $X \subset \mathbb{R}$  um subconjunto limitado superiormente. Um elemento  $\mathbf{b} \in \mathbb{R}$  é dito supremo de  $X$  se satisfaz as seguintes condições:

S1. Para todo  $x \in X$ , tem-se  $x \leq \mathbf{b}$ ;

S2. Dado  $c < \mathbf{b}$  em  $\mathbb{R}$  existe  $x \in X$  tal que  $c < x$ .

Analogamente, um elemento  $\mathbf{a} \in \mathbb{R}$  chama-se ínfimo de um conjunto  $Y \subset \mathbb{R}$ , limitado inferiormente, quando  $\mathbf{a}$  satisfaz as condições seguintes:

I1. Para todo  $y \in Y$ , tem-se  $\mathbf{a} \leq y$ ;

I2. Dado  $c \in \mathbb{R}$  com  $\mathbf{a} < c$ , existe  $y \in Y$  tal que  $y < c$ .

O supremo e ínfimo, quando existem são únicos e serão indicados, respectivamente por:  $\mathbf{b} = \sup X$  e  $\mathbf{a} = \inf Y$ .

**Definição 3.** Seja  $(x_n)$  uma seqüência limitada de números reais, digamos com  $\alpha \leq x_n \leq \beta$  para todo  $n \in \mathbb{N}$ . Escrevamos  $X_n = \{x_n, x_{n+1}, \dots\}$ . Temos  $[\alpha, \beta] \supset X_1 \supset X_2 \supset \dots \supset X_n \supset \dots$ . Logo, pondo  $\mathbf{a}_n = \inf X_n$  e  $\mathbf{b}_n = \sup X_n$ , vem

$$\alpha \leq \mathbf{a}_1 \leq \mathbf{a}_2 \leq \dots \leq \mathbf{a}_n \leq \dots \leq \mathbf{b}_n \leq \dots \leq \mathbf{b}_2 \leq \mathbf{b}_1 \leq \beta.$$

Existem, portanto, os limites

$$a = \lim a_n = \sup a_n = \sup_n \inf X_n,$$

$$b = \lim b_n = \inf b_n = \inf_n \sup X_n.$$

Escrevemos  $a = \liminf x_n$ ,  $b = \limsup x_n$ , diremos que  $a$  é o limite inferior e que  $b$  é o limite superior da sequência  $(x_n)$ . Tem-se evidentemente

$$\liminf x_n \leq \limsup x_n.$$

**Teorema 2. (Bolzano-Weierstrass).** *Toda sequência limitada em  $\mathbb{R}^n$  possui uma subsequência convergente.*

*Demonstração:* Ver [ 17 ], pág. 17. ◊

## 1.2 Definições e Alguns Resultados de Otimização

Dados  $C \subset \mathbb{R}^n$  e uma função  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ , o problema de minimizar  $f$  em  $C$  é escrito como

$$\min_{x \in C} f(x). \tag{1.1}$$

O conjunto  $C$  é chamado conjunto viável do problema, os pontos de  $C$  são chamados de pontos viáveis e  $f$  é chamada função objetivo. Quando  $C \neq \mathbb{R}^n$  dizemos que (1.1) é um problema restrito e quando  $C = \mathbb{R}^n$  dizemos que o problema é irrestrito.

**Definição 4.** *Dizemos que  $\bar{v} \in [-\infty, +\infty)$  definido por*

$$\bar{v} = \inf_{x \in C} f(x)$$

*é o valor ótimo de (1.1).*

**Definição 5.** *Considere ainda o problema (1.1). Dizemos que um ponto  $\bar{x}$  é*

a) *Minimizador global de  $f$  sobre  $C$ , se*

$$\bar{x} \in C \cap \text{dom}(f) \quad e \quad f(\bar{x}) = \inf_{x \in C} f(x);$$

b) *Minimizador local de  $f$  sobre  $C$ , se  $\bar{x} \in C \cap \text{dom}(f)$  e existe uma vizinhança  $V$  de  $\bar{x}$  tal que*

$$f(\bar{x}) \leq f(x), \quad \forall x \in C \cap V.$$



**Definição 6.** O domínio efetivo de uma função  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ , denotado por  $\text{dom}(f)$ , é o conjunto dado por  $\text{dom}(f) = \{\mathbf{x} \in \mathbb{R}^n; f(\mathbf{x}) < +\infty\}$ .

**Definição 7.** Uma função  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  é própria quando  $\text{dom}(f) \neq \emptyset$ .

**Definição 8.** Se  $\text{dom}(f) \neq \emptyset$ , então o conjunto dos seus minimizadores globais é dado por

$$\arg \min f = \left\{ \bar{\mathbf{x}} \in \mathbb{R}^n \mid f(\bar{\mathbf{x}}) = \inf_{\mathbf{x}} f(\mathbf{x}) \right\}.$$

**Exemplo 1.** Considere a seguinte função

$$f(\mathbf{x}) = \begin{cases} -\ln(\mathbf{x}), & \text{se } \mathbf{x} > 0, \\ +\infty, & \text{se } \mathbf{x} \leq 0. \end{cases}$$

Claramente,  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  é uma função própria.

A seguir, veremos as definições e proposições de sequências Fejér e quase-Fejér convergente, muito relevantes para a obtenção de alguns resultados nos capítulos posteriores. Sendo um destes, relacionados com a análise de convergência do método Proximal-Subgradiente incremental.

**Definição 9.** Uma sequência  $\{\mathbf{x}_k\} \subset \mathbb{R}^n$  é dita ser Fejér convergente a um conjunto não vazio  $\mathbf{U} \subset \mathbb{R}^n$  se para cada  $\mathbf{u} \in \mathbf{U}$  e para todo  $k \geq 0$  vale

$$\|\mathbf{x}_{k+1} - \mathbf{u}\|^2 \leq \|\mathbf{x}_k - \mathbf{u}\|^2.$$

**Definição 10.** Uma sequência  $\{\mathbf{x}_k\} \subset \mathbb{R}^n$  é quase-Fejér convergente ao conjunto não vazio  $\mathbf{U} \subset \mathbb{R}^n$  se para cada  $\mathbf{u} \in \mathbf{U}$  existe uma sequência  $\{\epsilon_k\} \subset \mathbb{R}$ , com  $\epsilon_k \geq 0$ , tal que  $\sum_{k=0}^{\infty} \epsilon_k < \infty$  e para todo  $k \geq 0$  vale

$$\|\mathbf{x}_{k+1} - \mathbf{u}\|^2 \leq \|\mathbf{x}_k - \mathbf{u}\|^2 + \epsilon_k.$$

**Proposição 1.** Se  $\{\mathbf{x}_k\} \subset \mathbb{R}^n$  é quase-Fejér convergente ao conjunto não vazio  $\mathbf{U} \subset \mathbb{R}^n$  então  $\{\mathbf{x}_k\}$  é limitada. E mais, se existir algum ponto de acumulação  $\bar{\mathbf{x}}$  de  $\mathbf{x}_k$  tal que  $\bar{\mathbf{x}} \in \mathbf{U}$ , então  $\lim \mathbf{x}_k = \bar{\mathbf{x}}$ .

*Demonstração:* Seja  $\mathbf{u} \in \mathbf{U}$ . Pela Definição 10 temos que

$$\|\mathbf{x}_k - \mathbf{u}\|^2 \leq \|\mathbf{x}_0 - \mathbf{u}\|^2 + \sum_{j=0}^{k-1} \epsilon_j \leq \|\mathbf{x}_0 - \mathbf{u}\|^2 + \sum_{j=0}^{\infty} \epsilon_j.$$

Assim temos que  $\{x_k\}$  é limitada. Agora, consideremos  $\bar{x} \in U$  um ponto de acumulação de  $\{x_k\}$  e  $\delta > 0$ . Seja  $\{x_{k_1}\}$  uma subsequência de  $\{x_k\}$ , tal que  $\lim x_{k_1} = \bar{x}$ . Pela mesma Definição, existe  $k_0$  tal que  $\sum_{j=k_0}^{\infty} \epsilon_j < \frac{\delta}{2}$  e existe  $k_1 \geq k_0$  tal que

$$\|x_{k_1} - \bar{x}\|^2 < \frac{\delta}{2} \text{ para todo } k \geq k_1.$$

Então para todo  $k \geq k_1$  vale

$$\|x_k - \bar{x}\|^2 \leq \|x_{k_1} - \bar{x}\|^2 + \sum_{j=k_1}^{k-1} \epsilon_j \leq \|x_{k_1} - \bar{x}\|^2 + \sum_{j=k_1}^{\infty} \epsilon_j < \frac{\delta}{2} + \frac{\delta}{2} = \delta$$

Dessa forma, temos que  $\lim x_k = \bar{x}$ . ◊

Claramente podemos ver que proposição acima é válida para o caso onde  $\{x_k\}$  é Fejér convergente ao conjunto  $U$ , basta considerar  $\epsilon_k = 0$ .

**Definição 11.** Um conjunto  $C \subset \mathbb{R}^n$  é chamado conjunto convexo se para quaisquer  $x, y \in C$  e  $\alpha \in [0, 1]$ , tem-se  $\alpha x + (1 - \alpha)y \in C$ .

**Exemplo 2.** O conjunto vazio, o espaço  $\mathbb{R}^n$  e o conjunto que contém um ponto só, são trivialmente convexos.

**Exemplo 3.** Uma bola em  $\mathbb{R}^n$ , isto é, um conjunto  $\{x \in \mathbb{R}^n \text{ tal que } \|x\| \leq c\}$ , onde  $c \in \mathbb{R}_+$ , também é um conjunto convexo.

**Definição 12.** a) Uma combinação afim de elementos  $x_1, \dots, x_k \in \mathbb{R}^n$  é um elemento da forma  $\sum_{i=1}^k \alpha_i x_i$ , onde os coeficientes satisfazem  $\sum_{i=1}^k \alpha_i = 1$ .

b) Uma variedade afim em  $\mathbb{R}^n$  é um conjunto contendo todas suas combinações afins.

c) Seja  $S \subset \mathbb{R}^n$  um conjunto não vazio. O fecho afim de  $S$  é denotado por  $\text{aff}(S)$  é a variedade afim gerada por  $S$ , isto é, a interseção de todas as variedades afins que contem  $S$ .

**Definição 13.** Seja  $C \subset \mathbb{R}^n$  um conjunto convexo. O interior relativo de  $C$ , denotado por  $\text{ri}(C)$  é o interior de  $C$  com relação a topologia relativa do fecho afim de  $S$ , isto é,  $x \in \text{ri}(C)$  se, e somente se,  $x \in \text{aff}(C)$  e existe  $\delta > 0$  tal que  $\text{aff}(C) \cap B(x, \delta) \subset C$ .

**Definição 14.** Sejam  $C \subset \mathbb{R}^n$  um conjunto convexo e  $\bar{x} \in C$ . O cone normal (cone de direções normais), no ponto  $\bar{x}$  em relação ao conjunto  $C$  é dado por

$$N_C(\bar{x}) = \{d \in \mathbb{R}^n \mid \langle d, x - \bar{x} \rangle \leq 0, \forall x \in C\}.$$

**Observação 1.** Se  $\bar{x} \in \text{int } C$ , então  $N_C(\bar{x}) = 0$ .

Com efeito, veja que se  $\bar{x} \in \text{int } C$ , então existe  $\delta > 0$  tal que,  $B(\bar{x}, \delta) \subset \mathbb{R}^n$  e pela definição de cone isto só é possível se  $d = 0$ .

**Observação 2.** Se  $C = \mathbb{R}^n$  então  $N_C(x) = 0, \forall x \in \mathbb{R}^n$ .

**Definição 15.** Seja  $C \subset \mathbb{R}^n$  um conjunto convexo, diz-se que a função  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  própria, é convexa em  $C$  quando para quaisquer  $x, y \in C$  e  $\alpha \in (0, 1)$ , tem-se

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

e a função  $f$  é dita estritamente convexa em  $C$ , quando a desigualdade acima é estrita para todos  $x \neq y$ , com  $x, y \in \text{dom}(f)$ . A função diz-se fortemente convexa em  $C$ , com módulo  $\gamma > 0$ , quando para quaisquer  $x, y \in C$  e  $\alpha \in (0, 1)$ , tem-se

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

Claramente vemos que uma função fortemente convexa em  $C$  é estritamente convexa em  $C$ , e uma estritamente convexa em  $C$  é convexa em  $C$ .

