UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

VANDO ANTÔNIO ADONA

Método Subgradiente Incremental para Otimização Convexa não Diferenciável

Goiânia
2014

sistema de bibliotecas ufg

UFG

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR AS TESES E DISSERTAÇÕES ELETRÔNICAS (TEDE) NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico: **[x] Dissertação** **[ ] Tese**

2. Identificação da Tese ou Dissertação

Autor (a):	Vando Antônio Adona				
E-mail:	vandomat32@hotmail.com				
Seu e-mail pode ser disponibilizado na página? [x] Sim [ ] Não					
Vínculo empregatício do autor	-				
Agência de fomento:	Conselho Nacional de Desenvolvimento Científico e Tecnológico	Sigla:	CNPq		
País:	Brasil	UF:	DF	CNPJ:	33.654.831
Título:	Método Subgradiente Incremental para Otimização Convexa não Diferenciável				
Palavras-chave:	Método subgradiente incremental, Otimização convexa, Otimização não diferenciável.				
Título em outra língua:	Incremental Subgradient Method for Nondifferentiable Convex Optimization				
Palavras-chave em outra língua:	Incremental subgradient method, Convex optimization, Nondifferentiable optimization.				
Área de concentração:	Otimização				
Data defesa: (dd/mm/aaaa)	18/12/2014				
Programa de Pós-Graduação:	Mestrado em Matemática				
Orientador (a):	Dr. Jefferson Divino Gonçalves de Melo				
E-mail:	jefferson@mat.ufg.br				
Co-orientador(a):*	-				

*Necessita do CPF quando não constar no SisPG

3. Informações de acesso ao documento:

Concorda com a liberação total do documento [x] SIM [ ] NÃO[1]

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF ou DOC da tese ou dissertação.

O sistema da Biblioteca Digital de Teses e Dissertações garante aos autores, que os arquivos contendo eletronicamente as teses e ou dissertações, antes de sua disponibilização, receberão procedimentos de segurança, criptografia (para não permitir cópia e extração de conteúdo, permitindo apenas impressão fraca) usando o padrão do Acrobat.

Assinatura do (a) autor (a)

Data: 18 / 12 / 2014

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.

Vando Antônio Adona

Método Subgradiente Incremental para Otimização Convexa não Diferenciável

Dissertação apresentada ao Programa de Pós–Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Mestre em Matemática.

Área de concentração: Otimização.

Orientador: Prof. Dr. Jefferson Divino Gonçalves de Melo

Goiânia
2014

Ficha catalográfica elaborada automaticamente
com os dados fornecidos pelo(a) autor(a), sob orientação do Sibi/UFG.

Adona, Vando Antônio
Método Subgradiente Incremental para Otimização Convexa não Diferenciável [manuscrito] / Vando Antônio Adona. - 2014.
66 f.

Orientador: Prof. Dr. Jefferson Divino Gonçalves de Melo.
Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto de Matemática e Estatística (IME) , Programa de Pós-Graduação em Matemática, Goiânia, 2014.
Bibliografia.
Inclui tabelas, algoritmos.

1. Método subgradiente incremental. 2. Otimização convexa. 3. Otimização não diferenciável. I. Melo, Jefferson Divino Gonçalves de, orient. II. Título.

VANDO ANTÔNIO ADONA

MÉTODO SUBGRADIENTE INCREMENTAL PARA OTIMIZAÇÃO CONVEXA NÃO DIFERENCIÁVEL

Dissertação defendida no Programa de Pós-Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás como requisito parcial para obtenção do título de Mestre em Matemática, aprovada no dia 18 de dezembro de 2014, pela Banca Examinadora constituída pelos professores:

Prof. Dr. Jefferson Divino Gonçalves de Melo
Instituto de Matemática e Estatística-UFG
Presidente da Banca

Prof. Dr. Max Leandro Nobre Gonçalves
Instituto de Matemática e Estatística-UFG

Prof. Dr. Gabriel Haeser
Instituto de Matemática e Estatística-USP

Prof. Dr. Jorge Barrios Ginart
Instituto de Matemática e Estatística-UFG

Vando Antônio Adona

Graduou-se em licenciatura em Matemática, no final do ano de 2012, pela Universidade Estadual de Goiás - UEG. Durante a graduação foi monitor de Cálculo I e Cálculo II. No ano de 2013 iniciou o Mestrado em Matemática na Universidade Federal de Goiás - UFG, onde foi bolsista do CNPq.

Agradecimentos

Primeiramente a Deus, por todas as conquistas que já me fez alcançar, por em nenhum momento sequer durante meus estudos me deixar faltar saúde, paz e nem disposição na busca do conhecimento, por sempre me dar esperança mesmo nas horas mais difíceis, por iluminar meu caminho a cada dia, e por me dar a certeza de que tudo que passamos em nossa vida, todos nossos esforços, são recompensados.

Agradeço a minha família, minha mãe Neosa e minhas irmãs Vanessa e Vânia, que sempre me apoiaram em todas as minhas decisões e nunca me deixou desencorajar, sempre incentivando meus estudos. Também agradeço a meu pai Dirceu, que não está mais aqui na terra, mas me ensinou muito enquanto esteve presente.

A todos os professores do IME/UFG, que são excelentes profissionais, em especial ao meu professor orientador Dr. Jefferson Divino Gonçalves de Melo, por dedicar bastante tempo compartilhando seu conhecimento comigo, sempre com muita paciência e disposição. Agradeço também ao professor pesquisador Dr. Jorge Barrios Ginart profissional muito generoso e que contribiu imensamente neste trabalho.

A todos meus colegas do curso de mestrado, com os quais aprendi muito e fiz grandes amizades.

Ao (CNPq) pela bolsa de estudos, que me auxiliou muito nesta jornada de estudos.

Agradeço também aos professores Dr. Gabriel Haeser e Dr. Max Leandro Nobre Gonçalves por suas participações na banca, pela leitura da dissertação e por suas sugestões para melhoria deste trabalho.

Resumo

Adona, Vando Antônio. **Método Subgradiente Incremental para Otimização Convexa não Diferenciável**. Goiânia , 2014. 66p. Dissertação de Mestrado. Instituto de Matemática e Estatística, Universidade Federal de Goiás.

Consideramos um problema de otimização cuja função objetivo consiste na soma de funções convexas, não necessariamente diferenciáveis. Estudamos um método subgradiente que executa a iteração de forma incremental, selecionando cada função componente de maneira sequencial e processando a iteração subgradiente individualmente. Analisamos diferentes alternativas para a escolha do comprimento de passo, destacando as propriedades de convergência para cada caso. Abordamos também o modelo incremental em outros métodos, considerando iteração proximal e combinações de iterações subgradiente e proximal. Esta abordagem incremental tem sido muito bem sucedida quando o número de funções componentes é grande.

Palavras–chave

Método subgradiente incremental, Otimização convexa, Otimização não diferenciável.

Abstract

Adona, Vando Antônio. **Incremental Subgradient Method for Nondifferentiable Convex Optimization**. Goiânia , 2014. 66p. MSc. Dissertation. Instituto de Matemática e Estatística, Universidade Federal de Goiás.

We consider an optimization problem for which the objective function is the sum of convex functions, not necessarily differentiable. We study a subgradient method that executes the iterations incrementally selecting each component function sequentially and processing the subgradient iteration individually. We analyze different alternatives for choosing the step length, highlighting the convergence properties for each case. We also analyze the incremental model in other methods, considering proximal iteration and combinations of subgradient and proximal iterations. This incremental approach has been very successful when the number of component functions is large.

Keywords

Incremental subgradient method, Convex optimization, Nondifferentiable optimization.

Sumário

CAPÍTULO 1

Introdução

Ao longo deste trabalho estudaremos métodos para resolver o seguinte problema de otimização

$$\min f(x) := \sum_{i=1}^{m} f_i(x), \quad \text{sujeito a} \quad x \in X, \tag{1-1}$$

onde $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \dots, m$ são funções convexas (não necessariamente diferenciáveis), e $X \subseteq \mathbb{R}^n$ é um conjunto não vazio, fechado e convexo. Para resolver o problema acima consideraremos o método do subgradiente incremental o qual pode ter melhor desempenho, comparado com o método clássico, quando o número de componentes m é muito grande. O método subgradiente incremental é um processo iterativo que gera uma sequência $\{x^k\}$ de pontos em $\mathbb{R}^n$, onde cada iteração x^k é obtida após m subiterações, sendo que, uma subiteração é calculada em relação a uma simples componente f_i. Portanto esta é a principal diferença entre o método incremental e o método do subgradiente convencional, o qual calcula cada iteração x^k em termos da função objetivo f.

Problemas na forma (1-1) aparecem em vários contextos e aplicações. Em seguida apresentaremos alguns exemplos onde o método incremental pode ser vantajoso.

Exemplo 1.1 (Mínimos Quadrados) *Consideraremos dois casos:*

1. *Regularização diferenciável*

$$\min f(x) := \sum_{i=1}^{m} \left(\langle a_i, x\rangle + b_i\right)^2 + \lambda \|x - \bar{x}\|^2, \quad \text{sujeito a} \quad x \in \mathbb{R}^n,$$

onde a_i e b_i são vetores em $\mathbb{R}^n$, para todo $i = 1, \dots, m$, λ é um número real e $\bar{x} \in \mathbb{R}^n$ é um vetor fixado;

2. *Regularização não diferenciável*

$$\min f(x) := \sum_{i=1}^{m} \left(\langle a_i, x\rangle + b_i\right)^2 + \lambda \sum_{j=1}^{n} |x_j|, \quad \text{sujeito a} \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n,$$

Exemplo 1.2 (Problema de Fermat-Weber) *Um problema básico em teoria local consiste em encontrar um ponto x cuja soma das distâncias ponderadas de x a um dado conjunto de pontos $y_1, \ldots, y_m$ é minimizado, ou seja,*

$$\min f(x) := \sum_{i=1}^{m} \lambda_i \|x - y_i\|, \quad \text{sujeito a} \quad x \in \mathbb{R}^n,$$

onde $\lambda_1, \ldots, \lambda_m$ são números positivos dados. Para mais detalhes sobre o problema de Fermat-Weber veja, por exemplo, [6].

Exemplo 1.3 (Problema dual) *Consideremos o problema*

$$\max \sum_{i=1}^{m} h_i(y_i), \quad \text{sujeito a} \quad \sum_{i=1}^{m} g_i(y_i) \geq 0, \quad y_i \in Y_i, \;\; i = 1, \ldots, m.$$

onde $h_i : \mathbb{R}^l \to \mathbb{R}$ e $g_i : \mathbb{R}^l \to \mathbb{R}^n$ são funções, $y_i \in \mathbb{R}^l$ e $Y_i \subseteq \mathbb{R}^l$. Então, considerando o vetor $x \in \mathbb{R}^n$ como multiplicador, em relação às funções restrições n-dimensionais, obtemos o seguinte problema dual

$$\min \sum_{i=1}^{m} f_i(x), \quad \text{sujeito a} \quad x \geq 0.$$

onde

$$f_i(x) = \sup_{y_i \in Y_i} \{h_i(y_i) + \langle x, g_i(y_i) \rangle\},$$

que possui a forma (1-1). Observe que as funções componentes f_i são supremos de funções afins, logo são convexas.

Um dos métodos utizados para solucionar o problema (1-1) é o método do subgradiente

$$x^{k+1} = P_X\left(x^k - \alpha_k \sum_{i=1}^{m} d^{i,k}\right) \tag{1-2}$$

onde $d^{i,k}$ é o subgradiente de f_i em x^k, $\alpha_k > 0$ e P_X é a projeção no conjunto X.

O método do subgradiente incremental apresentado em [14] por Nedić e Bertsekas, é semelhante ao método do subgradiente (1-2). A principal diferença é que em cada iteração, o vetor x^k é atualizado incrementalmente, atravéz de uma sequência de m passos. Cada passo é uma iteração subgradiente para uma única função componente f_i, e existe um passo por função componente. Assim uma iteração pode ser visto como um ciclo de m subiterações. Se x^k é um vetor obtido após k ciclos, o vetor x^{k+1} obtido após mais um ciclo é

$$x^{k+1} = \psi^{m,k} \tag{1-3}$$

onde $\psi^{m,k}$ é obtido após m passos

$$\psi^{i,k} = P_X\left(\psi^{i-1,k} - \alpha_k g^{i,k}\right), \quad g^{i,k} \in \partial f_i(\psi^{i-1,k}) \quad i = 1,\dots,m \tag{1-4}$$

iniciando com

$$\psi^{0,k} = x^k, \tag{1-5}$$

onde $\partial f_i(\psi^{i-1,k})$ denota o subdiferencial (conjunto de todos os subgradientes) de f_i no ponto $\psi^{i-1,k}$ (veja Definição 2.24). A atualização descrita em (1-4) se refere as subiterações do ciclo k.

O método do subgradiente incremental que apresentamos aqui, possui comportamento semelhante ao método do gradiente incremental, onde análises de convergência foram também apresentadas por Bertsekas em [1]. Estes métodos foram primeiramente estudados por Kibardin [11], e mais recentemente por Solodov e Zavriev [16], Nedić e Bertsekas [14], entre outros.

Destacamos que o método subgradiente incremental geralmente converge mais rápido que o método subgradiente (1-2), quando o ponto inicial é tomado distante do eventual limite. Contudo, próximo ao possível ponto limite de convergência, ele tipicamente converge lentamente pois objetiva minimizar cada componente separadamente. No caso especial quando todos os pontos estacionários de f são também pontos estacionários de todas as funções componentes f_i, o ciclo limite pode (dependendo da escolha de α_k) reduz-se a um único ponto, onde obtem-se a convergência. Por exemplo, se no problema (1-1), tivermos $X = \mathbb{R}$ e $f = \sum_{i=1}^{m} f_i$, onde $f_i : \mathbb{R} \to \mathbb{R}$ é dada por $f_i(x) = ix^2$, assim f_i é diferenciável, então o subgradiente será a derivada de f_i em qualquer ponto (ver Proposição 2.25). Logo para qualquer ponto $x^k \in \mathbb{R}$, se escolhermos $\alpha_k = 1/2$, então as subiterações do ciclo k fornecem todas o mesmo valor $x^* = 0$ que é o ponto ótimo solução do problema (1-1). No entanto, em geral, os pontos correspondentes às subiterações (1-4), são distintos.

Para melhor compreenssão desse trabalho, no Capítulo 2, faremos uma revisão de alguns conceitos básicos, como resultados de análise no espaço $\mathbb{R}^n$ e definições de sequência Féjer e quasi-Féjer convergente. Apresentaremos também alguns fatos essenciais no estudo de otimização, em especial tópicos sobre análise convexa.

No Capítulo 3, apresentaremos detalhadamente o método subgradiente incremental. Estudaremos propriedades de convergência para três tipos de regras de tamanho de passo: regra de tamanho de passo constante; regra de tamanho de passo pré-determinado, onde $\alpha_k \to 0$; e uma regra de tamanho de passo dinâmico, onde α_k depende do valor ótimo da função objetivo.

No Capítulo 4 examinaremos as propriedades do método incremental em outras situações. Adaptaremos o modelo incremental para o método do ponto proximal, denominado método proximal incremental. O objetivo deste tópico é fornecer, em caráter introdutório, o método proximal (também para o caso incremental) para que na seção 4.2, possamos apresentar o método incremental subgradiente-proximal, que é uma combinação de uma iteração proximal e uma subgradiente. Na seção 4.3 estudaremos alguns resultados de convergência sob a sequência $\{x^k\}$, com ordem cíclica de seleção de componente, gerada pelo método incremental subgradiente-proximal.

No Capítulo 5 faremos nossas considerações finais, destacando os resultados obtidos e sugerindo outros estudos.

CAPÍTULO 2

Conceitos básicos

Neste capítulo descreveremos, em carater introdutório, alguns conceitos que serão exigidos para leitura deste trabalho. Algumas demonstrações dos resultados apresentados serão omitidas, mas indicaremos onde podem ser encontradas. Inicialmente faremos uma revisão sobre conceitos de análise real e análise no espaço euclidiano $\mathbb{R}^n$, em particular definiremos sequência Féjer e quasi-Féjer convergente, assuntos bastante utilizados em otimização. Vários destes resultados podem ser obtidos em livros de análise, destacamos entre eles nossas referências utilizadas [12, 13]. Apresentaremos alguns tópicos de otimização, com ênfase a análise convexa, abordando assuntos como operador projeção, funções convexas não diferenciáveis, entre outros. Neste capítulo também utilizamos a referência [9].

2.1 Resultados básicos de Análise

Em várias ocasiões, no decorrer desse trabalho, utilizaremos as propriedades do ínfimo e do supremo de um conjuto, então é conveniente relembrar. Dado um conjunto $X \subset \mathbb{R}$ limitado, um elemento $a \in \mathbb{R}$ chama-se ínfimo do conjunto X se satisfaz as seguintes condições:

I1. Para todo $x \in X$ tem-se $a \leq x$;
I2. Dado $c \in \mathbb{R}$ com $a < c$, então existe $x \in X$, tal que $x < c$.

Também, por definição, um elemento $b \in \mathbb{R}$ é dito supremo do conjunto X, se podemos verificar as seguintes condições:

S1. Para todo $x \in X$, tem-se $x \leq b$;
S2. Dado $d \in \mathbb{R}$ com $d < b$, existe $x \in X$ tal que $d < x$.

O ínfimo a e o supremo b do conjunto X são únicos, e denotaremos esses números, respectivamente, por $a = \inf X$ e $b = \sup X$.

Relembraremos também as definições de limite superior e limite inferior. Seja $\{x_k\}$ uma sequência limitada em $\mathbb{R}$. Logo existem $\alpha, \beta \in \mathbb{R}$ satisfazendo $x_k \in [\alpha, \beta]$ para todo $k \in \mathbb{N}$. Escrevendo $X_k = \{x_k,\ x_{k+1},\ \ldots\}$. Temos que $[\alpha,\beta] \supseteq X_1 \supseteq X_2 \supseteq \cdots \supseteq X_k \supseteq \cdots$. Logo, definindo $a_k = \inf X_k$ e $b_k = \sup X_k$, obtemos

$$\alpha \leq a_1 \leq a_2 \leq \cdots \leq a_k \leq \cdots \leq b_k \leq \cdots \leq b_2 \leq b_1 \leq \beta.$$

Portanto, como as sequências $\{a_k\}$ e $\{b_k\}$ são monótonas e limitadas, é possível mostrar que existem os limites

$$a = \lim_{k\to\infty} a_k = \sup a_k = \sup_k \inf X_k = \sup_k \inf \left\{x_j; j \geq k\right\},$$

$$b = \lim_{k\to\infty} b_k = \inf b_k = \inf_k \sup X_k = \inf_k \sup \left\{x_j; j \geq k\right\}.$$

Dizemos que a é o limite inferior e b o limite superior da sequência $\{x_k\}$ e escreveremos $a = \liminf\limits_{k\to\infty} x_k$, $b = \limsup\limits_{k\to\infty} x_k$. Certamente temos que

$$\liminf_{k\to\infty} x_k \leq \limsup_{k\to\infty} x_k.$$

Aprensentaremos agora alguns conceitos básicos de análise no espaço euclidiano $\mathbb{R}^n$. Abordaremos assuntos relacionados a topologia como produto interno, norma e sequência.

Definição 2.1 *Um produto interno no espaço euclidiano $\mathbb{R}^n$ é uma função real definida em $\mathbb{R}^n \times \mathbb{R}^n$, que é bilinear, simétrica e definida positiva, onde a cada par de vetores $x,y \in \mathbb{R}^n$ faz corresponder um número real indicado por $\langle x,y\rangle$. Então, para quaisquer $x,y,z \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$ um produto interno deve satisfazer as seguintes propriedades:*

P1. $\langle x,y\rangle = \langle y,x\rangle$;
P2. $\langle x+y,z\rangle = \langle x,z\rangle + \langle y,z\rangle$;
P3. $\langle \alpha x,y\rangle = \alpha\langle x,y\rangle = \langle x,\alpha y\rangle$;
P4. $x \neq 0 \Rightarrow \langle x,x\rangle > 0$.

Neste trabalho utilizaremos o produto interno usual, isto é, dado $x = (x_1,\ldots,x_n)$, $y = (y_1,\ldots,y_n)$ vetores em $\mathbb{R}^n$

$$\langle x,y\rangle = \sum_{i=1}^{n} x_i y_i = x_1y_1 + x_2y_2 + \cdots + x_ny_n. \tag{2-1}$$

E fácil ver que essa relação é de fato um produto interno, ou seja, satisfaz as propriedades da definição (2.1).

Definição 2.2 *Uma norma no espaço $\mathbb{R}^n$ é uma função real positiva $\|.\| : \mathbb{R}^n \to \mathbb{R}$ que faz corresponder a um vetor $x \in \mathbb{R}^n$ um número real não negativo $\|x\|$, e que a cada par de vetores $x,\ y \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$ cumpre as seguintes condições:*

N1. $\|x+y\| \leq \|x\|+\|y\|$;
N2. $\|\alpha x\| = |\alpha|\,\|x\|$;
N3. $x \neq 0 \Rightarrow \|x\| > 0$.

Dado $x \in \mathbb{R}^n$, algumas das normas utilizadas em $\mathbb{R}^n$ são, $\|x\|_1 = |x_1|+\cdots+|x_n|$, $\|x\|_\infty = \max\{|x_i|\,; i=1,\ldots,n\}$ e $\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$. Pode ser mostrado que estas normas são equivalentes, no seguinte sentido, se uma sequência converge em uma delas, converge em todas as outras. Neste trabalho usaremos a norma euclidiana, ou seja, a norma proveniente do protudo interno (2-1), $\|x\| = \sqrt{\langle x,x\rangle}$.

Teorema 2.3 (Desigualdade de Cauchy-Schwarz) *Dados $x,y \in \mathbb{R}^n$ quaisquer, então vale a seguinte expressão*

$$|\langle x,y\rangle| \leq \|x\|\,\|y\|\,, \quad \forall\, x,y \in \mathbb{R}^n,$$

ocorrendo a igualdade se, e somente se, um dos vetores x,y é um múltiplo escalar do outro, ou seja, se existe $\alpha \in \mathbb{R}$, tal que $x = \alpha y$.

Prova. Ver página 5 de [13]. □

Destacaremos a seguir alguns fatos básicos sobre topologia no espaço euclidiano $\mathbb{R}^n$ que são bastante utilizados em otimização.