**Exemplo 4.** A função  $f : C \rightarrow \mathbb{R}$ , onde  $C \subseteq \mathbb{R}^n$ , dada por  $f(x) = \|x - x_0\|^2$ , para  $x_0 \in \mathbb{R}^n$  é fortemente convexa.

**Exemplo 5.** A função  $f : \mathbb{R} \rightarrow \mathbb{R}$ , dada por  $f(x) = e^x$  é estritamente convexa (mas não é fortemente convexa).

**Exemplo 6.** A função  $f : \mathbb{R} \rightarrow \mathbb{R}$ , dada por  $f(x) = x$  é convexa (mas não é estritamente convexa).

**Teorema 3.** Sejam  $C \subset \mathbb{R}^n$  um conjunto convexo e  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  uma função convexa e própria, então um mínimo local de  $f$  sobre  $C$  é também um mínimo global. E mais, se  $f$  é estritamente convexa, então existe no máximo um mínimo global de  $f$  sobre  $C$ .

*Demonstração:* Ver [6], página 87. ◇

**Teorema 4.** Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  uma função convexa. Então  $f$  é relativamente contínua em qualquer conjunto  $C$  relativamente aberto e convexo em  $\text{dom}(f)$ , em particular em  $\text{ri}(\text{dom}(f))$ .

*Demonstração:* Ver [21], página 82. ◇

**Definição 16.** *Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ , própria e  $C \subset \mathbb{R}^n$ . Dizemos que  $f$  é Lipschitz contínua sobre  $C$  se  $C \subset \text{dom}(f)$  e existe uma constante  $L > 0$  tal que*

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq L \|\mathbf{x} - \mathbf{y}\|, \quad \forall \mathbf{x}, \mathbf{y} \in C.$$

**Teorema 5.** *Seja  $f : C \rightarrow \mathbb{R} \cup \{+\infty\}$  uma função própria e convexa em  $C$  qualquer subconjunto compacto em  $\text{ri}(\text{dom}(f))$ , então  $f$  é Lipschitz-contínua em  $C$ .*

*Demonstração:* Ver [21], página 86. ◊

**Proposição 2.** *Seja  $C \subset \mathbb{R}^n$  um conjunto convexo. Se  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  é uma função convexa em  $C$  e  $h : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  é fortemente convexa em  $C$ , então  $f + h$  é fortemente convexa em  $C$ .*

*Demonstração:* Sejam  $\mathbf{x}, \mathbf{y} \in C$ , então

$$f(\alpha \mathbf{x} + (1 - \alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y}), \quad \text{se } \alpha \in [0, 1]$$

e

$$h(\alpha \mathbf{x} + (1 - \alpha)\mathbf{y}) \leq \alpha h(\mathbf{x}) + (1 - \alpha)h(\mathbf{y}) - \gamma \alpha (1 - \alpha) \|\mathbf{x} - \mathbf{y}\|^2, \quad \text{se } \alpha \in [0, 1],$$

então temos

$$\begin{aligned} (f + h)(\alpha \mathbf{x} + (1 - \alpha)\mathbf{y}) &= f(\alpha \mathbf{x} + (1 - \alpha)\mathbf{y}) + h(\alpha \mathbf{x} + (1 - \alpha)\mathbf{y}) \\ &\leq \alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y}) + \alpha h(\mathbf{x}) + (1 - \alpha)h(\mathbf{y}) - \gamma \alpha (1 - \alpha) \|\mathbf{x} - \mathbf{y}\|^2 \\ &= \alpha (f + h)(\mathbf{x}) + (1 - \alpha)(f + h)(\mathbf{y}) - \gamma \alpha (1 - \alpha) \|\mathbf{x} - \mathbf{y}\|^2. \end{aligned}$$

Logo,  $f + h$  é fortemente convexa em  $C$ . ◊

**Definição 17.** *Dizemos que a função  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  é semi-contínua inferiormente no ponto  $\mathbf{x} \in \text{dom}(f)$ , quando para qualquer sequência  $\{\mathbf{x}_k\} \subset \text{dom}(f)$  tal que  $\{\mathbf{x}_k\} \rightarrow \mathbf{x}$  ( $k \rightarrow \infty$ ), tem-se*

$$\liminf f(\mathbf{x}_k) \geq f(\mathbf{x}). \tag{1.2}$$

**Observação 3.** *Nas mesmas condições acima, dizemos que  $f$  é fechada quando a desigualdade (1.2) é válida para todo  $\mathbf{x} \in \mathbb{R}^n$ .*

**Definição 18.** *Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  uma função convexa. Dizemos que  $\mathbf{y} \in \mathbb{R}^n$  é um subgradiente de  $f$  no ponto  $\mathbf{x} \in \text{dom}(f)$  se*

$$f(\mathbf{z}) \geq f(\mathbf{x}) + \langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle, \quad \forall \mathbf{z} \in \mathbb{R}^n.$$

O conjunto de todos os subgradientes de  $f$  em  $\mathbf{x}$  é chamado subdiferencial de  $f$  em  $\mathbf{x}$  e sua notação é dada por  $\partial f(\mathbf{x})$ . Se  $f$  não é finito em  $\mathbf{x}$ , então definimos  $\partial f(\mathbf{x}) = \emptyset$ .

**Teorema 6.** *Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  uma função convexa e própria. Então para todo  $x \in \text{int}(\text{dom}(f))$ , o conjunto  $\partial f(x)$  é convexo, compacto e não vazio.*

*Demonstração:* Ver [21], página 217. ◇

**Teorema 7.** *Sejam  $f_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ ,  $i = 1, \dots, m$  funções convexas, próprias e assumamos que  $F = f_1 + \dots + f_m$  é própria. Então temos que:*

a)  $\partial F(x) \supset \partial f_1(x) + \dots + \partial f_m(x), \quad \forall x \in \mathbb{R}^n.$

b) Se  $\bigcap_{k=1}^m \text{ri}(\text{dom}(f_k)) \neq \emptyset$  temos que

$$\partial F(x) = \partial f_1(x) + \dots + \partial f_m(x), \quad \forall x \in \mathbb{R}^n.$$

*Demonstração:* Ver [21], página 223. ◇

**Definição 19.** *Para  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ , o epígrafo de  $f$  é o conjunto dado por*

$$E_f = \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq \alpha\}.$$

**Teorema 8.** *Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ , então  $f$  é fechada no  $\mathbb{R}^n$  se, e somente se, o epígrafo  $E_f$  é fechado em  $\mathbb{R}^n \times \mathbb{R}$ .*

*Demonstração:* Ver [28], página 9. ◇

**Teorema 9.** *Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ , então  $f$  é convexa no  $\mathbb{R}^n$  se, e somente se, o epígrafo  $E_f$  é um conjunto convexo em  $\mathbb{R}^n \times \mathbb{R}$ .*

*Demonstração:* Ver [28], página 40. ◇

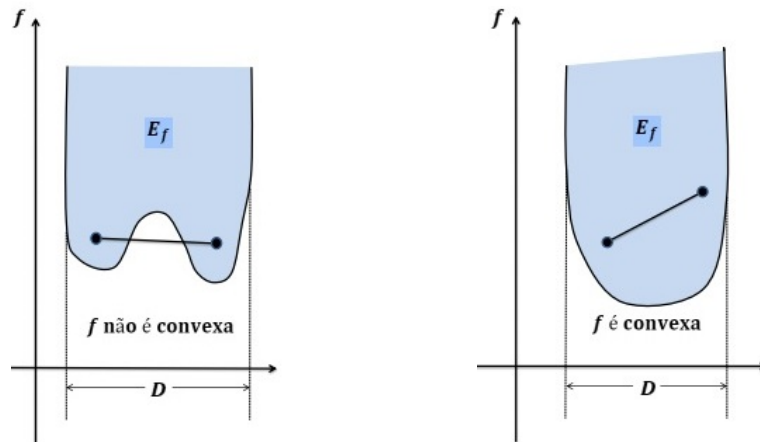


Figura 1.1: Convexidade da função  $f \Leftrightarrow$  convexidade do epígrafo de  $f$ .

**Definição 20.** Dado um subconjunto não vazio  $S \subset \mathbb{R}^n$ , a função  $I_S : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  definida por

$$I_S(x) = \begin{cases} 0, & \text{se } x \in S, \\ +\infty & \text{se } x \notin S \end{cases}$$

é chamada função indicadora de  $S$ . Se  $S$  é um conjunto convexo, fechado e não vazio então  $I_S$  é uma função convexa, fechada e própria.

**Observação 4.** Sejam  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  uma função convexa e  $C \subset \mathbb{R}^n$  um conjunto não vazio e convexo. Se  $\text{dom}(f) \cap C \neq \emptyset$  então  $\phi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ , definida por

$$\phi : \begin{cases} f(x), & \text{se } x \in C, \\ +\infty & \text{se } x \notin C \end{cases}$$

é uma função convexa e própria. Observe também que  $\phi = f + I_C$ . Assim, resolver o problema

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

é equivalente à resolver

$$\min_{x \in \mathbb{R}^n} \phi(x).$$

Na seguinte proposição observamos uma relação entre o operador normalizador e a função indicadora de um conjunto.

**Proposição 3.** O operador normalizador do conjunto  $S \subset \mathbb{R}^n$  ( $N_S(x)$ ) é definido por  $N_S(x) = \partial I_S(x)$ ,  $\forall x \in \mathbb{R}^n$ .

*Demonstração:* Seja  $y \in \partial I_S(x)$ , então:

a) Se  $x \notin S$  temos que  $I_S(x) = +\infty$  então, pela definição de subgradiente teríamos que

$$I_S(z) \geq \langle y, z - x \rangle, \quad \forall z \in \mathbb{R}^n,$$

o que é um absurdo, pois para  $y \in S$  teríamos:  $0 \geq +\infty$ . Logo  $\partial I_S(x) = \emptyset$ .

b) Se  $x \in S$  então, temos que  $I_S(x) = 0$ , daí

$$I_S(z) \geq \langle y, z - x \rangle, \quad \forall z \in \mathbb{R}^n.$$

Assim, temos dois casos:

i) Se  $z \notin S$  então,  $\langle \mathbf{y}, z - \mathbf{x} \rangle \leq +\infty$ .

ii) Se  $z \in S$  então,  $\langle \mathbf{y}, z - \mathbf{x} \rangle \leq 0$

Dessa forma, temos que:

$$\partial I_S(\mathbf{x}) = \begin{cases} \{\mathbf{y} \in \mathbb{R}^n; \langle \mathbf{y}, z - \mathbf{x} \rangle \leq 0, \forall z \in S\}, & \text{se } \mathbf{x} \in S, \\ \emptyset, & \text{se } \mathbf{x} \notin S. \end{cases}$$

E pela definição de subgradiente e normalizador, temos que o resultado é verdadeiro.  $\diamond$

**Teorema 10. (Condição de otimalidade para minimização convexa num conjunto convexo)** Sejam  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  uma função convexa, própria e fechada, e  $C \in \mathbb{R}^n$  um conjunto convexo, não vazio e fechado. Assuma que  $\text{ri}(C) \cap \text{ri}(\text{dom}(f)) \neq \emptyset$ . Então  $\bar{\mathbf{x}} \in \mathbb{R}^n$  é um minimizador de  $f$  em  $C$  se, e somente se,

$$\exists \mathbf{y} \in \partial f(\bar{\mathbf{x}}) \text{ tal que } \langle \mathbf{y}, \mathbf{x} - \bar{\mathbf{x}} \rangle \geq 0, \forall \mathbf{x} \in C,$$

ou, equivalentemente,

$$0 \in \partial f(\bar{\mathbf{x}}) + N_C(\bar{\mathbf{x}}). \tag{1.3}$$

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

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

*Demonstração:* Ver [1], página 120.  $\diamond$

**Observação 5.** Veja ainda que com a definição e propriedades de função indicadora, (1.3) é equivalente à

$$0 \in \partial (f + I_C)(\bar{\mathbf{x}}).$$

### 1.2.1 Operador Projeção

Uma projeção do ponto  $\mathbf{x} \in \mathbb{R}^n$  sobre um conjunto não vazio  $C \subset \mathbb{R}^n$  é o ponto de  $C$  que está mais próximo de  $\mathbf{x}$  (a distância aqui considerada é proveniente da norma euclidiana).