Sejam $a \in \mathbb{R}^n$ e $\delta > 0$ um número real. A bola aberta $B(a;\delta)$ de centro a e raio δ é o conjunto dos pontos $x \in \mathbb{R}^n$, cuja distância ao ponto a é menor que δ. Isto é,

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

Analogamente, definimos $B[a;\delta] = \{x \in \mathbb{R}^n; \|x-a\| \leq \delta\}$ como a bola fechada de centro a e raio δ.

Seja $U \subseteq \mathbb{R}^n$. Um ponto $a \in U$ chama-se ponto interior ao conjunto U, quando é centro de alguma bola aberta inteiramente contida em U. E dizemos que um conjunto $U \subseteq \mathbb{R}^n$ é um conjunto aberto, quando todos os pontos de U são pontos interiores a U. Obviamente, toda bola aberta é um conjunto aberto. Também, o espaço $\mathbb{R}^n$ é um conjunto aberto.

Um conjunto $X \subseteq \mathbb{R}^n$ é dito ser fechado se seu complentar $X^c = \mathbb{R}^n - X$ é aberto em $\mathbb{R}^n$. Dizemos também que um conjunto $K \subset \mathbb{R}^n$ é compacto quando ele for fechado e limitado.

Em seguida apresentaremos um importante teorema sobre sequências limitadas, devido a Bolzano-Weiertrass.

Teorema 2.4 (Teorema de Bolzano-Weierstrass) *Toda sequência limitada em $\mathbb{R}^n$ possui uma subsequência convergente.*

Prova. Ver página 17 de [13] □

Estabeleceremos agora alguns resultados que serão utilizados posteriormente na análise de convergência do método subgradiente incremental. Iniciamos com a noção de sequência quasi-Féjer convergente, para mais detalhes veja, por exemplo, [7, 8].

Definição 2.5 *Uma sequência $\{y^k\}$ em $\mathbb{R}^n$ é quasi-Fejér convergente a um conjunto $U \subset \mathbb{R}^n$ se para todo $u \in U$ existir uma sequência $\{\xi_k\}$ em $\mathbb{R}$ tal que*

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

Proposição 2.6 *Seja $\{y^k\}$ uma sequência em $\mathbb{R}^n$. Se $\{y^k\}$ é quasi-Féjer convergente a um conjunto não vazio $U \subset \mathbb{R}^n$, então $\{y^k\}$ é limitada. Se um ponto de acumulação y de $\{y^k\}$ pertence a U então $y = \lim\limits_{k \to \infty} y^k$.*

Prova. Seja $u \in U$, pela Definição 2.5 a sequência $\{y^k\}$ satisfaz

$$\|y^k - u\|^2 \leq \|y^0 - u\|^2 + \sum_{j=0}^{k-1} \xi_j \leq \|y^0 - u\|^2 + \sum_{j=0}^{\infty} \xi_j < \infty, \qquad k \in \mathbb{N}.$$

Então a sequência $\{y^k\}$ é limitada. Considere agora que $y \in U$ seja um ponto de acumulação de $\{y^k\}$, e $\{y^{k_l}\}$ a subsequência de $\{y^k\}$ tal que $\lim_{l \to \infty} y^{k_l} = y$. Dado $\varepsilon > 0$, como $y \in U$ existe uma sequência $\{\xi_k\}$ em $\mathbb{R}$ satisfazendo as propriedades da Definição 2.5. Então existe k_0 tal que

$$\sum_{k=k_0}^{\infty} \xi_k < \frac{\varepsilon}{2}.$$

Como $y^{k_l} \to y$ $(l \to \infty)$ podemos encontrar $\bar{l}$ tal que

$$k_{\bar{l}} \geq k_0 \qquad \text{e} \qquad \|y^{k_{\bar{l}}} - y\|^2 < \frac{\varepsilon}{2},$$

então para qualquer $k > k_{\bar{l}}$ temos

$$\|y^k - y\|^2 \leq \|y^{k_{\bar{l}}} - y\|^2 + \sum_{i=k_{\bar{l}}}^{k-1} \xi_i \leq \|y^{k_{\bar{l}}} - y\|^2 + \sum_{i=k_0}^{\infty} \xi_i < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Como ε foi escolhido arbitrariamente, segue que $\lim\limits_{k \to \infty} y^k = y$. □

Se $\{y^k\}$ satisfaz a Definição 2.5, com a sequência $\{\xi_k\}$ identicamente nula para todo $u \in U$, então para todo $u \in U$ temos $\|y^{k+1} - u\| \leq \|y^k - u\| \quad \forall\, k \geq 0$, e neste caso dizemos que a sequência $\{y^k\}$ é Féjer convergente.

2.2 Tópicos em otimização

Nesta seção discutiremos alguns conceitos básicos de otimização, como condições de otimalidade para probemas de otimização no caso diferenciável e não diferenciável.

Considere uma função $f : \mathbb{R}^n \to \mathbb{R}$. No estudo de otimização, um dos objetivos é encontrar um minimizador de f em $X \subseteq \mathbb{R}^n$. Esse é um problema restrito exposto da seguinte forma

$$\min f(x), \quad x \in X. \tag{2-2}$$

O conjunto X é chamado conjunto viável, e a função f é dita ser função objetivo do problema (2-2).

Definição 2.7 *Dada uma função $f : \mathbb{R}^n \to \mathbb{R}$ e um conjunto $X \subseteq \mathbb{R}^n$. Dizemos que um ponto $\bar{x} \in X$ é minimizador local do problema (2-2), se existir $\varepsilon > 0$, tal que,*

$$f(\bar{x}) \leq f(x), \quad \forall\, x \in X \cap B[\bar{x};\varepsilon]. \tag{2-3}$$

E dizemos que $\bar{x}$ é minimizador global do problema (2-2), se

$$f(\bar{x}) \leq f(x), \quad \forall\, x \in X. \tag{2-4}$$

Se $\bar{x}$ é um minimizador global, o valor $f(\bar{x})$ é o valor ótimo. Quando no problema (2-2) considerarmos $X = \mathbb{R}^n$ diremos que temos um problema irrestrito.

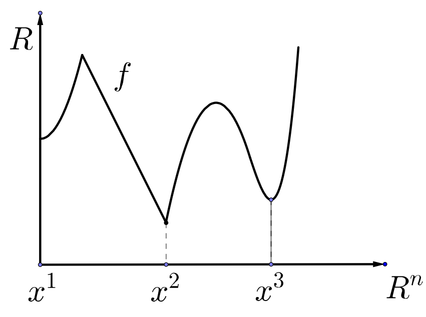

Figura 2.1: *x^1 e x^3 são minimizadores locais e x^2 é o minimizador global do problema (2-2), onde $X = \{x \in \mathbb{R}^n \,|\, x \geq 0\}$.*

Se uma função $f : \mathbb{R}^n \to \mathbb{R}$ é diferenciável num ponto $\bar{x} \in \mathbb{R}^n$, e se $\bar{x}$ satisfaz a condição

$$\nabla f(\bar{x}) = 0,$$

dizemos que $\bar{x}$ é um ponto estacionário ou crítico. Os pontos estacionários são os candidatos a serem minimizadores do problema (2-2). Como mostra o seguinte resultado.

Teorema 2.8 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável num ponto $\bar{x} \in \mathbb{R}^n$. Se $\bar{x}$ é um minimizador de f, então*

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

Prova. Veja página 16 de [9]. □

É importante ressaltar que a recíproca do Teorema 2.8 não é verdadeira. Por exemplo, a função $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = x^3$ satisfaz $f'(0) = 0$, mas o ponto $x = 0$ não é minimizador de f.

É possível mostrar que toda função $f : U \to \mathbb{R}$ definida em um aberto $U \subseteq \mathbb{R}^n$, diferenciável em um ponto, é contínua nesse ponto.

Para funções contínuas, definidas em conjunto fechado e limitado de $\mathbb{R}^n$, o próximo resultado mostra uma condição de existência de pontos ótimos.

Teorema 2.9 (Weierstrass) *Toda função real contínua $f : K \to \mathbb{R}$ definida num conjunto compacto $K \subset \mathbb{R}^n$, atinge seu máximo e seu mínimo em K, isto é, existem pontos $x_0, x_1 \in K$ tais que $f(x_0) \leq f(x) \leq f(x_1)$ para qualquer $x \in K$.*

Prova. Ver página 45 de [13]. □

O conjunto de nível de uma função será utilizado posteriormente para fornecer uma caracterização das função convexa.

Definição 2.10 *O conjunto de nível da função $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ associado a $c \in \mathbb{R}$, é o conjunto dado por*

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

2.2.1 Análise convexa

Iniciamos esta subseção com a definição de conjunto convexo. Em termos geométricos, um conjunto $D \subseteq \mathbb{R}^n$ é convexo se, qualquer que seja os pontos $x, y \in D$ o segmento de reta que une esses pontos está inteiramente contido em D. De forma precisa, a definição é a seguinte.

Definição 2.11 *Um conjunto $D \subset \mathbb{R}^n$ é chamado convexo, se para todo $x, y \in D$ e $\lambda \in [0, 1]$ tivermos*

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

O espaço euclidiano $\mathbb{R}^n$ e qualquer bola $B(a; \delta)$ são exemplos de conjuntos convexos.

Uma característica elementar sobre conjuntos convexos é que a interseção de conjuntos convexos é ainda um conjunto convexo.

Proposição 2.12 *Se $D_i \subset \mathbb{R}^n$ são conjuntos convexos para todo $i \in I$, onde I é um conjunto de índices, então a interseção $D = \cap_{i \in I} D_i$ também é um conjunto convexo.*

Prova. Sejam $x, y \in D$ e $\lambda \in [0, 1]$. Pela definição de interseção de conjuntos temos que $x, y \in D_i$ para todo $i \in I$. Como os conjuntos D_i são convexos para todo $i \in I$, obtemos $\lambda x + (1-\lambda)y \in D_i$ para todo $i \in I$. Novamente pela definição de interseção de conjuntos, segue que $\lambda x + (1-\lambda)y \in D$, ou seja, D é convexo. □

Definição 2.13 *Seja $D \subset \mathbb{R}^n$ um conjunto qualquer. O fecho convexo de D, denotado por* conv D*, é o menor conjunto convexo em $\mathbb{R}^n$ que contém D, ou seja,* conv D *é a interseção de todos os conjuntos convexos em $\mathbb{R}^n$ que contêm D.*

Segue da Proposição 2.12 que conv D é um conjunto convexo para qualquer conjunto $D \subset \mathbb{R}^n$. Além disso é óbvio que se D é convexo, então conv $D = D$

Uma classe de funções onde todo ponto estacionário é solução do problema (2-2) é a classe das funções convexas.

Definição 2.14 (Função Convexa) *Dada uma função $f : D \to \mathbb{R}$ definida no conjunto convexo $D \subset \mathbb{R}^n$. Diz-se que f é convexa em D, se para todo $\lambda \in [0, 1]$ tivermos:*

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y), \quad \forall\, x, y \in D.$$

As vezes dizemos apenas que f é convexa, ficando subentendido que f é uma função convexa em todo seu domínio. Observa-se da definição anterior, que o gráfico de uma função convexa deve estar localizado abaixo ou sobre o segmento de reta que liga os pontos $(x, f(x))$ e $(y, f(y))$, $\forall\, x, y \in D$.

(a) *f é convexa* (b) *f não é convexa*

Figura 2.2: *A função em (b) não é convexa pois o segmento pontilhado está abaixo do gráfico de f.*

Toda função convexa possui conjunto de nível convexo para todo número real, isso é o que afirma o próximo resultado, mas vale destacar que a recíproca não é verdadeira.