Assim temos que, a projeção de  $\mathbf{x}$  sobre  $C$  é uma solução (global) do problema

$$\min_{\mathbf{y} \in C} \|\mathbf{y} - \mathbf{x}\|.$$

E ainda, quando  $C$  é fechado esta projeção existe para qualquer  $\mathbf{x} \in \mathbb{R}^n$ . O teorema a seguir nos garante a unicidade da projeção (sob algumas condições) e o subseqüente nos dará uma propriedade que será usada no Capítulo 3.

**Teorema 11. (Teorema da projeção)** *Seja  $C \in \mathbb{R}^n$  um conjunto convexo, fechado e não vazio. Então para todo  $x \in \mathbb{R}^n$ , a projeção de  $x$  sobre  $C$ , denotada por  $P_C(x)$ , existe e é única. Além disso,  $\bar{x} = P_C(x)$  se, e somente se,*

$$\bar{x} \in C, \quad \langle x - \bar{x}, y - \bar{x} \rangle \leq 0, \forall y \in C, \quad (1.4)$$

ou, equivalentemente,

$$\bar{x} \in C, \quad x - \bar{x} \in N_C(\bar{x}).$$

*Demonstração:* Ver [13], página 94. ◇

**Observação 6.** *Veja que se  $C = \mathbb{R}^n$ , então  $P_C(x) = x, \forall x \in \mathbb{R}^n$ .*

**Teorema 12. (O operador projeção é Lipschitz)** *Seja  $C \subset \mathbb{R}^n$  um conjunto convexo, fechado e não vazio. Então para  $x, y \in \mathbb{R}^n$  quaisquer, temos:*

$$\|P_C(x) - P_C(y)\| \leq \|x - y\|$$

*Demonstração:* Temos que  $P_C(x) \in C$  e  $P_C(y) \in C$  e usando (1.4) para  $x$  e  $y$ , respectivamente, temos

$$\langle x - P_C(x), P_C(y) - P_C(x) \rangle \leq 0,$$

$$\langle y - P_C(y), P_C(x) - P_C(y) \rangle \leq 0.$$

*Somando as desigualdades acima obtemos*

$$\begin{aligned} 0 &\geq \langle y - P_C(y) - x + P_C(x), P_C(x) - P_C(y) \rangle \\ &= \|P_C(x) - P_C(y)\|^2 + \langle y - x, P_C(x) - P_C(y) \rangle, \end{aligned}$$

*e usando a desigualdade de Cauchy-Schwarz na expressão acima temos*

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

*Se  $P_C(x) = P_C(y)$  o resultado é claro. Caso contrário, dividimos os dois lados da desigualdade acima por  $\|P_C(x) - P_C(y)\|$  e obtemos a desigualdade desejada.* ◇



### 1.2.2 Operador Proximal

O operador proximal é o elemento chave nos algoritmos de otimização convexa, por exemplo, calcular o operador proximal de uma função é a operação base do método proximal. A seguir veremos algumas propriedades deste operador.

**Definição 21.** *Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ . Dado um parâmetro positivo  $\alpha$ , o operador proximal de uma função  $f$  em um ponto  $\mathbf{u} \in \mathbb{R}^n$  é definido por*

$$\text{prox}_f(\mathbf{u}) := \arg \min_{\mathbf{x} \in \mathbb{R}^n} \left( f(\mathbf{x}) + \frac{1}{2\alpha} \|\mathbf{x} - \mathbf{u}\|^2 \right)$$

**Observação 7.** *Para  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  própria, fechada e convexa e  $\mathbf{u}, \mathbf{p} \in \mathbb{R}^n$  temos que*

$$\mathbf{p} = \text{prox}_f(\mathbf{u}) \text{ se, e somente se } \mathbf{u} - \mathbf{p} \in \partial f(\mathbf{p}).$$

*Com efeito, veja que pela definição acima e pelo Teorema 10 temos que*

$$\mathbf{p} = \text{prox}_f(\mathbf{u}) \Leftrightarrow 0 \in \partial \left( f + \frac{1}{2} \|\cdot - \mathbf{u}\|^2 \right)(\mathbf{p}).$$

*Veja que,*

$$\partial \left( f + \frac{1}{2} \|\cdot - \mathbf{u}\|^2 \right)(\mathbf{p}) = \partial f(\mathbf{p}) + \partial \left( \frac{1}{2} \|\mathbf{p} - \mathbf{u}\|^2 \right) = \partial f(\mathbf{p}) + \mathbf{p} - \mathbf{u}.$$

*Assim temos que,*

$$0 \in \partial f(\mathbf{p}) + \mathbf{p} - \mathbf{u} \Leftrightarrow \mathbf{u} - \mathbf{p} \in \partial f(\mathbf{p}).$$

**Definição 22.**  $T : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^n)$ , onde  $\mathcal{P}(\mathbb{R}^n)$  é o conjunto das partes de  $\mathbb{R}^n$ , é dito monótono se

$$0 \leq \langle \mathbf{x} - \mathbf{y}, \mathbf{u} - \mathbf{v} \rangle, \tag{1.5}$$

*para todo  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ,  $\mathbf{u} \in T(\mathbf{x})$  e  $\mathbf{v} \in T(\mathbf{y})$ .*

**Proposição 4.** *Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  uma função convexa, então  $\partial f$  é monótono.*

*Demonstração:* Ver [27], página 110. ◊

**Definição 23.** *Seja  $D \subset \mathbb{R}^n$ . Dizemos que um operador  $T : D \rightarrow \mathbb{R}^n$  é firmemente não expansivo se*

$$\|T(\mathbf{x}) - T(\mathbf{y})\|^2 \leq \langle T(\mathbf{x}) - T(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle, \quad \forall \mathbf{x}, \mathbf{y} \in D.$$

**Observação 8.** Para  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  é própria, fechada e convexa temos que o operador

$$\text{prox}_f : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

é firmemente não expansivo.

Com efeito, veja que  $\partial f$  é um operador monótono, dessarte temos que

$$\begin{aligned} \mathbf{y} = \text{prox}_f(\mathbf{x}), \mathbf{v} = \text{prox}_f(\mathbf{u}) &\Rightarrow \mathbf{x} - \mathbf{y} \in \partial f(\mathbf{y}), \mathbf{u} - \mathbf{v} \in \partial f(\mathbf{v}) \\ &\Rightarrow 0 \leq \langle \mathbf{x} - \mathbf{y} - \mathbf{u} + \mathbf{v}, \mathbf{y} - \mathbf{v} \rangle \\ &\Rightarrow \|\mathbf{y} - \mathbf{v}\|^2 \leq \langle \mathbf{x} - \mathbf{u}, \mathbf{y} - \mathbf{v} \rangle. \end{aligned}$$

Veja ainda que, o operador  $\text{prox}_f$  é Lipschitz-contínuo, basta usar a Desigualdade de Cauchy-Schwarz para última desigualdade acima. E assim, temos

$$\|\mathbf{y} - \mathbf{v}\|^2 \leq \|\mathbf{x} - \mathbf{u}\| \|\mathbf{y} - \mathbf{v}\|.$$

Dividindo a equação acima por  $\|\mathbf{y} - \mathbf{v}\|$  temos a validade desta última observação.

# Capítulo 2

## Métodos Incrementais

Neste capítulo, fazemos uma breve abordagem sobre alguns métodos usados em problemas de otimização convexa em grande escala do tipo

$$\min_{\mathbf{x} \in \mathcal{C}} \sum_{i=1}^m f_i(\mathbf{x}) \quad (2.1)$$

onde  $f_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ ,  $i = 1, \dots, m$ , são funções convexas (não necessariamente diferenciáveis) e  $\mathcal{C}$  é um conjunto fechado e convexo. Quando o número  $m$  de componentes é grande é preferível usar métodos incrementais, caracterizados por atuar em cada função  $f_i$  a cada iteração. Os métodos incrementais tem se tornado preferíveis em problemas como em (2.1), por causa de sua rápida taxa de convergência inicial. Tais métodos surgiram de problemas oriundos de treinamento de redes neurais, regularização dos mínimos quadrados, problema de Fermat-Weber em teoria da localização, problema dual, estimativa de máxima verossimilhança, etc.

É importante ressaltar que, para a análise de convergência, neste trabalho a ordem em que as componentes  $\{f_i, \mathbf{h}_i\}$  serão escolhidas para as iterações, será a cíclica. Na ordem cíclica, as componentes  $\{f_i, \mathbf{h}_i\}$  são tomadas na ordem determinista fixa  $1, \dots, m$ , de modo que

$$i_k = (k \text{ módulo } m) + 1,$$

ou seja,  $i_k$  equivalente à 1 mais o resto da divisão de  $k$  por  $m$ . Um bloco contínuo de iterações envolvendo  $\{f_1, \mathbf{h}_1\}, \dots, \{f_m, \mathbf{h}_m\}$ , nesta ordem e exatamente uma vez é chamado ciclo. Vamos assumir que o passo  $\alpha_k$  é constante ao longo de um ciclo, ou seja, para todo  $k$ , onde  $i_k \equiv 1$ , temos que  $\alpha_k = \alpha_{k+1} = \dots = \alpha_{k+m-1}$ .

## 2.1 Método do Subgradiente Incremental

Veremos a seguir o método do subgradiente projetado, em sua versão clássica. Por simplicidade, vamos nos referir à ele como método do subgradiente.

### 2.1.1 Método do Subgradiente

O método do Subgradiente é um algoritmo usado em problemas de minimização de funções convexas não diferenciáveis, este método foi originalmente desenvolvido por Shor [24], na União Soviética, nas décadas de 60 e 70. Considere o seguinte problema

$$\min_{x \in C} F(x), \quad (2.2)$$

onde  $F : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  é uma função convexa, fechada e própria e  $C$  um conjunto convexo, não vazio e fechado. Vamos considerar a seguinte hipótese:  $C \subset \text{ri}(\text{dom}(F))$ . Assim, iniciando com um ponto em  $x_0 \in C$ , este método gera uma sequência  $\{x_k\}$  de pontos no  $\mathbb{R}^n$ , da forma

$$x_{k+1} = P_C \left( x_k - \alpha_k \tilde{\nabla} F(x_k) \right). \quad (2.3)$$

Onde  $\tilde{\nabla} F(x_k)$  é algum subgradiente de  $F$  em  $x_k$  e  $\alpha_k$  é um escalar positivo.

Este método possui algumas propriedades de convergência, para mais detalhes consulte [5].

### 2.1.2 Método do Subgradiente Incremental(SI)

Este método é voltado para problemas como (2.1), ele pode ser visto como uma generalização do Método do Gradiente Incremental (GI), cujo algoritmo é da seguinte forma: Escolha  $x_0 \in C$

**Passo 1** Seja

$$\Phi_{0,m} = x_k$$

**Passo 2** Após obter  $\nabla f_{i_k}(\Phi_{i-1,k})$  obtenha

$$\Phi_{i,k} = P_C(\Phi_{i-1,k} - \alpha_k \nabla f_{i_k}(\Phi_{i-1,k})), \quad i = 1, \dots, m.$$

**Passo 3** Atualize  $x_{k+1} = \Phi_{k,m}$

Tome  $k := k + 1$  e retorne para o **Passo 1**.

Onde  $\alpha_k$  é um passo positivo.

Veja que este algoritmo é um ciclo de  $m$  subiterações, assim se  $x_k$  é um ponto obtido após  $k$  ciclos, então o ponto  $x_{k+1}$  obtido após mais um ciclo é dado por  $\Phi_{k,m}$  (que é obtido após  $m$  passos) e este ciclo foi iniciado com  $x_k = \Phi_{0,k}$ . O método SI foi primeiramente proposto na União Soviética por Kibardin [15], onde o problema (2.1), foi considerado o caso irrestrito. Posteriormente, considerando o problema no caso restrito, Solodov e Zavriev [25] analisaram a convergência deste método, considerando a função objetivo Lipschitz contínua e  $C$  adotado como convexo e compacto. Nedić e Bertsekas [20] também propuseram uma análise de convergência deste método, porém neste exige a convexidade da função  $f$ . O método do subgradiente incremental tem a seguinte forma

$$x_{k+1} = P_C \left( x_k - \alpha_k \tilde{\nabla} f_{i_k}(x_k) \right). \quad (2.4)$$

Onde  $\tilde{\nabla} f_{i_k}(x_k)$  é um subgradiente de  $f_{i_k}$  em  $x_k$ . Ele opera de forma análoga ao método GI, em ciclos (de  $m$  subiterações) e ao invés de usar o gradiente  $\nabla f_{i_k}(x_k)$ , da componente  $f_{i_k}$ , é usado um subgradiente arbitrário  $\tilde{\nabla} f_{i_k}(x_k)$ . A semelhança está também na necessidade de diminuição do passo  $\alpha_k$  para que o método tenha convergência. É importante ressaltar que o método SI converge mais rapidamente quando o ponto inicial é tomado distante do ponto convergente, mas convergem lentamente se for tomado próximo do ponto limite de convergência, pois objetiva minimizar cada função componente separadamente, veja [3].

## 2.2 Método Proximal Incremental

Visando um melhor entendimento do método proximal, veremos agora a sua versão para uma única função e algumas de suas propriedades.

### 2.2.1 Método do Ponto Proximal

O algoritmo de ponto proximal foi proposto por Martinet [19] e Rockafellar [22], na década de 70 e ele é usado resolver problemas como o seguinte

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

sendo  $F : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  uma função convexa, fechada e própria  $C \subset \mathbb{R}^n$  um conjunto convexo, fechado e não vazio. Consideremos aqui a seguinte hipótese:  $C \subset \text{ri}(\text{dom}(F))$ . Iniciando com um ponto  $x_0 \in C$ , este método gera sequências,  $\{x_k\} \subset \mathbb{R}^n$ , da seguinte forma

$$x_{k+1} = \arg \min_{x \in C} \left\{ F(x) + \frac{1}{2\alpha_k} \|x - x_k\|^2 \right\}, \quad (2.5)$$

onde  $\alpha_k$  é uma sequência de escalares positivos. Observe que este método está bem definido, para isso considere

$$F_k(x) = F(x) + I_C(x) + \frac{1}{2\alpha_k} \|x - x_k\|^2,$$

Veja que o problema (2.5) é equivalente à

$$x_{k+1} = \arg \min_{x \in \mathbb{R}^n} F_k(x).$$

Dessa forma, sendo  $F_k$  a soma de funções convexas,  $F$  e  $I_C$ , com uma função estritamente convexa,  $\|x - x_k\|^2$ , temos que  $F_k$  é função estritamente convexa e como consequência, possui um único mínimo global, garantindo assim a boa definição do algoritmo.

**Teorema 13.** *Seja  $\{x_k\}$  a sequência gerada pela iteração (2.5), então*

$$\frac{1}{\alpha_k} (x_k - x_{k+1}) \in \partial F(x_{k+1}) + N_C(x_{k+1}).$$

*Demonstração:* Se  $x_{k+1}$  é um minimizador de  $F_k$  então

$$\begin{aligned} 0 \in \partial F_k(x_{k+1}) &= \partial \left( F(x_{k+1}) + I_C(x_{k+1}) + \frac{1}{2\alpha_k} \|x_{k+1} - x_k\|^2 \right) \\ &= \partial F(x_{k+1}) + N_C(x_{k+1}) + \frac{1}{\alpha_k} (x_{k+1} - x_k). \end{aligned}$$

Assim, obtemos que

$$\frac{1}{\alpha_k} (x_k - x_{k+1}) \in \partial F(x_{k+1}) + N_C(x_{k+1}).$$

Provando a validade do teorema. ◊

**Lema 1.** *Se  $\{x_k\}$  é a sequência gerada pela iteração acima, então para todo  $x \in C$  vale*

$$\|x_{k+1} - x\|^2 \leq \|x_k - x\|^2 - \|x_k - x_{k+1}\|^2 - 2\alpha_k (F(x_{k+1}) - F(x)) \quad (2.6)$$

$$\leq \|x_k - x\|^2 - 2\alpha_k (F(x_{k+1}) - F(x)). \quad (2.7)$$

*Demonstração:* Seja  $x \in C$  então

$$\|x_k - x\|^2 = \|x_k - x_{k+1} + x_{k+1} - x\|^2 \quad (2.8)$$

$$= \|x_k - x_{k+1}\|^2 - 2\langle x_k - x_{k+1}, x - x_{k+1} \rangle + \|x_{k+1} - x\|^2. \quad (2.9)$$

Temos, pelo Teorema 13, que

$$\frac{1}{\alpha_k}(\mathbf{x}_k - \mathbf{x}_{k+1}) \in \partial F(\mathbf{x}_{k+1}) + \mathbf{N}_C(\mathbf{x}_{k+1}),$$

usando esta relação e as definições de  $\partial F(\mathbf{x}_{k+1})$  e  $\mathbf{N}_C(\mathbf{x}_{k+1})$  obtemos,

$$F(\mathbf{x}_{k+1}) + \frac{1}{\alpha_k} \langle \mathbf{x}_k - \mathbf{x}_{k+1}, \mathbf{x} - \mathbf{x}_{k+1} \rangle \leq F(\mathbf{x}). \quad (2.10)$$

Combinando (2.10) com a igualdade (2.8) – (2.9) temos

$$\|\mathbf{x}_{k+1} - \mathbf{x}\|^2 \leq \|\mathbf{x}_k - \mathbf{x}\|^2 - \|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2 - 2\alpha_k (F(\mathbf{x}_{k+1}) - F(\mathbf{x})).$$

A desigualdade (2.5) é clara. ◇

**Lema 2.** A sequência  $\{F(\mathbf{x}_k)\}$  é monótona não crescente.

*Demonstração:* Usando o Lema 1 e fazendo  $\mathbf{x} = \mathbf{x}_k$  temos que

$$0 \leq \|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \leq -\alpha_k (F(\mathbf{x}_{k+1}) - F(\mathbf{x}_k)).$$

Claramente obtemos,

$$F(\mathbf{x}_{k+1}) \leq F(\mathbf{x}_k),$$

garantindo então, veracidade deste lema. ◇

O próximo Teorema diz respeito à convergência do método em questão e para um melhor esclarecimento deste, vamos assumir que

$$F^* = \inf_{\mathbf{x} \in C} F(\mathbf{x}) \quad \text{e} \quad S = \{\bar{\mathbf{x}} \in C \mid F(\bar{\mathbf{x}}) = F^*\},$$

Podendo o conjunto  $S$  ser vazio.

**Teorema 14.** Sejam  $\{\mathbf{x}_k\}$  é a sequência gerada pelo algoritmo (2.5). Se a sequência  $\{\alpha_k\}$

é tal que  $\sum_{k=0}^{\infty} \alpha_k = +\infty$ , então

$$\liminf_k F(\mathbf{x}_k) = F^*.$$

E além disso, se o conjunto  $S$  é não vazio então  $\lim_k \mathbf{x}_k = \bar{\mathbf{x}}$ , com  $\bar{\mathbf{x}} \in S$ .

*Demonstração:* Vamos demonstrar inicialmente a primeira parte. Assim, suponha por absurdo que tal limite não seja verdadeiro, então existe  $\epsilon > 0$  tal que

$$\liminf F(\mathbf{x}_k) > F^* + 2\epsilon.$$

Assim, existe  $\bar{\mathbf{y}} \in C$  de tal forma que,

$$\liminf F(\mathbf{x}_k) > F(\bar{\mathbf{y}}) + 2\epsilon.$$

Seja  $k_0$  suficientemente grande, assim, para todo  $k \geq k_0$  obtemos

$$F(x_k) \geq \liminf F(x_k) - \epsilon.$$

Combinando estas duas desigualdades obtemos

$$F(x_k) - F(\bar{y}) \geq \epsilon. \quad (2.11)$$

Pelo Lema 1, temos que

$$2\alpha_k (F(x_{k+1}) - F(x)) \leq \|x_k - x\|^2 - \|x_{k+1} - x\|^2.$$

Fazendo  $k = 0, 1, \dots, n$  na desigualdade acima e somando-as obtemos

$$\begin{aligned} 2 \sum_{k=0}^n \alpha_k (F(x_{k+1}) - F(x)) &\leq \sum_{k=0}^n (\|x_k - x\|^2 - \|x_{k+1} - x\|^2) \\ &= \|x_0 - x\|^2 - \|x_{n+1} - x\|^2. \end{aligned}$$

Dessa forma, a expressão abaixo é verdadeira

$$2 \sum_{k=0}^n \alpha_k (F(x_{k+1}) - F(x)) \leq \|x_0 - x\|^2 - \|x_{n+1} - x\|^2,$$

então

$$2 \sum_{k=0}^n \alpha_k (F(x_{k+1}) - F(x)) \leq \|x_0 - x\|^2.$$

Fazendo  $\bar{y} = x$  nesta última relação e considerando (2.11) temos,

$$2 \sum_{k=0}^n \alpha_k \epsilon \leq \|x_0 - \bar{y}\|^2. \quad (2.12)$$

Um absurdo, para  $k$  suficientemente grande.

Agora, para demonstrar a segunda parte, suponha que  $F$  tenha um minimizador, ou seja,  $S \neq \emptyset$  e suponha que  $\bar{y} \in S$ , então pelo Lema 1, temos que

$$\|x_{k+1} - \bar{y}\|^2 \leq \|x_k - \bar{y}\|^2.$$

Assim, temos que  $\{x_k\}$  é Fejér convergente ao conjunto  $S$  e pela definição de Fejér convergente temos que esta sequência é limitada e dessa forma, pelo Teorema de Weierstrass, ela possui uma subsequência convergente, sejam  $\{x_{k_j}\}$  esta subsequência e  $\bar{x} = \lim_k x_{k_j}$ . Assim,  $\bar{x}$  é um ponto de acumulação de  $\{x_k\}$  e pela Proposição de sequências Fejér convergentes temos que,  $\bar{x} = \lim_k x_k$ . E como  $F$  é fechada temos que

$$\liminf_k F(x_k) \geq F(\bar{x}),$$

logo  $\bar{x} \in S$ . ◇



### 2.2.2 Método Proximal Incremental

A seguinte iteração proximal incremental resolve problemas como em (2.1)

$$\mathbf{x}_{k+1} = \arg \min_{\mathbf{x} \in \mathbf{C}} \left\{ f_{i_k}(\mathbf{x}) + \frac{1}{2\alpha_k} \|\mathbf{x} - \mathbf{x}_k\|^2 \right\}. \quad (2.13)$$

Onde  $f_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  são funções convexas,  $\alpha_k$  são sequências escalares positivas e  $\mathbf{C}$  é um conjunto fechado, convexo. Vamos considerar a seguinte hipótese:  $\mathbf{C} \subset \text{dom}(f_i)$ . O método em (2.3), tem uma relação direta com o algoritmo de minimização proximal proposto por Martinet [19] e Rockafellar [22]. Estes tipos de métodos são considerados, geralmente, mais estáveis que o método do subgradiente, pelo fato de ser possível obter a limitação das sequências geradas por ele. No entanto, Bertsekas em [4] menciona que este método pode requer uma estrutura especial, no que diz respeito as funções componentes. E dessa forma, faz sentido considerar um método que combine as iterações proximais e as iterações de subgradiente/gradiente, como veremos no Capítulo 3. Para Ma [23], assumir que uma certa função  $f$  tem uma estrutura especial, significa dizer seu operador proximal dado por

$$\text{prox}_f(\mathbf{u}) := \arg \min_{\mathbf{x} \in \mathbb{R}^n} \left( f(\mathbf{x}) + \frac{1}{2\lambda} \|\mathbf{x} - \mathbf{u}\|^2 \right)$$

com parâmetro  $\lambda > 0$ , é relativamente fácil de computar. Exemplo:

**Exemplo 7.** Se  $f(\mathbf{x}) = 0$  então  $\text{prox}_f(\mathbf{u}) = \mathbf{u}$ .

**Exemplo 8.** Se  $f(\mathbf{x}) = I_{\mathbf{C}}(\mathbf{x})$  então  $\text{prox}_f(\mathbf{u}) = \arg \min_{\mathbf{x} \in \mathbf{C}} \|\mathbf{x} - \mathbf{u}\|^2 = \mathbf{P}_{\mathbf{C}}(\mathbf{u})$ .