Teorema 2.15 *Considere $D \subseteq \mathbb{R}^n$ um conjunto convexo. E seja $f : D \to \mathbb{R}$ convexa. Então, para todo número real c, o conjunto de nível $L_{f,D}(c)$ é convexo.*

Prova. Seja $c \in \mathbb{R}$ um número qualquer fixo. Se $L_{f,D} = \emptyset$, o teorema estará demonstrado (o conjunto vazio é convexo trivialmente). Considere agora que o conjunto de nível de f associado a $c \in \mathbb{R}$ seja diferente do vazio. Sejam $x, y \in L_{f,D}$ então $x, y \in D$, $f(x) \leq c$ e $f(y) \leq c$. Pela convexidade de D, $\lambda x + (1-\lambda)y \in D$, para qualquer $\lambda \in [0,1]$. Pela convexidade de f em D

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

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

Com o objetivo de caracterizar as funções convexas, definiremos epígrafo de uma função, para em seguida apresentar um teorema que relaciona convexidade de conjuntos e funções.

Definição 2.16 *O epígrafo da função $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ é o conjunto em $\mathbb{R}^n \times \mathbb{R}$ dado por*

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

Teorema 2.17 *Seja $D \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : D \to \mathbb{R}$ é convexa se, e somente se, o epígrafo de f é um conjunto convexo em $\mathbb{R}^n \times \mathbb{R}$.*

Prova. Suponhamos inicialmente que o epígrafo E_f seja convexo. Tomemos $x, y \in D$ quaisquer. Observa-se claramente que $(x, f(x)), (y, f(y)) \in E_f$. Da hipótese de ser E_f convexo, para todo $\lambda \in [0,1]$ temos

$$(\lambda x + (1-\lambda)y, \lambda f(x) + (1-\lambda) f(y)) = \lambda(x, f(x)) + (1-\lambda)(y, f(y)) \in E_f.$$

Pela definição de epígrafo, isto é equivalente a

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

logo, f é convexa. Para demonstrar a recíproca, considere f convexa. Sejam os pontos $(x, c_1), (y, c_2) \in E_f$. Como $f(x) \leq c_1$ e $f(y) \leq c_2$, pela convexidade de f, para todo $\lambda \in [0,1]$ segue que,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \leq \lambda c_1 + (1-\lambda)c_2,$$

onde obtemos,

$$\lambda(x,c_1)+(1-\lambda)(y,c_2)=(\lambda x+(1-\lambda)y,\lambda c_1+(1-\lambda)c_2)\in E_f.$$

O que mostra que E_f é convexo. □

Tanto a convexidade de conjunto como de funções são características importantes em problemas de otimização, sendo que estas hipóteses serão consideradas em capítulos posteriores. Por agora, destacamos o seguinte resultado.

Teorema 2.18 *Considere $f:D\to\mathbb{R}$ uma função convexa definida no conjunto convexo $D\subseteq\mathbb{R}^n$. Então todo minimizador local em D é minimizador global em D.*

Prova. Ver página 77 de [9]. □

Na otimização não diferenciável e convexa, frequentemente necessitamos da noção de cone normal.

Definição 2.19 *Seja D um conjunto convexo, e $\bar{x}\in D$. O cone normal ao conjunto D no ponto $\bar{x}$ é dado por*

$$N_D(\bar{x})=\{y\in\mathbb{R}^n|\ \langle y,x-\bar{x}\rangle\leq 0\quad\forall\,x\in D\}.$$

(a) (b)

Figura 2.3: *(a) e (b) são exemplos de cone mormal ao conjunto convexo D no ponto $\bar{x}$.*

Trataremos agora sobre projeções ortogonais de um ponto $x\in\mathbb{R}^n$ sobre um conjunto $D\subset\mathbb{R}^n$.

Se $D\subset\mathbb{R}^n$, a projeção (ortogonal) de um ponto $y\in\mathbb{R}^n$ sobre D é um ponto em D que está mais próximo de y utilizando a distância medida pela norma euclidiana. Assim podemos dizer que a projeção de y sobre D é uma solução do problema de otimização

$$\min\|x-y\|\quad\text{sujeito a}\quad x\in D.$$

Quando o conjunto D é não vazio e fechado, existe uma projeção de y sobre D. Se além disso, D é convexo, a projeção é única como mostra o seguinte resultado.

Teorema 2.20 (Teorema da Projeção) *Seja $D \subset \mathbb{R}^n$ um conjunto fechado, não vazio e convexo, então para todo ponto $x \in \mathbb{R}^n$ a projeção de x sobre D existe e é única. Escrevemos $P_D(x)$ significando a projeção de x sobre D. Além disso, $\bar{x} = P_D(x)$ se, e somente se,*

$$\bar{x} \in D, \qquad \langle x - \bar{x}, y - \bar{x} \rangle \leq 0 \quad \forall\, y \in D, \tag{2-5}$$

ou, equivalentemente,

$$\bar{x} \in D, \qquad x - \bar{x} \in N_D(\bar{x}).$$

Prova. Veja página 102 de [9]. □

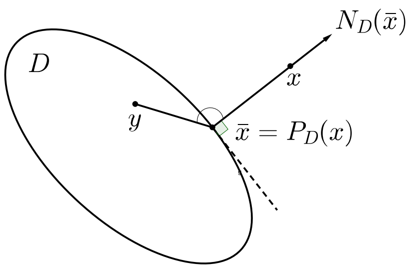

Figura 2.4: *Para todo $y \in D$, $\langle x - \bar{x}, y - \bar{x} \rangle \leq 0$, ou equivalentemente $x - \bar{x} \in N_D(\bar{x})$.*

Mostraremos a seguir que o operador projeção é não expansivo, isso significa que a distância entre dois pontos de $\mathbb{R}^n$ não pode ser menor que a distância das projeções desses pontos sobre um conjunto fechado e convexo.

Teorema 2.21 *Seja D um conjunto convexo e fechado. Então para todo $x, y \in \mathbb{R}^n$ temos*

$$\|P_D(x) - P_D(y)\| \leq \|x - y\|.$$

Prova. Se $P_D(x) = P_D(y)$, ou seja, se $\|P_D(x) - P_D(y)\| = 0$ o resultado vale trivialmente. Suponha então que $P_D(x) \neq P_D(y)$. Usando a relação (2-5) com $P_D(x) \in D$ e $P_D(y) \in D$ obtemos duas desigualdades

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

Somando membro a membro essas expressões, obtemos da linearidade do produto interno

$$0 \geq \langle y - P_D(y) - x + P_D(x), P_D(x) - P_D(y) \rangle = \|P_D(x) - P_D(y)\|^2 + \langle y - x, P_D(x) - P_D(y) \rangle,$$

usando a desigualdade de Cauchy-Schwarz, podemos concluir que

$$\|P_D(x) - P_D(y)\|^2 \leq \langle x - y, P_D(x) - P_D(y) \rangle \leq \|x - y\| \, \|P_D(x) - P_D(y)\| \, ,$$

dividindo ambos os lados dessa relação por $\|P_D(x) - P_D(y)\|$ obtem-se o resultado desejado. □

2.2.2 Funções convexas

Descreveremos mais detalhadamente assuntos sobre funções convexas (não necessariamente diferenciáveis). Iniciamos mostrando que a soma de um número finito de funções convexas é convexa.

Proposição 2.22 *Considere $D \subseteq \mathbb{R}^n$ um conjunto convexo e sejam $f_i : D \to \mathbb{R}$ funções convexas para todo $i = 1, \ldots, m$. Então, dados quaisquer números reais não negativos α_i, $i = 1, \ldots, m$, a função $f : D \to \mathbb{R}$ definida por*

$$f(x) = \sum_{i=1}^{m} \alpha_i f_i(x)$$

é convexa.

Prova. Sejam $x, y \in D$ arbitrários. Como as funções f_i, $i = 1, \ldots, m$ são convexas e $\alpha_i \geq 0$, para qualquer $\lambda \in [0, 1]$ temos

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= \sum_{i=1}^{m} \alpha_i f_i(\lambda x + (1-\lambda)y) \leq \sum_{i=1}^{m} \alpha_i \left(\lambda f_i(x) + (1-\lambda) f_i(y)\right) \\ &= \lambda \sum_{i=1}^{m} \alpha_i f_i(x) + (1-\lambda) \sum_{i=1}^{m} \alpha_i f_i(y) = \lambda f(x) + (1-\lambda) f(y). \end{aligned}$$

Isso mostra que f é convexa em D. □

Enunciaremos a seguir um importante resultado relativo a continuidade de funções convexas.

Teorema 2.23 *Sejam $U \subseteq \mathbb{R}^n$ um conjunto convexo e aberto e $f : U \to \mathbb{R}$ uma função convexa. Então f é localmente Lipschitz contínua em U. Em particular f é contínua em U.*

Prova. Veja página 144 de [9]. □

Para funções convexas não diferenciável, o conceito de subgradiente é uma generalização do vetor gradiente. Esta noção é bastante utilizada na otimização convexa.

Definição 2.24 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Dizemos que $y \in \mathbb{R}^n$ é um subgradiente de f no ponto $x \in \mathbb{R}$ se*

$$f(z) \geq f(x) + \langle y, z - x \rangle \qquad \forall\, z \in \mathbb{R}^n. \tag{2-6}$$

O conjunto de todos os subgradientes de f em x é dito subdiferencial de f em x, e é denotado por $\partial f(x)$.

Cada subgradiente $y \in \partial f(x)$ define uma aproximação linear $T : \mathbb{R}^n \to \mathbb{R}$ dada por $T(z) = f(x) + \langle y, z - x \rangle$, cujo gráfico se encontra abaixo do gráfico de f, e tal que $T(x) = f(x)$.

Quando $f : \mathbb{R}^n \to \mathbb{R}$ é convexa e diferenciável em x, a desigualdade (2-6) é satisfeita com $y = \nabla f(x)$. Com efeito, dado $z \in \mathbb{R}^n$ qualquer, pondo $v = z - x$, se $t \in (0,1]$ segue da convexidade de f

$$f(x+tv) = f(tz + (1-t)x) \leq t f(z) + (1-t) f(x) \;\;\Rightarrow\;\; t(f(z) - f(x)) \geq f(x+tv) - f(x).$$

Dividindo os dois lados desta última desigualdade por $t > 0$, e passando o limite com $t \to 0^+$, obtemos

$$f(z) - f(x) \geq \lim_{t \to 0^+} \frac{f(x+tv) - f(x)}{t} = \langle \nabla f(x), v \rangle = \langle \nabla f(x), z - x \rangle\,,$$

onde primeira igualdade acima segue de resultados sobre derivada direcional, quando a função é diferenciável, veja páginas 120 e 139 de [13].

Teorema 2.25 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Então para cada $x \in \mathbb{R}^n$, o subdiferencial de f em x é um conjunto não vazio, convexo e compacto. Além disso, f é diferenciável no ponto $x \in \mathbb{R}^n$ se, e somente se, $\partial f(x) = \{\nabla f(x)\}$.*

Prova. Veja página 173 e 176 de [9]. □

O próximo resultado generaliza o Teorema 2.8 para funções convexas não necessariamente diferenciáveis.