A proposição a seguir nos dá resultados que serão muito úteis, no próximo capítulo. A parte a) obtemos uma aproximação entre as iterações proximais e de subgradiente e na parte b) obtemos uma desigualdade que será usada na análise de convergência do método no capítulo seguinte.

**Proposição 5.** Sejam  $\mathbf{C}$  um conjunto não vazio, convexo, fechado e  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  uma função convexa, fechada e própria tal que  $\mathbf{C} \subset \text{ri}(\text{dom}(f))$ . Considere a seguinte iteração proximal

$$\mathbf{x}_{k+1} = \arg \min_{\mathbf{x} \in \mathbf{C}} \left\{ f(\mathbf{x}) + \frac{1}{2\alpha_k} \|\mathbf{x} - \mathbf{x}_k\|^2 \right\} \quad (2.14)$$

Para qualquer  $\mathbf{x}_k \in \mathbf{C}$  e  $\alpha_k > 0$ , temos:

(a) A iteração pode ser escrita como

$$\mathbf{x}_{k+1} = \mathbf{P}_C(\mathbf{x}_k - \alpha_k \tilde{\nabla} f(\mathbf{x}_{k+1})), \quad (2.15)$$

onde  $\tilde{\nabla} f(\mathbf{x}_{k+1})$  é algum subgradiente de  $f$  em  $\mathbf{x}_{k+1}$ .

(b) Para todo  $\mathbf{y} \in C$ , temos que

$$\begin{aligned} \|\mathbf{x}_{k+1} - \mathbf{y}\|^2 &\leq \|\mathbf{x}_k - \mathbf{y}\|^2 - 2\alpha_k (f(\mathbf{x}_{k+1}) - f(\mathbf{y})) - \|\mathbf{x}_k - \mathbf{x}_{k+1}\|^2 \\ &\leq \|\mathbf{x}_k - \mathbf{y}\|^2 - 2\alpha_k (f(\mathbf{x}_{k+1}) - f(\mathbf{y})) \end{aligned} \quad (2.16)$$

*Demonstração:*

(a) Veja que mostrar que o método em (2.14) está bem definido e é equivalente a mostrar a boa definição de

$$\mathbf{x}_{k+1} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \left\{ f(\mathbf{x}) + I_C(\mathbf{x}) + \frac{1}{2\alpha_k} \|\mathbf{x} - \mathbf{x}_k\|^2 \right\}.$$

Para mostrar este último, considere

$$f_k(\mathbf{x}) = f(\mathbf{x}) + I_C(\mathbf{x}) + \frac{1}{2\alpha_k} \|\mathbf{x} - \mathbf{x}_k\|^2.$$

Veja que  $f$  e  $I_C$  são funções convexas no  $\mathbb{R}^n$  e  $g(\mathbf{x}) = \frac{1}{2\alpha_k} \|\mathbf{x} - \mathbf{x}_k\|^2$  é fortemente convexa, portanto  $f_k$  é fortemente convexa no  $\mathbb{R}^n$  o que implica que  $f_k$  possui um único minimizador global, daí temos a boa definição do método. Temos que  $f_k$  é própria e  $\text{ri}(\text{dom}(f)) \cap \text{ri}(\text{dom}(I_C)) \cap \text{ri}(\text{dom}(g)) \neq \emptyset$ , então pelo Teorema 7 temos que

$$\partial f_k(\mathbf{x}) = \partial f(\mathbf{x}) + \partial I_C(\mathbf{x}) + \partial g(\mathbf{x}),$$

assim,

$$\partial f_k(\mathbf{x}) = \partial f(\mathbf{x}) + \mathbf{N}_C(\mathbf{x}) + \frac{1}{\alpha_k}(\mathbf{x} - \mathbf{x}_k).$$

Como  $\mathbf{x}_{k+1}$  é minimizador de  $f_k$  temos

$$0 \in \partial f_k(\mathbf{x}_{k+1}) = \partial f(\mathbf{x}_{k+1}) + \mathbf{N}_C(\mathbf{x}_{k+1}) + \frac{1}{\alpha_k}(\mathbf{x}_{k+1} - \mathbf{x}_k).$$

Da relação acima, temos que existem vetores  $\mathbf{v}_1 \in \partial f(\mathbf{x}_{k+1})$  e  $\mathbf{v}_2 \in \mathbf{N}_C(\mathbf{x}_{k+1})$  tal que

$$\frac{1}{\alpha_k}(\mathbf{x}_k - \mathbf{x}_{k+1}) = \mathbf{v}_1 + \mathbf{v}_2,$$

portanto,

$$\frac{1}{\alpha_k}(\mathbf{x}_k - \mathbf{x}_{k+1}) \in \partial f(\mathbf{x}_{k+1}) + \mathbf{N}_C(\mathbf{x}_{k+1}). \quad (2.17)$$

Dessa forma, de (2.17) podemos concluir que existe  $\tilde{\nabla}f(x_{k+1}) \in \partial f(x_{k+1})$  tal que

$$(x_k - x_{k+1}) - \alpha_k \tilde{\nabla}f(x_{k+1}) \in N_C(x_{k+1})$$

e pelo Teorema da Projeção temos que a relação acima é equivalente à

$$x_{k+1} = P_C \left( x_k - \alpha_k \tilde{\nabla}f(x_{k+1}) \right).$$

(b) Veja que  $\|x_k - y\|^2 = \|x_k - x_{k+1} + x_{k+1} - y\|^2$  e expandindo esta norma temos que

$$\|x_k - y\|^2 = \|x_k - x_{k+1}\|^2 - 2\langle x_k - x_{k+1}, y - x_{k+1} \rangle + \|x_{k+1} - y\|^2. \quad (2.18)$$

Por (2.17) temos que existem vetores  $v_1 \in \partial f(x_{k+1})$  e  $v_2 \in N_C(x_{k+1})$  tais que

$$\frac{1}{\alpha_k}(x_k - x_{k+1}) = v_1 + v_2.$$

Usando a definição de subgradiente e cone normal temos que para todo  $y \in C$  vale

$$f(x_{k+1}) + \langle v_1, y - x_{k+1} \rangle \leq f(y) \text{ e } \langle v_2, y - x_{k+1} \rangle \leq 0$$

e somando estas desigualdades obtemos

$$f(x_{k+1}) + \frac{1}{\alpha_k} \langle x_k - x_{k+1}, y - x_{k+1} \rangle \leq f(y).$$

Da desigualdade acima temos que

$$\langle x_k - x_{k+1}, y - x_{k+1} \rangle \leq \alpha_k (f(y) - f(x_{k+1})).$$

Combinando esta ultima desigualdade com (2.18) temos que a desigualdade em (2.16) é válida.

◇

# Capítulo 3

## Método Proximal-Subgradiente Incremental

Este novo método foi estudado por Bertsekas [3], [4] e os resultados obtidos nesse trabalho tem como base os trabalhos deste autor.

Em problemas de minimização convexa, podem ocorrer casos em que algumas funções tenham uma estrutura especial, de tal forma que aplicar o método proximal pode ser mais vantajoso que o método do subgradiente, como pode haver também outros em que certas funções sejam inapropriadas para as iterações proximais, sendo mais viáveis às iterações de subgradiente, isto porque, a seguinte minimização

$$\mathbf{x}_{k+1} = \arg \min_{\mathbf{x} \in C} \left\{ f_{i_k}(\mathbf{x}) + \frac{1}{2\alpha_k} \|\mathbf{x} - \mathbf{x}_k\|^2 \right\}$$

pode tornar-se inconveniente para estas funções. Levando isto em conta, faz sentido considerar uma combinação entre as iterações proximais e de subgradiente. E pensando nisto, para o problema a seguir, vamos supor aqui que as funções  $f_i$  sejam apropriadas para as iterações proximais, enquanto que as funções  $h_i$  não são e assim, sendo preferíveis as iterações que envolvem o subgradiente.

$$\min_{\mathbf{x} \in C} F(\mathbf{x}) = \sum_{i=1}^m F_i(\mathbf{x}). \quad (3.1)$$

Onde  $F_i$  é da forma

$$F_i(\mathbf{x}) = f_i(\mathbf{x}) + h_i(\mathbf{x})$$

$f_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  e  $h_i : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ ,  $i = 1, \dots, m$  são funções convexas, fechadas e próprias, e  $C$  é um conjunto não vazio, fechado e convexo. Vamos supor aqui que  $C$

satisfaça a seguinte condição:

$$C \subset \text{ri}(\text{dom}(f_i)) \quad \text{e} \quad C \subset \text{ri}(\text{dom}(h_i)).$$

E assim nosso algoritmo para problemas como em (3.1) será da seguinte maneira

$$z_k = \arg \min_{x \in C} \left\{ f_{i_k}(x) + \frac{1}{2\alpha_k} \|x - x_k\|^2 \right\} \quad (3.2)$$

$$x_{k+1} = P_C \left( z_k - \alpha_k \tilde{\nabla} h_{i_k}(z_k) \right). \quad (3.3)$$

A boa definição deste algoritmo é garantida pelo fato das funções  $f_i$  serem convexas e  $\|x - x_k\|^2$  ser uma função estritamente convexa e, de acordo com Teorema 3, o mínimo em (3.2) existe e neste caso é único. Quanto ao subdiferencial  $\partial h_i(z_k)$  é diferente do vazio, pois as funções  $h_i$  são convexas.

Quando as funções  $f_i$  ou  $h_i$  são todas nulas, temos casos especiais como as iterações proximais (3.2) ou de subgradiente (3.3), respectivamente. Claramente, podemos ver que as sequências  $\{x_k\}$  e  $\{z_k\}$  estão restritas ao conjunto  $C$ . Mas, como vimos na Observação 4, podemos transformar um problema com restrições em um problema irrestrito, facilitando assim os cálculos. Dessa forma haverá casos onde podemos trabalhar com irrestrições tanto para as iterações proximais quanto para as de subgradiente, o que nos leva aos seus respectivos algoritmos

$$z_k = \arg \min_{x \in \mathbb{R}^n} \left\{ f_{i_k}(x) + \frac{1}{2\alpha_k} \|x - x_k\|^2 \right\} \quad (3.4)$$

$$x_{k+1} = P_C \left( z_k - \alpha_k \tilde{\nabla} h_{i_k}(z_k) \right) \quad (3.5)$$

e

$$z_k = x_k - \alpha_k \tilde{\nabla} h_{i_k}(x_k). \quad (3.6)$$

$$x_{k+1} = \arg \min_{x \in C} \left\{ f_{i_k}(x) + \frac{1}{2\alpha_k} \|x - z_k\|^2 \right\}. \quad (3.7)$$

Veja que diante dos algoritmos citados e com a parte (a) da Proposição 5, obtemos os seguintes resultados:

(1) A iteração (3.2)-(3.3) pode ser escrita como

$$z_k = P_C \left( x_k - \alpha_k \tilde{\nabla} f_{i_k}(z_k) \right), \quad x_{k+1} = P_C \left( z_k - \alpha_k \tilde{\nabla} h_{i_k}(z_k) \right). \quad (3.8)$$

(2) A iteração (3.4)-(3.5) pode ser escrita como

$$z_k = x_k - \alpha_k \tilde{\nabla} f_{i_k}(z_k), \quad x_{k+1} = P_C \left( z_k - \alpha_k \tilde{\nabla} h_{i_k}(z_k) \right). \quad (3.9)$$

(3) A iteração (3.6)-(3.7) pode ser escrita como

$$z_k = x_k - \alpha_k \tilde{\nabla} h_{i_k}(x_k), \quad x_{k+1} = P_C \left( z_k - \alpha_k \tilde{\nabla} f_{i_k}(x_{k+1}) \right). \quad (3.10)$$

Observe que pelas atualizações acima, o subgradiente  $\tilde{\nabla} h_{i_k}$  pode ser qualquer vetor no subdiferencial de  $h_{i_k}$ , mas  $\tilde{\nabla} f_{i_k}$  é um vetor específico, em  $\partial f_{i_k}$ , que satisfaz Proposição 5(a). Observe também que a iteração (3.9) ainda pode ser escrita da seguinte forma

$$x_{k+1} = P_C \left( x_k - \alpha_k \tilde{\nabla} F_{i_k}(z_k) \right).$$

e assim temos o método do subgradiente para minimizar sobre o conjunto  $C$  a função objetivo

$$F(x) = \sum_{i=1}^m F_i(x) = \sum_{i=1}^m (f_i(x) + h_i(x)).$$

### 3.1 Análise de Convergência para o Método com a Ordem Cíclica

Nesta seção, faremos uma análise da convergência do método para a ordem cíclica. Denotaremos por  $F^*$  o valor ótimo:

$$F^* = \inf_{x \in C} F(x),$$

e por  $S$  o conjunto de soluções ótimas (podendo ser vazio):

$$S = \{\bar{x} \in C \mid F(\bar{x}) = F^*\}.$$

E mais, para um conjunto  $C$  não vazio, fechado e convexo, denotamos por  $\text{dist}(\cdot; C)$  como sendo a função distância dada por

$$\text{dist}(x; C) = \min_{z \in C} \|x - z\|, \quad x \in \mathbb{R}^n.$$

Vamos assumir as seguintes hipóteses,

**Hipótese 1.** (Para as iterações (3.8) e (3.9)) Existe uma constante  $\mathbf{c} \in \mathbb{R}$  tal que para todo  $\mathbf{k}$  temos

$$\max \left\{ \|\tilde{\nabla} f_{i_{\mathbf{k}}}(z_{\mathbf{k}})\|, \|\tilde{\nabla} h_{i_{\mathbf{k}}}(z_{\mathbf{k}})\| \right\} \leq \mathbf{c}. \quad (3.11)$$

Além disso, para todo  $\mathbf{k}$  que marca o início de um ciclo (isto é  $\mathbf{k} > 0$ , com  $i_{\mathbf{k}} \equiv 1$ ), temos que para todo  $\mathbf{j} = 1, 2, \dots, \mathbf{m}$ ,

$$\max \{f_{\mathbf{j}}(x_{\mathbf{k}}) - f_{\mathbf{j}}(z_{\mathbf{k}+\mathbf{j}-1}), h_{\mathbf{j}}(x_{\mathbf{k}}) - h_{\mathbf{j}}(z_{\mathbf{k}+\mathbf{j}-1})\} \leq \mathbf{c} \|x_{\mathbf{k}} - z_{\mathbf{k}+\mathbf{j}-1}\|. \quad (3.12)$$

**Hipótese 2.** (Para a iteração (3.10)) Existe uma constante  $\mathbf{c} \in \mathbb{R}$  tal que para todo  $\mathbf{k}$  tem-se

$$\max \left\{ \|\tilde{\nabla} f_{i_{\mathbf{k}}}(x_{\mathbf{k}+1})\|, \|\tilde{\nabla} h_{i_{\mathbf{k}}}(x_{\mathbf{k}})\| \right\} \leq \mathbf{c}. \quad (3.13)$$

Além disso, para todo  $\mathbf{k}$  que marca o início de um ciclo (isto é  $\mathbf{k} > 0$ , com  $i_{\mathbf{k}} \equiv 1$ ), temos que para todo  $\mathbf{j} = 1, 2, \dots, \mathbf{m}$ ,

$$\max \{f_{\mathbf{j}}(x_{\mathbf{k}}) - f_{\mathbf{j}}(x_{\mathbf{k}+\mathbf{j}-1}), h_{\mathbf{j}}(x_{\mathbf{k}}) - h_{\mathbf{j}}(x_{\mathbf{k}+\mathbf{j}-1})\} \leq \mathbf{c} \|x_{\mathbf{k}} - x_{\mathbf{k}+\mathbf{j}-1}\|, \quad (3.14)$$

$$f_{\mathbf{j}}(x_{\mathbf{k}+\mathbf{j}-1}) - f_{\mathbf{j}}(x_{\mathbf{k}+\mathbf{j}}) \leq \mathbf{c} \|x_{\mathbf{k}+\mathbf{j}-1} - x_{\mathbf{k}+\mathbf{j}}\|. \quad (3.15)$$

Veja que da convexidade de  $f_{\mathbf{j}}$  e  $h_{\mathbf{j}}$ , garantimos a existência de seus respectivos subgradientes em  $x_{\mathbf{k}}$ , logo existem  $\tilde{\nabla} f_{\mathbf{j}}(x_{\mathbf{k}}) \in \partial f_{\mathbf{j}}(x_{\mathbf{k}})$  e  $\tilde{\nabla} h_{\mathbf{j}}(x_{\mathbf{k}}) \in \partial h_{\mathbf{j}}(x_{\mathbf{k}})$  de forma que

$$f_{\mathbf{j}}(x_{\mathbf{k}}) - f_{\mathbf{j}}(y) \leq \langle \tilde{\nabla} f_{\mathbf{j}}(x_{\mathbf{k}}), x_{\mathbf{k}} - y \rangle \leq \|\tilde{\nabla} f_{\mathbf{j}}(x_{\mathbf{k}})\| \|x_{\mathbf{k}} - y\|, \quad \forall y \in \mathbb{R}^n$$

e

$$h_{\mathbf{j}}(x_{\mathbf{k}}) - h_{\mathbf{j}}(y) \leq \langle \tilde{\nabla} h_{\mathbf{j}}(x_{\mathbf{k}}), x_{\mathbf{k}} - y \rangle \leq \|\tilde{\nabla} h_{\mathbf{j}}(x_{\mathbf{k}})\| \|x_{\mathbf{k}} - y\|, \quad \forall y \in \mathbb{R}^n.$$

e assim, se estes subgradientes forem limitados em norma por  $\mathbf{c}$ , temos que as condições (3.12) e (3.14) são satisfeitas. A seguir temos algumas condições que implicam nestas hipóteses:

- a) Para o algoritmo (3.8): Se  $f_i$  e  $h_i$  são Lipschitz contínuas em  $\mathbf{C}$ .
- b) Para os algoritmos (3.9) e (3.10): Se  $f_i$  e  $h_i$  são Lipschitz contínuas em  $\mathbb{R}^n$ .
- c) Se as sequências  $\{x_{\mathbf{k}}\}$  e  $\{z_{\mathbf{k}}\}$  são limitadas (pois  $f_i$  e  $h_i$  são funções de valores reais e convexas, logo serão Lipschitz contínuas sobre qualquer conjunto limitado que contenha  $\{x_{\mathbf{k}}\}$  e  $\{z_{\mathbf{k}}\}$ ).

**Proposição 6.** *Seja  $\{x_k\}$  a sequência gerada pelos algoritmos (3.8)-(3.10), com a ordem cíclica da seleção de componentes. Então para todo  $y \in C$  e para todo  $k$  que marca o início de um ciclo( ou seja  $i_k = 1$ ), temos*

$$\|x_{k+m} - y\|^2 \leq \|x_k - y\|^2 - 2\alpha_k(F(x_k) - F(y)) + \alpha_k^2 \beta m^2 c^2. \quad (3.16)$$

Onde  $\beta = \frac{1}{m} + 4$  no caso (3.8) e (3.9) e  $\beta = \frac{5}{m} + 4$  no caso (3.10).

*Demonstração:* Primeiro vamos provar o resultado para os algoritmos (3.8) e (3.9). Para isso, veja que, pela Proposição 5(b), temos que

$$\|z_k - y\|^2 \leq \|x_k - y\|^2 - 2\alpha_k (f_{i_k}(z_k) - f_{i_k}(y)) \quad \forall y \in C. \quad (3.17)$$

Veja também que devido a projeção  $(P_C)$  ser Lipschitz temos

$$\begin{aligned} \|x_{k+1} - y\|^2 &= \|P_C(z_k - \alpha_k \tilde{\nabla} h_{i_k}(z_k)) - y\|^2 \\ &\leq \|z_k - \alpha_k \tilde{\nabla} h_{i_k}(z_k) - y\|^2 \\ &= \|z_k - y\|^2 - 2\alpha_k \langle \tilde{\nabla} h_{i_k}(z_k), z_k - y \rangle + \alpha_k^2 \|\tilde{\nabla} h_{i_k}(z_k)\|^2. \end{aligned}$$

Da relação acima, diante da Hipótese 1 e da definição de subgradiente obtemos

$$\|x_{k+1} - y\|^2 \leq \|z_k - y\|^2 - 2\alpha_k (h_{i_k}(z_k) - h_{i_k}(y)) + \alpha_k^2 c^2. \quad (3.18)$$

Combinando (3.17) e (3.18) e usando a definição de  $F_i = f_i + h_i$  obtemos

$$\|x_{k+1} - y\|^2 \leq \|x_k - y\|^2 - 2\alpha_k (F_{i_k}(z_k) - F_{i_k}(y)) + \alpha_k^2 c^2. \quad (3.19)$$

Seja  $k$  o início de um ciclo, isto é  $i_k \equiv 1$ . Considerando a ordem cíclica, na iteração  $k+j-1$ , a seleção de componentes é  $\{f_j, h_j\}$ , onde  $j = 1, \dots, m$ . Usando (3.19) e substituindo  $k$  por  $k+1, \dots, k+m-1$  temos

$$\begin{aligned} \|x_{k+1} - y\|^2 &\leq \|x_k - y\|^2 - 2\alpha_k (F_1(z_k) - F_1(y)) + \alpha_k^2 c^2 \\ \|x_{k+2} - y\|^2 &\leq \|x_{k+1} - y\|^2 - 2\alpha_k (F_2(z_{k+1}) - F_2(y)) + \alpha_k^2 c^2 \\ &\dots \\ \|x_{k+m} - y\|^2 &\leq \|x_{k+m-1} - y\|^2 - 2\alpha_k (F_m(z_{k+m-1}) - F_m(y)) + \alpha_k^2 c^2. \end{aligned}$$

Combinando as desigualdades acima teremos

$$\|x_{k+m} - y\|^2 \leq \|x_k - y\|^2 - 2\alpha_k \sum_{j=1}^m (F_j(z_{k+j-1}) - F_j(y)) + m\alpha_k^2 c^2.$$



Sabemos que  $F = \sum_{j=1}^m F_j$ , com isto e adicionando  $\sum_{j=1}^m (F_j(x_k) - F_j(z_{k+j-1}))$  no lado direito da relação anterior obtemos ainda que

$$\begin{aligned} \|x_{k+m} - y\|^2 &\leq \|x_k - y\|^2 - 2\alpha_k (F(x_k) - F(y)) + m\alpha_k^2 c^2 \\ &\quad + 2\alpha_k \sum_{j=1}^m (F_j(x_k) - F_j(z_{k+j-1})). \end{aligned} \quad (3.20)$$

Da relação (3.12) conseguimos temos

$$F_j(x_k) - F_j(z_{k+j-1}) \leq 2c\|x_k - z_{k+j-1}\| \quad \forall j, \text{ tal que, } 1 \leq j \leq m. \quad (3.21)$$

Observe também que

$$\|x_k - z_{k+j-1}\| \leq \|x_k - x_{k+1}\| + \dots + \|x_{k+j-1} - z_{k+j-1}\|.$$