Teorema 2.26 *Um ponto $\bar{x} \in \mathbb{R}^n$ é solução do problema de minimização de uma função convexa $f : \mathbb{R}^n \to \mathbb{R}$ restrita a um conjunto convexo $D \subseteq \mathbb{R}^n$ se, e somente se,*

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

ou, de forma equivalente,

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

Em particular, $\bar{x}$ é solução do problema de minimização de f em $\mathbb{R}^n$ se, e somente se,

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

Prova. Veja página 177 de [9]. $\square$

Outra característica importante do subdiferencial de uma função convexa é que, sobre conjuntos limitados o subdiferencial também é limitado.

Proposição 2.27 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Se $D \subset \mathbb{R}^n$ é um conjunto limitado, então o conjunto $\cup_{x \in D} \partial f(x)$ também é limitado.*

Prova. Veja página 179 de [9]. $\square$

Proposição 2.28 *Sejam $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, funções convexas, e considere que $f : \mathbb{R}^n \to \mathbb{R}$ é definida por $f(x) = \sum_{i=1}^{m} f_i(x)$. Então para todo $x \in \mathbb{R}^n$*

$$\partial f(x) = \sum_{i=1}^{m} \partial f_i(x).$$

Prova. Veja página 180 de [9]. $\square$

Se $f_i : \mathbb{R}^n \to \mathbb{R}$ são funções convexas para $i = 1, \ldots, m,$ é possível mostrar que a função $f(x) = \max\limits_{1 \le i \le m} f_i(x)$ também é convexa. No entanto, mesmo que f_i com $i = 1, \ldots, m$, seja diferenciável, pode ser que f não seja. O próximo teorema mostra, em particular, como obter o subdiferencial de f.

Teorema 2.29 *Seja Z um conjunto compacto. Suponha que $g : \mathbb{R}^n \times Z \to \mathbb{R}$ seja uma função contínua tal que $g(\cdot, z) : \mathbb{R}^n \to \mathbb{R}$ seja convexa para todo $z \in Z$. Seja*

$$f : \mathbb{R}^n \to \mathbb{R}, \qquad f(x) = \max_{z \in Z} g(x, z).$$

Se $g(\cdot, z)$ é diferenciável para todo $z \in Z$ e $\frac{\partial g}{\partial x}(x, \cdot)$ é contínua em Z para todo $x \in \mathbb{R}^n$, então

$$\partial f(x) = \operatorname{conv}\left\{ \frac{\partial g}{\partial x}(x, \bar{z}) \,|\; \bar{z} \in Z(x) \right\}, \quad x \in \mathbb{R}^n,$$

onde

$$Z(x) = \left\{ \bar{z} \in Z \,|\; f(x) = g(x, \bar{z}) = \max_{z \in Z} g(x, z) \right\}.$$

Prova. Veja página 181 de [9]. $\square$

O Teorema 2.29 se aplica, por exemplo, para a função $f : \mathbb{R} \to \mathbb{R}$, dada por $f(x) = |x|$, pois é fácil ver que $f(x) = \max_{z \in Z} g(x,z)$, onde $Z = \{1,2\}$, $g(x,1) = x$ e $g(x,2) = -x$. Para $x > 0$, $f(x) = x = g(x,1)$, $Z(x) = \{1\}$ e temos que

$$f'(x) = \frac{\partial g}{\partial x}(x,1) = 1.$$

Para $x < 0$, $f(x) = -x = g(x,2)$, $Z(x) = \{2\}$ e temos que

$$f'(x) = \frac{\partial g}{\partial x}(x,2) = -1.$$

Para $x = 0$, $f(0) = 0 = g(0,1) = g(0,2)$, $Z(0) = \{1,2\}$ e temos que

$$\partial f(0) = \operatorname{conv}\left\{\frac{\partial g}{\partial x}(0,1), \frac{\partial g}{\partial x}(0,2)\right\} = \operatorname{conv}\{1,-1\} = [-1,1].$$

Algumas vezes, em métodos computacionais de otimização, faz-se necessário o cálculo de um valor aproximado do subgradiente. Surge então o conceito de ε-subgradiente.

Definição 2.30 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa, e ε um número não negativo. Dizemos que $y \in \mathbb{R}^n$ é um ε-subgradiente de f no ponto $x \in \mathbb{R}$ se*

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

O conjunto de todos os ε-subgradientes de f em x chama-se ε-subdiferencial de f em x, e é denotado por $\partial_\varepsilon f(x)$.

CAPÍTULO 3

Método do Subgradiente Incremental

Neste capítulo descreveremos o foco central desse trabalho. Apresentaremos o algoritmo do subgradiente incremental e estudaremos alguns resultados de convergência para ele. Para estes resultados utilizamos [14] como referência principal. As técnicas apresentadas neste capítulo, embora sejam para otimização convexa não diferenciável, podem ser de grande interesse na resolução de problemas em que a função objetivo seja diferenciável, ou até mesmo não convexa (veja, por exemplo, seção 5.2 de [9]).

Considere novamente um problema de otimização descrito da forma

$$\min f(x) := \sum_{i=1}^{m} f_i(x), \quad \text{sujeito a} \quad x \in X, \tag{3-1}$$

onde $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$ são funções convexas e $X \subseteq \mathbb{R}^n$ é um conjunto não vazio, fechado e convexo. Em seguida apresentaremos o método subgradiente incremental o qual será usado para resolver o problema (3-1).

Algorithm 3.1: Algoritmo do subgradiente incremental

Escolha $x^0 \in \mathbb{R}^n$ e defina $k = 0$.

Passo 1. Seja

$$\psi^{0,k} = x^k. \tag{3-2}$$

Passo 2. Para cada $i = 1, \ldots, m$ obtenha $g^{i,k} \in \partial f_i(\psi^{i-1,k})$ e calcule

$$\psi^{i,k} = P_X\left(\psi^{i-1,k} - \alpha_k g^{i,k}\right). \tag{3-3}$$

Passo 3. Atualize

$$x^{k+1} = \psi^{m,k}. \tag{3-4}$$

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

No Algoritmo 3.1 a atualização descrita em (3-3) denota as m subiterações do k-ésimo ciclo, e $\partial f_i(\psi^{i-1,k})$ é o subdiferencial (conjunto de todos os subgradientes) de f_i

no ponto $\psi^{i-1,k}$.

É importante destacar que na otimização não diferenciável, mais especialmente em métodos subgradientes, em geral, critérios de parada não são simples. Esta deficiência pode ser vista quando tentamos adaptar critério de parada utilizados em outros métodos. Por exemplo, no caso diferenciável a condição $\|\nabla f(x^k)\| \leq \varepsilon$ onde $\varepsilon > 0$ é suficientemente pequeno, é bastante utilizada. No entanto, afirmamos que esse tipo de regra não pode ser estendida ao caso não diferenciável, como mostra o seguinte exemplo. Seja $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = |x|$. Temos que, para $x^k \neq 0$ suficientemente próximo da solução $x^* = 0$, tem-se $|g| = 1$ para todo $g \in \partial f(x^k)$. Até mesmo se $x^k = 0$, pode acontecer de $|g| > \varepsilon$, pois $\partial f(0) = [-1,1]$.

Todos os resultados desse capítulo objetivam solucionar o problema (3-1). Assim, sempre que nos referirmos a função f ficará subentendido se tratar da função, escrita como soma das f_i's, do problema (3-1). Também, por simplicidade, escreveremos muitas vezes método incremental ao invéz de método do subgradiente incremental.

Em princípio, analisando as iterações (3-2)-(3-4) no Algoritmo 3.1, obtemos uma estimativa que é frequentemente exigida em métodos subgradiente, pois como afirma o próximo resultado, sob certas hipóteses, a soma dos subgradientes $g^{i,k}$ é um ε_k-subgradiente de f em x^k, onde $\varepsilon_k > 0$ é um valor que depende de α_k.

Proposição 3.1 *Considere o problema* (3-1)*, e seja* $\{x^k\}$ *uma sequência gerada pelo Algoritmo incremental 3.1. Suponha que, os valores*

$$C_i := \sup_{k \geq 0} \left\{ \|g\| \mid g \in \partial f_i(x^k) \cup \partial f_i(\psi^{i-1,k}) \right\}, \quad i = 1, \ldots, m \tag{3-5}$$

sejam finitos. Então $\sum_{i=1}^{m} g^{i,k} \in \partial_{\varepsilon_k} f(x^k)$*, onde* $g^{i,k}$ *são os subgradientes no* ***Passo 2*** *do Algoritmo 3.1, e*

$$\varepsilon_k := 2\alpha_k \sum_{i=2}^{m} C_i \left(\sum_{j=1}^{i-1} C_j \right) \tag{3-6}$$

Prova. Observe inicialmente, que para todo $z \in \mathbb{R}^n$, pela linearidade do produto interno, podemos escrever

$$\langle \sum_{i=1}^{m} g^{i,k}, z - x^k \rangle = \sum_{i=1}^{m} \left(\langle g^{i,k}, z - \psi^{i-1,k} \rangle + \langle g^{i,k}, \psi^{i-1,k} - x^k \rangle \right).$$

Usando a desigualdade de Cauchy-Shwarz no segundo membro do lado direito dessa igualdade e sendo, $g^{i,k} \in \partial f_i(\psi^{i-1,k}) \;\Rightarrow\; f_i(z) \geq f_i(\psi^{i-1,k}) + \left\langle g^{i,k}, z - \psi^{i-1,k} \right\rangle$, para cada

$i = 1, \ldots, m$ segue então, com algumas manipulações algébricas, que

$$\begin{aligned}\langle \sum_{i=1}^{m} g^{i,k}, z - x^k \rangle &\leq \sum_{i=1}^{m} \left(f_i(z) - f_i(\psi^{i-1,k}) + \|g^{i,k}\| \|\psi^{i-1,k} - x^k\| \right) \\ &= f(z) - f(x^k) + \sum_{i=1}^{m} \left(f_i(x^k) - f_i(\psi^{i-1,k}) + \|g^{i,k}\| \|\psi^{i-1,k} - x^k\| \right),\end{aligned}$$

como para $i = 1$, $\psi^{i-1,k} = \psi^{0,k} = x^k$, na soma do lado direito da relação acima podemos omitir o primeiro índice, ou seja, temos

$$\langle \sum_{i=1}^{m} g^{i,k}, z - x^k \rangle \leq f(z) - f(x^k) + \sum_{i=2}^{m} \left(f_i(x^k) - f_i(\psi^{i-1,k}) + \|g^{i,k}\| \|\psi^{i-1,k} - x^k\| \right). \quad (3\text{-}7)$$

Como f_i é convexa, para cada $i = 1, \ldots, m$, existe $\tilde{g}^{i,k} \in \partial f_i(x^k)$ tal que

$$f_i(\psi^{i-1,k}) \geq f_i(x^k) + \langle \tilde{g}^{i,k}, \psi^{i-1,k} - x^k \rangle,$$

o qual combinado com a desigualdade de Cauchy-Schwarz, implica que para cada $i = 2, \ldots, m$

$$f_i(x^k) - f_i(\psi^{i-1,k}) \leq \langle \tilde{g}^{i,k}, x^k - \psi^{i-1,k} \rangle \leq \|\tilde{g}^{i,k}\| \|\psi^{i-1,k} - x^k\|. \quad (3\text{-}8)$$

Assim, podemos combinar a relação (3-7) em (3-8) para obter a seguinte desigualdade

$$\langle \sum_{i=1}^{m} g^{i,k}, z - x^k \rangle \leq f(z) - f(x^k) + \sum_{i=2}^{m} \left(\|\tilde{g}^{i,k}\| + \|g^{i,k}\| \right) \|\psi^{i-1,k} - x^k\|. \quad (3\text{-}9)$$