Pela definição dos algoritmos (3.8) e (3.9), pela propriedade Lipschitz de  $P_C$  e assumindo que vale a parte (3.11) da Hipótese 1 obtemos

$$\begin{aligned} \|x_q - x_{q+1}\| &= \|x_q - P_C(z_q - \alpha_q \tilde{\nabla} h_{i_q}(z_q))\| \\ &\leq \|x_q - z_q - \alpha_q \tilde{\nabla} h_{i_q}(z_q)\| \\ &\leq \|x_q - z_q\| + \alpha_q \|\tilde{\nabla} h_{i_q}(z_q)\| \\ &= \alpha_q \|\tilde{\nabla} f_{i_q}(z_q)\| + \alpha_q \|\tilde{\nabla} h_{i_q}(z_q)\| \\ &= 2\alpha_q c \quad \forall q = k, \dots, k+j-2. \end{aligned}$$

Assim,

$$\|x_k - z_{k+j-1}\| \leq 2\alpha_k c(j-1) + \alpha_k c = \alpha_k c(2j-1). \quad (3.22)$$

Combinando (3.21) e (3.22) obtemos

$$F_j(x_k) - F_j(z_{k+j-1}) \leq 2\alpha_k c^2(2j-1). \quad (3.23)$$

De (3.20) e (3.23) concluímos que

$$\begin{aligned} \|x_{k+m} - y\| &\leq \|x_k - y\|^2 - 2\alpha_k (F(x_k) - F(y)) + m\alpha_k^2 c^2 + 4\alpha_k^2 c^2 \sum_{j=1}^m (2j-1) \\ &= \|x_k - y\|^2 - 2\alpha_k (F(x_k) - F(y)) + \alpha_k^2 m^2 c^2 \left( \frac{1}{m} + 4 \right). \end{aligned}$$

Agora, vamos provar o resultado para o algoritmo (3.10). De tal algoritmo obtemos

$$\begin{aligned} \|z_k - y\|^2 &= \|x_k - y - \alpha_k \tilde{\nabla} h_{i_k}(x_k)\|^2 \\ &\leq \|x_k - y\|^2 - 2\alpha_k \langle \tilde{\nabla} h_{i_k}(x_k), x_k - y \rangle + \alpha_k^2 \|\tilde{\nabla} h_{i_k}(x_k)\|^2. \end{aligned}$$

Veja que,

$$h_{i_k}(y) - h_{i_k}(x_k) \geq -\langle \tilde{\nabla} h_{i_k}(x_k), x_k - y \rangle$$

com isto e fazendo uso da Hipótese 2 conseguimos ainda que

$$\|z_k - y\|^2 \leq \|x_k - y\|^2 - 2\alpha_k (h_{i_k}(x_k) - h_{i_k}(y)) + \alpha_k^2 c^2. \quad (3.24)$$

comparando (3.10) com (2.15) e usando a Proposição 5(b) temos

$$\|x_{k+1} - y\|^2 \leq \|z_k - y\|^2 - 2\alpha_k (f_{i_k}(x_{k+1}) - f_{i_k}(y)). \quad (3.25)$$

De (3.24) e (3.25) obtemos

$$\begin{aligned} \|x_{k+1} - y\|^2 &\leq \|x_k - y\|^2 - 2\alpha_k (f_{i_k}(x_{k+1}) + h_{i_k}(x_k) - f_{i_k}(y) - h_{i_k}(y)) + \alpha_k^2 c^2 \\ &= \|x_k - y\|^2 - 2\alpha_k (F_{i_k}(x_k) - F_{i_k}(y)) + 2\alpha_k (f_{i_k}(x_k) - f_{i_k}(x_{k+1})) + \alpha_k^2 c^2. \end{aligned}$$

Considere que  $k$  marca o começo de um ciclo ( $i_k \equiv 1$ ). Usando um procedimento análogo ao anterior (dos algoritmos (3.8) e (3.9)), repentinando o processo e substituindo  $k$  por  $k+1, \dots, k+m-1$ , temos que

$$\begin{aligned} \|x_{k+m} - y\|^2 &\leq \|x_k - y\|^2 - 2\alpha_k \sum_{j=1}^m (F_j(x_{k+j-1}) - F_j(y)) \\ &\quad + 2\alpha_k \sum_{j=1}^m (F_j(x_{k+j-1}) - F_j(x_{k+j})) + m\alpha_k^2 c^2. \end{aligned}$$

Adicionando  $\sum_{j=1}^m (F_j(x_k) - F_j(x_k))$  no lado direito da desigualdade acima obtemos ainda que

$$\begin{aligned} \|x_{k+m} - y\|^2 &\leq \|x_k - y\|^2 - 2\alpha_k [F(x_k) - F(y)] + 2\alpha_k \sum_{j=1}^m (F_j(x_k) - F_j(x_{k+j-1})) \\ &\quad + 2\alpha_k \sum_{j=1}^m (f_j(x_{k+j-1}) - f_j(x_{k+j})) + m\alpha_k^2 c^2. \end{aligned} \quad (3.26)$$

Usando a Hipótese 2, para os somatórios em (3.26) temos

$$F_j(x_k) - F_j(x_{k+j-1}) \leq 2c \|x_k - x_{k+j-1}\|.$$

Pela Desigualdade Triangular ganhamos que

$$F_j(x_k) - F_j(x_{k+j-1}) \leq 2c (\|x_k - x_{k+1}\| + \dots + \|x_{k+j-2} - x_{k+j-1}\|).$$

Devido a propriedade lipschitziana de  $P_C$ , a definição do algoritmo em questão e valendo-se de (3.13), vemos que cada norma acima é limitada por  $2\alpha_k c$ . Assim,

$$F_j(x_k) - F_j(x_{k+j-1}) \leq 4\alpha_k c^2 (j-1).$$

E mais, deste algoritmo e da Hipótese 2, parte (3.15), temos

$$f_j(\mathbf{x}_{k+j-1}) - f_j(\mathbf{x}_{k+j}) \leq c \|\mathbf{x}_{k+j-1} - \mathbf{x}_{k+j}\| \leq 2\alpha_k c^2.$$

Combinando as duas relações acima e somando-as obtemos

$$\begin{aligned} 2\alpha_k \sum_{j=1}^m (F_j(\mathbf{x}_k) - F_j(\mathbf{x}_{k+j-1})) + 2\alpha_k \sum_{j=1}^m (f_j(\mathbf{x}_{k+j-1}) - f_j(\mathbf{x}_{k+j})) \\ \leq 8\alpha_k^2 c^2 \sum_{j=1}^m (j-1) + 4\alpha_k^2 c^2 m \\ = 4\alpha_k^2 c^2 m^2 + 4\alpha_k^2 c^2 m \\ = \left(4 + \frac{4}{m}\right) \alpha_k^2 c^2 m^2. \end{aligned}$$

Por isto e pela desigualdade em (3.26) temos a validade da relação (3.16), onde  $\beta = \left(4 + \frac{5}{m}\right)$ .  $\diamond$

A Proposição 6 nos garante ainda que, considerando a ordem cíclica, dada a iterada  $\mathbf{x}_k$  no começo de um ciclo e qualquer  $\mathbf{y} \in \mathbf{C}$  com menor custo que  $\mathbf{x}_k$  (ou seja,  $F(\mathbf{y}) \leq F(\mathbf{x}_k)$ ), os algoritmos rendem um ponto  $\mathbf{x}_{k+m}$  (no final do ciclo) que estará mais próximo de  $\mathbf{y}$  que  $\mathbf{x}_k$ , desde que o passo  $\alpha_k$  seja suficientemente pequeno, tal que satisfaça

$$\alpha_k \leq \frac{2(F(\mathbf{x}_k) - F(\mathbf{y}))}{\beta m^2 c^2}.$$

Em particular, assumindo que exista a solução ótima  $\bar{\mathbf{x}}$ , para qualquer  $\epsilon > 0$  temos que

$$F(\mathbf{x}_k) \leq F(\bar{\mathbf{x}}) + \frac{\alpha_k \beta m^2 c^2}{2} + \epsilon,$$

ou a distância quadrada à  $\bar{\mathbf{x}}$  será estritamente decrescida por  $2\alpha_k \epsilon$ ,

$$\|\mathbf{x}_{k+m} - \bar{\mathbf{x}}\|^2 < \|\mathbf{x}_k - \bar{\mathbf{x}}\|^2 - 2\alpha_k \epsilon.$$

A seguinte proposição nos mostra que, para o passo  $\alpha_k$  constante, a convergência é estabilizada na vizinhança do valor ótimo.

**Proposição 7.** *Seja  $\{\mathbf{x}_k\}$  a sequência gerada por qualquer um dos algoritmos (3.8)–(3.10), com a ordem cíclica da seleção de componentes e seja  $\alpha_k = \alpha$ , para algum  $\alpha > 0$  constante. Temos que*

a) Se  $F^* = -\infty$ , então

$$\liminf_{k \rightarrow \infty} F(\mathbf{x}_k) = F^*.$$

b) Se  $F^* > -\infty$ , então

$$\liminf_{k \rightarrow \infty} F(x_k) \leq F^* + \frac{\alpha\beta m^2 c^2}{2}.$$

Onde  $c$  e  $\beta$  são constantes da Proposição 6.

*Demonstração:* Consideremos inicialmente a parte b). Suponha por absurdo que o resultado não é válido, então existe  $\epsilon > 0$  tal que

$$\liminf F(x_{km}) > F^* + \frac{\alpha\beta m^2 c^2}{2} + 2\epsilon. \quad (3.27)$$

Considere  $\bar{y} \in C$  tal que

$$\liminf F(x_{km}) - \frac{\alpha\beta m^2 c^2}{2} - 2\epsilon \geq F(\bar{y}),$$

e pela definição de limite inferior podemos encontrar  $k_0$  (suficientemente grande) tal que  $\forall k \geq k_0$  temos

$$F(x_{km}) - \frac{\alpha\beta m^2 c^2}{2} - 2\epsilon \geq F(\bar{y}),$$

e assim, da desigualdade acima obtemos

$$-(F(x_{km}) - F(\bar{y})) \leq -\left(\frac{\alpha\beta m^2 c^2}{2} + 2\epsilon\right). \quad (3.28)$$

Usando a Proposição 6, onde  $y = \bar{y}$  teremos

$$\|x_{(k+1)m} - \bar{y}\|^2 \leq \|x_{km} - \bar{y}\|^2 - 2\alpha(F(x_{km}) - F(\bar{y})) + \alpha^2\beta m^2 c^2. \quad (3.29)$$

E usando (3.28) em (3.29) e fazendo algumas manipulações obteremos

$$\|x_{(k+1)m} - \bar{y}\|^2 \leq \|x_{km} - \bar{y}\|^2 - 4\epsilon\alpha.$$

Desta desigualdade, obtemos que para todo  $k \geq k_0$ ,

$$\begin{aligned} \|x_{(k+1)m} - \bar{y}\|^2 &\leq \|x_{(k-1)m} - \bar{y}\|^2 - 8\alpha\epsilon \\ &\leq \dots \\ &\leq \|x_{k_0} - \bar{y}\|^2 - 4(k+1-k_0)\alpha\epsilon. \end{aligned}$$

Um absurdo para  $k$  suficientemente grande. Veja que a prova da parte a) é análoga à demonstração acima, basta supor por absurdo que a igualdade em a) não é válida e usar o fato de  $F^* = -\infty$  em (3.27) e deforma similar à b) garantimos a validade do resultado.  $\diamond$

Na proposição a seguir, podemos ver o número de iterações necessárias para garantir um certo nível de otimalidade.

**Proposição 8.** *Seja  $\{x_k\}$  uma sequência gerada como na Proposição 6 e suponha que  $S \neq \emptyset$ . Então para  $\epsilon > 0$  temos*

$$\min_{0 \leq k \leq N} F(x_k) \leq F^* + \frac{\alpha\beta m^2 c^2 + \epsilon}{2}. \quad (3.30)$$

Sendo  $N$  é dado por

$$N = m \left\lfloor \frac{\text{dist}(x_0; S)^2}{\alpha\epsilon} \right\rfloor,$$

e  $\left\lfloor \frac{\text{dist}(x_0; S)^2}{\alpha\epsilon} \right\rfloor$  definido como o maior inteiro menor ou igual à  $\frac{\text{dist}(x_0; S)^2}{\alpha\epsilon}$ .