Usando as definições (3-2)-(3-4), a não expansividade do operador projeção e ainda que $x^k = \psi^{0,k} \in X$ temos, para $i = 2, \ldots, m$

$$\begin{aligned}\left\| \psi^{i-1,k} - x^k \right\| &\leq \left\| \psi^{i-2,k} - \alpha_k g^{i-1,k} - \psi^{0,k} \right\| \\ &\leq \left\| P_X\left(\psi^{i-3,k} - \alpha_k g^{i-2,k} \right) - P_X\left(\psi^{0,k} \right) \right\| + \alpha_k \left\| g^{i-1,k} \right\| \\ &\leq \left\| \psi^{i-3,k} - \alpha_k g^{i-2,k} - \psi^{0,k} \right\| + \alpha_k \left\| g^{i-1,k} \right\| \\ &\vdots \\ &\leq \alpha_k \left\| g^{1,k} \right\| + \cdots + \alpha_k \left\| g^{i-1,k} \right\| = \alpha_k \sum_{j=1}^{i-1} \left\| g^{j,k} \right\|. \quad (3\text{-}10)\end{aligned}$$

Agora, combinando a hipótese (3-5) com as relações (3-9) e (3-10) podemos concluir que

$$\langle \sum_{i=1}^{m} g^{i,k}, z - x^k \rangle \leq f(z) - f(x^k) + \sum_{i=2}^{m} 2C_i \left(\alpha_k \sum_{j=1}^{i-1} C_j \right),$$

e pela definição de ε_k em (3-6) obtemos, para todo $z \in \mathbb{R}^n$

$$f(z) \geq f(x^k) + \langle \sum_{i=1}^{m} g^{i,k}, z - x^k \rangle - \varepsilon_k,$$

ou seja, $\sum_{i=1}^{m} g^{i,k} \in \partial_{\varepsilon_k} f(x^k)$. □

Assim, pela Proposição 3.1, se os subgradientes $g^{i,k}$ e $\tilde{g}^{i,k}$ são limitados para todo $k \geq 0$, e consequentemente os valores C_i serão finitos, então ε_k é limitado e tende a zero se $\alpha_k \to 0$. Segue que, se escolhermos o comprimento de passo α_k tal que $\alpha_k \to 0$, então, sob hipóteses usuais, as propriedades de convergência do método subgradiente (veja página 349 de [10]) podem ser estendidas para o método incremental.

3.1 Análise de convergência

Antes de passarmos às considerações sobre os resultados de convergência, destacamos que no decorrer desse trabalho adotaremos as seguintes notações

$$f^* = \inf_{x \in X} f(x), \quad X^* = \{x \in X; f(x) = f^*\}, \quad \operatorname{dist}(x, X^*) = \inf_{x \in X^*} \|x - x^*\|.$$

Além disso, para obter os resultados de convergência nessa seção, precisamos da seguinte hipótese sobre os subdiferenciais das funções componentes.

Hipótese 3.2 (Subgradientes limitados) *Existem números reais $C_1, \ldots, C_m$ tais que*

$$\|g\| \leq C_i \quad \forall g \in \partial f_i(x^k) \cup \partial f_i(\psi^{i-1,k}), \quad i = 1, \ldots, m \quad k = 0, 1 \ldots.$$

Sendo cada f_i uma função real convexa sobre o espaço inteiro $\mathbb{R}^n$, o subdiferencial $\partial f_i(x)$ é não vazio, convexo e compacto para todo x e todo i (veja Teorema 2.25). Logo, se o conjunto X é compacto, ou as sequências $\{\psi^{i,k}\}$ são limitadas, então a Hipótese 3.2 é satisfeita, já que o conjunto $\cup_{x \in D} \partial f_i(x)$ é limitado para todo conjunto limitado D (veja Proposição 2.27).

Introduziremos agora o Lema que fornecerá uma estimativa que será usada repetidamente nas análises de convergência.

Lema 3.3 *Assuma a Hipótese 3.2 e seja $\{x^k\}$ uma sequência gerada pelo Algoritmo incremental 3.1. Então para todo $y \in X$ e $k \geq 0$, temos*

$$\|x^{k+1} - y\|^2 \leq \|x^k - y\|^2 - 2\alpha_k \left(f(x^k) - f(y) \right) + \alpha_k^2 C^2$$

onde $C = \sum_{i=1}^{m} C_i$, *e* C_i *é a constante da Hipótese 3.2*

Prova. Usando a propriedade não expansiva do operador projeção, e a iteração do método incremental, temos para todo $y \in X$

$$\begin{aligned} \|\psi^{i,k} - y\|^2 &\leq \|\psi^{i-1,k} - \alpha_k g^{i,k} - y\|^2 \\ &= \|\psi^{i-1,k} - y\|^2 - 2\alpha_k \langle \psi^{i-1,k} - y, g^{i,k} \rangle + \alpha_k^2 \|g^{i,k}\|^2. \end{aligned} \tag{3-11}$$

Como $g^{i,k} \in \partial f_i(\psi^{i-1,k})$, a desigualdade do subgradiente (Definição 2.24) implica que

$$\langle \psi^{i-1,k} - y, g^{i,k} \rangle \geq f_i(\psi^{i-1,k}) - f_i(y)$$

multiplicando a última desigualdade por $-2\alpha_k \leq 0$ e substituindo em (3-11), obtemos

$$\|\psi^{i,k} - y\|^2 \leq \|\psi^{i-1,k} - y\|^2 - 2\alpha_k \left(f_i(\psi^{i-1,k}) - f_i(y) \right) + \alpha_k^2 C_i^2 \qquad \forall\, i,k$$

onde usamos também a limitação do subdiferencial (Hipótese 3.2). Agora, como a última relação vale para todo $i = 1, \ldots, m$, e como $x^{k+1} = \psi^{m,k}$ e $x^k = \psi^{0,k}$, segue então que para todo $y \in X$ e todo k

$$\begin{aligned} \|x^{k+1} - y\|^2 &\leq \|x^k - y\|^2 - 2\alpha_k \sum_{i=1}^{m} \left(f_i(\psi^{i-1,k}) - f_i(y) \right) + \alpha_k^2 \sum_{i=1}^{m} C_i^2 \\ &= \|x^k - y\|^2 - 2\alpha_k \left(f(x^k) - f(y) + \sum_{i=1}^{m} \left(f_i(\psi^{i-1,k}) - f_i(x^k) \right) \right) \\ &\quad + \alpha_k^2 \sum_{i=1}^{m} C_i^2. \end{aligned} \tag{3-12}$$

Como f_i é convexa, então para todo $i = 1, \ldots, m$, existe $\tilde{g}^{i,k} \in \partial f_i(x^k)$, i. e.,

$$f_i(\psi^{i-1,k}) - f_i(x^k) \geq \langle \tilde{g}^{i,k}, \psi^{i-1,k} - x^k \rangle$$

multiplicando a última desigualdade por $-2\alpha_k \leq 0$ e usando Cauchy-Schwarz e Hipótese 3.2, temos

$$-2\alpha_k \left(f_i(\psi^{i-1,k}) - f_i(x^k) \right) \leq 2\alpha_k \|\tilde{g}^{i,k}\| \|\psi^{i-1,k} - x^k\| \leq 2\alpha_k C_i \|\psi^{i-1,k} - x^k\|. \tag{3-13}$$

Agora, usando um argumento similiar ao da Proposição 3.1 (veja equação (3-10)), podemos obter, para $i = 2, \ldots, m$

$$\|\psi^{i-1,k} - x^k\| \leq \alpha_k \sum_{j=1}^{i-1} \|g^{i,k}\|$$

Daí, combinando a última desigualdade com (3-12) e (3-13), temos

$$\begin{aligned}
\|x^{k+1}-y\|^2 &\leq \|x^k-y\|^2 - 2\alpha_k\left(f(x^k)-f(y)\right) + 2\alpha_k^2\sum_{i=2}^{m} C_i\left(\sum_{j=1}^{i-1}\|g^{i,k}\|\right) + \alpha_k^2\sum_{i=1}^{m} C_i^2 \\
&\leq \|x^k-y\|^2 - 2\alpha_k\left(f(x^k)-f(y)\right) + \alpha_k^2\left[2\sum_{i=2}^{m} C_i\left(\sum_{j=1}^{i-1} C_j\right) + \sum_{i=1}^{m} C_i^2\right].
\end{aligned}$$

Usando a definição da constante C, o último termo do lado direito da relação acima pode ser escrito como

$$2\sum_{i=2}^{m} C_i\left(\sum_{j=1}^{i-1} C_j\right) + \sum_{i=1}^{m} C_i^2 = \left[\sum_{i=1}^{m} C_i\right]^2 = C^2,$$

portanto, para todo $y \in X$ e todo $k \geq 0$, concluímos que

$$\|x^{k+1}-y\|^2 \leq \|x^k-y\|^2 - 2\alpha_k\left(f(x^k)-f(y)\right) + \alpha_k^2 C^2.$$

□

Entre outras coisas, o Lema 3.3 garante que dado um iterando x^k e um ponto $y \in X$, tal que a função objetivo avaliada nesse ponto seja menor ou igual ao valor avaliado em x^k, ou seja $f(y) \leq f(x^k)$, o iterando seguinte x^{k+1} não poderá estar mais longe de y que x^k desde que o passo α_k seja um valor pequeno (menor ou igual a $2(f(x^k)-f(y))/C^2)$). A seguir discutiremos algumas escolhas de comprimento de passo α_k e as análises de convergência associadas.

3.1.1 Regra comprimento de passo constante

O primeiro caso considerado é a regra de comprimento de passo constante, onde α_k é um valor fixo positivo α para toda iteração k. Esse tipo de comprimento de passo serve como uma motivação para outros casos que serão considerados.