*Demonstração:* Suponha, por absurdo, que (3.30) não é válida, então para todo  $k$ , onde  $0 \leq km \leq N$  tem-se

$$F(x_{km}) > F^* + \frac{\alpha\beta m^2 c^2 + \epsilon}{2}.$$

Usando a desigualdade acima e a Proposição 6, onde  $\alpha_k = \alpha$  e  $y$  é algum vetor de  $S$  que minimiza a distância de  $x_{jm}$  à  $S$ , obtemos

$$\begin{aligned} \text{dist}(x_{(k+1)m}; S)^2 &\leq \text{dist}(x_{km}; S)^2 - 2\alpha(F(x_{km}) - F^*) + \alpha^2\beta m^2 c^2 \\ &\leq \text{dist}(x_{km}; S)^2 - (\alpha^2\beta m^2 c^2 + \alpha\epsilon) + \alpha^2\beta m^2 c^2 \\ &< \text{dist}(x_{km}; S)^2 - \alpha\epsilon. \end{aligned}$$

Nesta última relação, fazendo  $k = 0, 1, \dots, \frac{N}{m}$  e somando as desigualdades, obtemos ainda que

$$\text{dist}(x_{N+m}; S)^2 \leq \text{dist}(x_0; S)^2 - \left(\frac{N}{m} + 1\right) \alpha\epsilon.$$

E assim temos

$$\left(\frac{N}{m} + 1\right) \alpha\epsilon \leq \text{dist}(x_0; S)^2$$

Um absurdo, pela forma como definimos  $N$ . ◊

Nesta próxima proposição, veremos que para o passo  $\alpha_k$  convergindo à zero, mas satisfazendo  $\sum_{k=0}^{\infty} \alpha_k = \infty$ , obtemos um resultado à respeito da convergência.

**Proposição 9.** *Seja  $\{x_k\}$  uma sequência gerada pelos algoritmos (3.8) – (3.10), com a ordem cíclica da relação de componentes e seja  $\alpha_k$  o passo satisfazendo*

$$\lim \alpha_k = 0, \quad \sum_{k=0}^{\infty} \alpha_k = \infty.$$

Então

$$\liminf_k F(x_k) = F^*.$$

Além disso, se  $S \neq \emptyset$  e

$$\sum_{k=0}^{\infty} \alpha_k^2 < \infty,$$

então  $\{x_k\}$  converge para algum  $x^* \in S$ .

*Demonstração:* Veja que para mostrar a primeira parte, é suficiente provar que  $\liminf F(x_{km}) = F^*$ , dessa forma suponha por absurdo que tal limite não seja válido. Então existe  $\epsilon > 0$  tal que

$$\liminf_k F(x_{km}) > F^* + 2\epsilon.$$

Assim, pela definição de ínfimo e da continuidade de  $F$  temos que existe  $\bar{y} \in C$  tal que

$$\liminf_k F(x_{km}) > F(\bar{y}) + 2\epsilon. \quad (3.31)$$

Seja  $k_0$  suficientemente grande, tal que para todo  $k \geq k_0$  temos que

$$F(x_{km}) \geq \liminf_k F(x_{km}) - \epsilon. \quad (3.32)$$

Combinando (3.31) e (3.32) tem-se

$$F(x_{km}) - F(\bar{y}) \geq \epsilon. \quad (3.33)$$

Usando (3.33) na Proposição 6 e fazendo  $y = \bar{y}$  obtemos

$$\begin{aligned} \|x_{(k+1)m} - \bar{y}\|^2 &\leq \|x_{km} - \bar{y}\|^2 - 2\alpha_{km} (F(x_{km}) - F(\bar{y})) + \alpha_{km}^2 \beta m^2 c^2 \\ &< \|x_{km} - \bar{y}\|^2 - \alpha_{km} (2\epsilon + \beta \alpha_{km} m^2 c^2). \end{aligned}$$

Temos que  $\alpha_k \rightarrow 0$  e como  $k_0$  é suficientemente grande, podemos assim assumi-lo de forma que

$$2\epsilon - \beta \alpha_k m^2 c^2 \geq \epsilon, \quad \forall k \geq k_0.$$

E ainda, para todo  $k \geq k_0$  obtemos

$$\begin{aligned} \|x_{(k+1)m} - \bar{y}\|^2 &\leq \|x_{km} - \bar{y}\|^2 - \epsilon \alpha_{km} \\ &\leq \|x_{(k-1)m} - \bar{y}\|^2 - \epsilon (\alpha_{km} + \alpha_{(k-1)m}) \\ &\leq \dots \\ &\leq \|x_{k_0 m} - \bar{y}\|^2 - \epsilon \sum_{l=k_0}^k \alpha_{lm}. \end{aligned}$$

Um absurdo, quando  $k \rightarrow \infty$ . Logo  $\liminf F(x_{km}) = F^*$ . Para provar a segunda parte, observe que pela Proposição 6, obtemos que para todo  $x^* \in S$  e  $k \geq 0$

$$\begin{aligned} \|x_{(k+1)m} - x^*\|^2 &\leq \|x_{km} - x^*\|^2 - 2\alpha_{km} (F(x_{km}) - F(x^*)) + \alpha_{km}^2 \beta m^2 c^2 \\ &\leq \|x_{km} - x^*\|^2 + \alpha_{km}^2 \beta m^2 c^2. \end{aligned}$$

Dessa forma, pela hipótese de  $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$ , pela proposição e definição de quase-Fejér convergência temos que  $\{x_{km}\}$  é limitada, logo possui uma subsequência convergente. Considere  $\{x_{k_j m}\}$  esta subsequência de  $\{x_{km}\}$  e  $\bar{x} = \lim \{x_{k_j m}\}$ . Assim,

$$F(\bar{x}) = \liminf_j F(x_{k_j m}) = \liminf_k F(x_{km}) = F^*,$$

garantindo que  $\bar{x} \in S$  e dessa forma obtemos que,  $\lim x_{km} = \bar{x}$ . Para mostrar que  $\{x_k\}$  converge à  $\bar{x}$ , note que para os algoritmos em questão provamos que  $\|x_{k+1} - x_k\| \leq 2\alpha_k c$  e como  $\{\alpha_k^2\}$  é somável temos que  $\|x_{k+1} - x_k\| \rightarrow 0$ . Como  $\{x_{km}\}$  converge para  $\bar{x}$ , temos que  $\{x_k\}$  também irá convergir para  $\bar{x}$ .  $\diamond$

# Capítulo 4

## Considerações Finais

Nesta dissertação, que teve por base os trabalhos de Bertsekas [4], vimos um método relativamente novo, voltado para problemas que envolvem a minimização da soma de um grande número de funções. Este método faz uma unificação de algoritmos: o algoritmo proximal incremental e o algoritmo do subgradiente incremental, mostrando assim, que ele pode ser bem conveniente, no caso onde algumas das funções nessa minimização forem mais propícias à serem usadas as iterações proximais e outras as iterações de subgradiente. Vimos também, algumas propriedades de convergência deste método. Um ponto interessante também, seria avaliar se a combinação de outros métodos proveria resultados semelhantes, para o problema de minimização em larga escala.



# Referências Bibliográficas

- [1] Auslender, A. ; Teboulle, M. Asymptotic Cones and Functions in Optimization and Variational Inequalities. Springer- Verlag, New York; (2003).
- [2] Ben-Tal, A.; Nemirovski, A. Optimization I-II. Technion-Israel Institute of Thecnology; (2004).
- [3] Bertsekas, D.P. Incremental Gradient, Subgradient, and Proximal Methods for Convex Optimization: A Survey. In: Labaratory for Information and Decision Systems Report LIDS-P-2848. MIT; (2010).
- [4] Bertsekas, D. P. Incremental Proximal Methods for Large Scale Convex Optimization. Lab. for Info. and Decision Systems Report LIDS-P-2847, MIT, Cambridge, MA. to appear in Math. Programming J; (2010).
- [5] Bertsekas, D. P. Convex Optimization Theory. Athena Scientific. Belmont, MA; (2009).
- [6] Bertsekas, D. P; Nedić, A.; Ozdaglar, A.Convex Analysis and Optimization. Athena Scientific. Belmont, MA; (2003).
- [7] Borwein, Jonathan M.; Lewis, Adrian S. Convex Analysis and Optimization. Spriger-Verlag, New York; (2000).
- [8] Combettes, P.L; Wajs, V.R. Signal Recovery by Proximal Forward-Backward Splitting. Multiscale Modeling and Simulation 4, pp. 1168-1200; (2005).
- [9] Greene, W. Econometric Analysis, 4<sup>a</sup> ed. Prentice Hall, Upper Sadle River, NJ; (2000).
- [10] Hiriart-Urruty,Jean-Baptiste; Lemaréchal, Claude. Convex Analysis and Minimization Algorithms I. Springer-Verlag; (1993).

- [11] Iusem, A. N., Métodos de Ponto Proximal em Otimização, 20° Colóquio Brasileiro de Matemática, IMPA, Rio de Janeiro; (1995).
- [12] Iusem, A. N.; Svaiter, B. F. A Proximal Regularization of the Steepest Descent Method, vol 29, n°2, pp.123-130; (1995).
- [13] Izmailov, A.; Solodov, M. Otimização, Volume 1, Rio de Janeiro: IMPA; (2005).
- [14] Izmailov; A., Solodov, M. Otimização, Volume 2, Rio de Janeiro, IMPA; (2007).
- [15] Kibardin, V. M. Decomposition into Functions in the Minimization Problem. Automation and Remote Control, Vol. 40, pp. 1311-1323; (1980).
- [16] Lima, Elon L.: Análise Real, Volume 2, Funções de  $n$  várias variáveis. 11ª ed. Rio de Janeiro: IMPA; (2012).
- [17] Lima, Elon L., Curso de Análise, Volume 2. Projeto Euclides. IMPA, Rio de Janeiro, Brasil; (2009).
- [18] Lima, Ronaldo F. de, Topologia e Análise no Espaço . Ronaldo Freire de Lima. Rio de Janeiro: SBM; (2015).
- [19] Martinet, B. Régularisation d'Inéquations Variationelles Par Approximations Successives. Revue Française. d'Infomatique et Recherche Opérationelle 4, pp. 154-159; (1970).
- [20] Nedić, A.; Bertsekas, D.P. Incremental Subgradient Methods for Nondifferentiable Optimization. SIAM J. Optim. Vol 12, pp. 109-138; (2001).
- [21] Rockafellar ,R.T. Convex Analysis. Princeton University Press, Princeton, New Jersey; (1970).
- [22] Rockafellar, R.T. Monotone Operators and the Proximal Point Algorithm. SIAM J. Control.Optim. Vol 14, pp. 877-898; (1976).
- [23] S, Ma. Alternating Proximal Gradient Method for Convex Minimization. Disponível em [http://www.optimization-online.org/DB\\_FILE/2012/09/3608.pdf](http://www.optimization-online.org/DB_FILE/2012/09/3608.pdf); (2012).
- [24] Shor, N.Z. Minimization Methods for Nondifferentiable Functions, Springer-Verlag, Berlim; (1985).

- 
- [25] Solodov, M.V.; Zavriev, S.K. Error Stability Properties of Generalized Gradient-Type Algorithms. *J. Opt. Theory Appl.* Vol 98, pp. 663-680; (1998).
- [26] Sra, Suvrit; Nowozin, Sebastian ; Wright, Stephen J. *Optimization for Machine Learning*. Massachusetts; (2012).
- [27] Tiel, Jan van. *Convex Analysis*; (1984).
- [28] Rockafellar, R. T; Wets ,R.J.B. *Variational Analysis*. Springer-Verlag, New York; (1997).
- [29] Zvi, Drezner; Hamacher, Horst W. *Facility Location: Applications and Theory*. Springer- Verlag, Berlin. (2004).