Teorema 3.4 *Assuma a Hipótese 3.2, e seja $\{x^k\}$ uma sequência gerada pelo Algoritmo incremental 3.1 com o comprimento de passo $\alpha_k = \alpha$ para todo $k \geq 0$. Então as seguintes afirmações são válidas:*

(a) Se $f^* = -\infty$, então

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

(b) Se $f^* > -\infty$, então

$$\liminf_{k\to\infty} f(x^k) \leq f^* + \frac{\alpha C^2}{2},$$

onde $C = \sum_{i=1}^{m} C_i$, *e* C_i *é a constante da Hipótese 3.2.*

Prova. Provaremos (a) e (b) simultaneamente. Para isso suponha que o resultado não seja verdadeiro, logo existe $\varepsilon > 0$ tal que

$$\liminf_{k\to\infty} f(x^k) > f^* + \frac{\alpha C^2}{2} + 2\varepsilon.$$

Como f é contínua existe também $\hat{y} \in X$ tal que

$$\liminf_{k\to\infty} f(x^k) \geq f(\hat{y}) + \frac{\alpha C^2}{2} + 2\varepsilon. \tag{3-14}$$

Por definição de $\liminf_{k\to\infty} f(x^k) = \sup_k \inf \left\{ f(x^j) \mid j \geq k \right\}$, podemos encontrar k_0 suficientemente grande tal que para todo $k \geq k_0$

$$f(x^k) \geq \liminf_{k\to\infty} f(x^k) - \varepsilon. \tag{3-15}$$

Somando termo a termo as desigualdades (3-14) e (3-15) obtemos

$$f(x^k) - f(\hat{y}) \geq \frac{\alpha C^2}{2} + \varepsilon.$$

Agora, combinando o Lema 3.3 com $y = \hat{y}$ e $\alpha_k = \alpha$ e a última desigualdade, obtemos para todo $k \geq k_0$

$$\begin{aligned}
\|x^{k+1} - \hat{y}\|^2 &\leq \|x^k - \hat{y}\|^2 - 2\alpha\left(f(x^k) - f(\hat{y})\right) + \alpha^2 C^2 \\
&\leq \|x^k - \hat{y}\|^2 - 2\alpha\left(\frac{\alpha C^2}{2} + \varepsilon\right) + \alpha^2 C^2 \\
&= \|x^k - \hat{y}\|^2 - 2\alpha\varepsilon.
\end{aligned}$$

Sendo essa relação válida para todo $k \geq k_0$, podemos concluir que

$$\|x^{k+1} - \hat{y}\|^2 \leq \|x^k - \hat{y}\|^2 - 2\alpha\varepsilon \leq \|x^{k-1} - \hat{y}\|^2 - 4\alpha\varepsilon \leq \cdots \leq \|x^{k_0} - \hat{y}\|^2 - 2\,(k + 1 - k_0)\,\alpha\varepsilon,$$

o qual não é verdadeiro para k suficientemente grande. Dessa contradição, segue que $\liminf_{k\to\infty} f(x^k) \leq f^* + \alpha C^2/2$. Em particular se $f^* = -\infty$ temos $\liminf_{k\to\infty} f(x^k) = f^*$ concluindo a demonstração. □

3.1.2 Regra comprimento de passo pré-determinado

Trabalharemos agora com o caso onde o comprimento de passo é fixado a priori e satisfaz $\alpha_k > 0$ e $\alpha_k \to 0$ ($k \to \infty$). Ressaltamos que o primeiro resultado a ser introduzido é semelhante ao clássico resultado de convergência do método do subgradiente (veja, por exemplo, o Teorema 5.1.1 página 349 de [10]).

Teorema 3.5 *Assuma a Hipótese 3.2, e seja $\{x^k\}$ uma sequência gerada pelo Algoritmo incremental 3.1, tal que o comprimento de passo α_k satisfaz*

$$\alpha_k > 0, \qquad \lim_{k\to\infty} \alpha_k = 0, \qquad \sum_{k=0}^{\infty} \alpha_k = \infty.$$

Então,

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

Prova. Suponha que a afirmação não seja verdadeira, ou seja, que $\liminf_{k\to\infty} f(x^k) > f^*$. Então existe $\varepsilon > 0$ tal que

$$\liminf_{k\to\infty} f(x^k) > f^* + 2\varepsilon.$$

Seja $\hat{y} \in X$ tal que

$$\liminf_{k\to\infty} f(x^k) \geq f(\hat{y}) + 2\varepsilon. \tag{3-16}$$

Como estamos supondo que $\liminf_{k\to\infty} f(x^k) > f^*$ podemos afirmar, em qualquer caso ($f^* = -\infty$ ou $f^* > -\infty$), que $\liminf_{k\to\infty} f(x^k) > -\infty$. Logo, pela definição de $\liminf_{k\to\infty} f(x^k) = \sup_k \inf\left\{f(x^j) \mid j \geq k\right\}$, existe um índice k_0 tal que $k \geq k_0$ implica em

$$\inf\left\{f(x^j) \mid j \geq k\right\} \geq \liminf_{k\to\infty} f(x^k) - \varepsilon.$$

Mas, para todo k, $f(x^k) \geq \inf\left\{f(x^j) \mid j \geq k\right\}$. Assim, para todo $k \geq k_0$

$$f(x^k) \geq \liminf_{k\to\infty} f(x^k) - \varepsilon. \tag{3-17}$$

Somando termo a termo (3-16) e (3-17) temos

$$f(x^k) - f(\hat{y}) \geq \varepsilon, \qquad \forall\, k \geq k_0.$$

Agora, combinando o Lema 3.3 com $y = \hat{y}$ e a última desigualdade, obtemos para todo $k \geq k_0$

$$\|x^{k+1} - \hat{y}\|^2 \leq \|x^k - \hat{y}\|^2 - 2\alpha_k \left(f(x^k) - f(\hat{y})\right) + \alpha_k^2 C^2 \leq \|x^k - \hat{y}\|^2 - 2\alpha_k \varepsilon + \alpha_k^2 C^2.$$

Como $\alpha_k \to 0$ existe $k_1 \geq k_0$ tal que para todo $k \geq k_1$, $\alpha_k C^2 \leq \varepsilon \Rightarrow \alpha_k^2 C^2 \leq \alpha_k \varepsilon$. Assim, para todo $k \geq k_1$ segue que

$$\|x^{k+1} - \hat{y}\|^2 \leq \|x^k - \hat{y}\|^2 - \alpha_k \varepsilon \;\Leftrightarrow\; \varepsilon \alpha_k \leq \|x^k - \hat{y}\|^2 - \|x^{k+1} - \hat{y}\|^2.$$

Podemos concluir então, que para todo $k \geq k_1$ vale

$$\varepsilon \sum_{j=k_1}^{k} \alpha_j \leq \sum_{j=k_1}^{k} \left(\|x^k - \hat{y}\|^2 - \|x^{k+1} - \hat{y}\|^2 \right) = \|x^{k_1} - \hat{y}\|^2 - \|x^{k+1} - \hat{y}\|^2 \leq \|x^{k_1} - \hat{y}\|^2,$$

mas isso é uma contradição quando $k \to \infty$, pois por hipótese $\sum_{k=0}^{\infty} \alpha_k = \infty$. □

Quando assumimos ainda que o conjunto solução X^* é não vazio e limitado, a conclusão do Teorema 3.5 pode ser fortalecida como mostraremos a seguir.

Teorema 3.6 *Assuma a Hipótese 3.2 e considere que X^* é não vazio e limitado. Suponha ainda que $\{x^k\}$ é uma sequência gerada pelo Algoritmo incremental 3.1 tal que o comprimento de passo α_k satisfaz*

$$\alpha_k > 0, \qquad \lim_{k \to \infty} \alpha_k = 0, \qquad \sum_{k=0}^{\infty} \alpha_k = \infty.$$

Então,

$$\lim_{k \to \infty} \operatorname{dist}\left(x^k, X^*\right) = 0 \qquad \lim_{k \to \infty} f(x^k) = f^*.$$

Prova. A idéia consiste em mostrar que uma vez que a iteração x^k entra em um conjunto de nível, ele não pode ir para longe desse conjunto. Para isso, fixamos $\gamma > 0$. Como $\alpha_k \to 0$, existe k_0 tal que $\alpha_k^2 C^2 \leq \alpha_k \gamma \quad \forall\, k \geq k_0$. Distinguimos dois casos.
Caso 1: $f(x^k) > f^* + \gamma$. Pelo Lema 3.3 obtemos para todo $x^* \in X^*$

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - 2\alpha_k \left(f(x^k) - f^* \right) + \alpha_k^2 C^2$$

então, como $f(x^k) - f^* > \gamma$ e $\alpha_k^2 C^2 \leq \alpha_k \gamma$, temos

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \alpha_k \gamma \leq \|x^k - x^*\|^2. \tag{3-18}$$

Extraindo a raiz quadrada em ambos os lados de (3-18) e tomando o ínfimo sobre os $x^* \in X^*$, conclui-se que

$$\operatorname{dist}\left(x^{k+1}, X^*\right) \leq \operatorname{dist}\left(x^k, X^*\right). \tag{3-19}$$

Caso 2: $f(x^k) \leq f^* + \gamma$. Nesse caso, vemos que x^k pertence ao conjunto de nível

$$L_\gamma := L_{f,X}(f^* + \gamma) = \{y \in X \mid f(y) \leq f^* + \gamma\}$$

que é limitado (em vista da limitação de X^*), assim podemos afirmar que

$$\operatorname{dist}\left(x^k, X^*\right) \leq \max_{y \in L_\gamma} \operatorname{dist}(y, X^*) := d(\gamma) < \infty. \tag{3-20}$$

Usando Lema 3.3 com $y = x^k$ temos que $\left\|x^{k+1} - x^k\right\|^2 \leq \alpha_k^2 C^2$, o qual implica que $\left\|x^{k+1} - x^k\right\| \leq \alpha_k C$. Logo, para todo $x^* \in X^*$

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

ao tomarmos o ínfimo sobre $x^* \in X^*$, obtemos

$$\operatorname{dist}\left(x^{k+1}, X^*\right) \leq \operatorname{dist}\left(x^k, X^*\right) + \alpha_k C,$$

combinando (3-20) com a última expressão, segue que

$$\operatorname{dist}\left(x^{k+1}, X^*\right) \leq \operatorname{dist}\left(x^k, X^*\right) + \alpha_k C \leq d(\gamma) + \alpha_k C. \tag{3-21}$$

Agora, vemos que o **Caso 1** não pode ocorrer para todo $k \geq k_0$, pois por (3-18) teríamos

$$0 \leq \|x^{k+1} - x^*\|^2 \leq \|x^{k_0} - x^*\|^2 - \gamma \sum_{j=k_0}^{k} \alpha_j \;\Rightarrow\; \gamma \sum_{j=k_0}^{k} \alpha_j \leq \|x^{k_0} - x^*\|^2,$$

o que seria uma contradição quando $k \to \infty$, pois por hipótese $\sum_{k=0}^{\infty} \alpha_k = \infty$. Logo, o **Caso 2** necessariamente ocorre para infinito k. Seja então $k_1 < k_2 < \cdots < k_j < \cdots$ os indices consecutivos a partir de k_0 em que ocorre o **Caso 2**. Por (3-21), temos

$$\operatorname{dist}\left(x^{k_j+1}, X^*\right) \leq \operatorname{dist}\left(x^{k_j}, X^*\right) + \alpha_{k_j} C \leq d(\gamma) + \alpha_{k_j} C \qquad \forall\, j = 1, 2, \ldots. \tag{3-22}$$

Desta forma, fixado um j, vemos que para os índices $k = k_j + 1, \ldots, k_{j+1} - 1$ ocorreu o **Caso 1**, então por (3-19) temos

$$\operatorname{dist}\left(x^{k_{j+1}}, X^*\right) \leq \operatorname{dist}\left(x^{k_{j+1}-1}, X^*\right) \leq \cdots \leq \operatorname{dist}\left(x^{k_j+1}, X^*\right). \tag{3-23}$$

A partir das desigualdades (3-22) e (3-23) podemos concluir que

$$\operatorname{dist}\left(x^k, X^*\right) \leq d(\gamma) + \alpha_{k_j} C \qquad k_j + 1 \leq k \leq k_{j+1}. \tag{3-24}$$

Portanto, para cada $\gamma > 0$ fixado inicialmente, podemos obter índices da sequência $\{x^k\}$ satisfazendo (3-24). Assim, dado $\varepsilon > 0$ arbitrário, em vista da continuidade da função f, e sendo X^* não vazio e limitado, segue que o conjunto de nível de f é compacto. Então, $\lim\limits_{\gamma \to 0} d(\gamma) = 0$. Logo existe $\gamma_0 > 0$ tal que $d(\gamma_0) < \varepsilon/2$. Para esse γ_0 existem índices $k_1 < k_2 < \cdots < k_j < \cdots$ da sequência $\{x^k\}$ tal que ocorre a desigualdade (3-24) (com γ_0 no lugar de γ). Como $\alpha_k \to 0$ $(k \to \infty)$ qualquer subsequência de α_k também tende a zero, em particular $\alpha_{k_j} \to 0$ $(j \to \infty)$, então existe j_0 tal que $\alpha_{k_{j_0}} C < \varepsilon/2, \quad \forall\, j \geq j_0$, por (3-24) podemos concluir que

$$\operatorname{dist}\left(x^k, X^*\right) \leq d(\gamma_0) + \alpha_{k_{j_0}} C < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \qquad \forall\, k \geq k_{j_0} + 1,$$

isto prova que $\lim\limits_{k \to \infty} \operatorname{dist}\left(x^k, X^*\right) = 0$. Novamente, usando a continuidade da função f esta relação implica também que $\lim\limits_{k \to \infty} f(x^k) = f^*$. □

O Teorema 3.6 não garante convergência de toda a sequência $\{x^k\}$. Em seguida mostraremos que, com hipótese adicional no comprimento de passo, esta convergência é garantida.

Teorema 3.7 *Assuma a Hipótese 3.2 e considere que X^* é não vazio. Suponha ainda que $\{x^k\}$ é uma sequência gerada pelo Algoritmo incremental 3.1 tal que o comprimento de passo α_k satisfaz*

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

Então a sequência $\{x^k\}$ converge para alguma solução ótima.

Prova. Seja $x^* \in X^*$, claro que $f(x^k) \geq f^* = f(x^*) \Rightarrow f(x^k) - f(x^*) \geq 0 \;\; \forall\, k \geq 0$. Então, usando o Lema 3.3 com $y = x^*$ temos que

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 - 2\alpha_k \left(f(x^k) - f(x^*)\right) + \alpha_k^2 C^2 \\ &\leq \|x^k - x^*\|^2 + \alpha_k^2 C^2. \end{aligned}$$

Como essa desigualdade vale para todo $x^* \in X^*$ e todo $k \geq 0$, segue da Definição 2.5 que a sequência $\{x^k\}$ é quasi-Féjer convergente ao conjunto não vazio X^* (pois por hipótese $C^2 \sum_{k=0}^{\infty} \alpha_k^2 < \infty$). Pela Proposição 2.6 $\{x^k\}$ é limitada, e portanto possui um ponto de acumulação $\bar{x}$. Agora, pelo Teorema 3.6 temos que $\bar{x} \in X^*$. Então, segue novamente da Proposição 2.6 que toda a sequência $\{x^k\}$ converge para $\bar{x}$. □

3.1.3 Regra comprimento de passo dinâmico para f^* conhecido

No caso particular quando f^* é conhecido, uma alternativa interessante para o comprimento de passo α_k, que foi proposto por Polyak [15] no método do subgradiente, é dado por

$$\alpha_k = \gamma_k \frac{f(x^k) - f^*}{\|g^k\|^2}, \qquad 0 < \underline{\gamma} \leq \gamma_k \leq \bar{\gamma} < 2,$$

onde $g^k \in \partial f(x^k)$. No problema (3-1) onde f é a soma das f_i's, suponha que conhecemos f^* previamente. Para o Algoritmo do subgradiente incremental 3.1, visando evitar o cálculo de $\|g^k\|$ a cada iteração, consideramos uma variação para este comprimento de passo, especificamente

$$\alpha_k = \gamma_k \frac{f(x^k) - f^*}{C^2}, \qquad 0 < \underline{\gamma} \leq \gamma_k \leq \bar{\gamma} < 2, \tag{3-25}$$

onde

$$C = \sum_{i=1}^{m} C_i, \qquad C_i \geq \sup_{k \geq 0} \left\{ \|g\| \,; g \in \partial f_i(x^k) \cup \partial f_i(\psi^{i-1,k}) \right\}, \quad i = 1, \ldots, m. \tag{3-26}$$

Para o comprimento de passo (3-25)-(3-26), mostraremos a seguir, um resultado de convergência que também foi obtido no Teorema 3.7.

Teorema 3.8 *Assuma a Hipótese 3.2 e considere que X^* é não vazio. Então a sequência $\{x^k\}$ gerada pelo Algoritmo incremental 3.1 com a regra de comprimento de passo (3-25)-(3-26) converge para alguma solução ótima.*

Prova. Usando o Lema 3.3 com $y = x^* \in X^*$, segue que

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - 2\alpha_k \left(f(x^k) - f^* \right) + \alpha_k^2 C^2, \quad \forall\, k \geq 0. \tag{3-27}$$

Da limitação de γ_k em (3-25) temos $0 < \underline{\gamma} \leq \gamma_k \leq \bar{\gamma} < 2 \;\Leftrightarrow\; 0 < 2 - \bar{\gamma} \leq 2 - \gamma_k \leq 2 - \underline{\gamma} < 2$. Destas duas equivalências obtemos $0 < \underline{\gamma}(2 - \bar{\gamma}) \leq \gamma_k (2 - \gamma_k)$. Daí, usando a definição de α_k (dada em (3-25)) em (3-27), resulta que

$$\begin{aligned}
\|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 - 2\gamma_k \left(\frac{f(x^k) - f^*}{C} \right)^2 + \gamma_k^2 \left(\frac{f(x^k) - f^*}{C} \right)^2 \\
&= \|x^k - x^*\|^2 - \gamma_k (2 - \gamma_k) \left(\frac{f(x^k) - f^*}{C} \right)^2 \\
&\leq \|x^k - x^*\|^2 - \underline{\gamma}(2 - \bar{\gamma}) \left(\frac{f(x^k) - f^*}{C} \right)^2 \leq \|x^k - x^*\|^2.
\end{aligned} \tag{3-28}$$

Sendo a relação (3-28) válida para todo $k \geq 0$ e todo $x^* \in X^*$, segue que a sequência $\{x^k\}$ é Féjer convergente ao conjunto não vazio X^*. Pela Proposição 2.6, a sequência $\{x^k\}$ é limitada e portanto possui um ponto de acumulação $\bar{x}$. Seja $\{x^{k_j}\}$ uma subsequência de $\{x^k\}$ tal que $\lim\limits_{j\to\infty} x^{k_j} = \bar{x}$. Da relação (3-28), temos

$$\|x^{k_j+1} - x^*\|^2 \leq \|x^{k_j} - x^*\|^2 - \underline{\gamma}(2-\overline{\gamma})\left(\frac{f(x^{k_j}) - f^*}{C}\right)^2,$$

aplicando o limite com $j \to \infty$, segue da continuidade de f que

$$\|\bar{x} - x^*\|^2 \leq \|\bar{x} - x^*\|^2 - \underline{\gamma}(2-\overline{\gamma})\left(\frac{f(\bar{x}) - f^*}{C}\right)^2,$$

o qual implica que $\bar{x} \in X^*$. Portanto, a Proposição 2.6 também garante que $\bar{x} = \lim\limits_{k\to\infty} x^k$. $\square$

3.1.4 Regra comprimento de passo dinâmico para f^* desconhecido

Na maioria dos problemas práticos o valor f^* não é conhecido. Neste caso podemos modificar o comprimento de passo (3-25) substituindo f^* por uma estimativa. Isto induz a regra comprimento de passo dada por

$$\alpha_k = \gamma_k \frac{f(x^k) - f_k^{lev}}{C^2}, \qquad 0 < \underline{\gamma} \leq \gamma_k \leq \overline{\gamma} < 2, \qquad \forall\, k \geq 0, \tag{3-29}$$

onde C é definido em (3-26) e f_k^{lev} é uma estimativa de f^*.

Discutiremos dois processos para atualização de f_k^{lev}, em ambos ele será a diferença entre o melhor valor da função $\min\limits_{0\leq j\leq k} f(x^j)$ e uma constante positiva δ_k que será ajustada durante a execução do algoritmo.

No primeiro processo de ajustamento, f_k^{lev} é dado por

$$f_k^{lev} = \min_{0\leq j\leq k} f(x^j) - \delta_k, \tag{3-30}$$

e δ_k é atualizado da seguinte forma

$$\delta_{k+1} = \begin{cases} \rho\delta_k, & \text{se} \quad f(x^{k+1}) \leq f_k^{lev}, \\ \max\{\beta\delta_k, \delta\}, & \text{se} \quad f(x^{k+1}) > f_k^{lev}, \end{cases} \tag{3-31}$$

onde δ_0, δ, β e ρ são constantes positivas fixadas com $0 < \delta \leq \delta_0$, $\beta < 1$ e $\rho \geq 1$. Se, em uma dada iteração ocorrer, $f(x^{k+1}) \leq f_k^{lev}$, aumentamos δ_{k+1} ou mantemos ele no mesmo valor, dependendo da escolha de ρ. Por outro lado, se $f(x^{k+1}) > f_k^{lev}$, δ_{k+1} é reduzido até o possível valor δ. De qualquer forma, isso garante que o passo α_k de (3-29) é positivo, pois para todo $k \geq 0$

$$f(x^k) \geq \min_{0\leq j\leq k} f(x^j) - \delta_k + \delta_k = f_k^{lev} + \delta_k \;\Rightarrow\; f(x^k) - f_k^{lev} \geq \delta_k \geq \delta, \tag{3-32}$$

então

$$\alpha_k = \gamma_k \frac{f(x^k) - f_k^{lev}}{C^2} \geq \underline{\gamma} \frac{\delta}{C^2}.$$

Para esse método, apresentamos um resultado semelhante ao obtido com o comprimento de passo constante dado no Teorema 3.4.

Teorema 3.9 *Assuma a Hipótese 3.2, e seja $\{x^k\}$ uma sequência gerada pelo Algoritmo incremental 3.1 com a regra comprimento de passo dinâmico (3-29) e processo de ajustamento (3-30)-(3-31). Então as seguintes afirmações são válidas:*

(a) Se $f^* = -\infty$, então

$$\inf_{k\geq 0} f(x^k) = f^*.$$

(b) Se $f^* > -\infty$, então

$$\inf_{k\geq 0} f(x^k) \leq f^* + \delta.$$

Prova. De forma análoga à demonstração do Teorema 3.4, mostraremos os ítens (a) e (b) simultaneamente. Para isso, suponha por contradição que

$$\inf_{k\geq 0} f(x^k) > f^* + \delta. \tag{3-33}$$

Observe que, se ocorrer $f(x^k) \leq f_{k-1}^{lev}$ temos

$$f(x^k) \leq f_{k-1}^{lev} = \min_{0\leq j\leq k-1} f(x^j) - \delta_{k-1} \leq \min_{0\leq j\leq k-1} f(x^j) - \delta$$

pois $0 < \delta \leq \delta_k, \quad \forall\, k \geq 0$. Isso mostra que o atual valor da função $\min_{0\leq j\leq k} f(x^j)$ decresce por pelo menos δ. Logo, se a desigualdade ocorre para infintos índices, digamos $k_1 < k_2 < \cdots < k_i < \cdots$, escrevendo $f(x^{k_0}) = \min_{0\leq j\leq k_1-1} f(x^j)$ temos

$$\begin{aligned}
f(x^{k_1}) &\leq f_{k_1-1}^{lev} = \min_{0\leq j\leq k_1-1} f(x^j) - \delta_{k_1-1} \leq f(x^{k_0}) - \delta \\
f(x^{k_2}) &\leq f_{k_2-1}^{lev} = \min_{0\leq j\leq k_2-1} f(x^j) - \delta_{k_2-1} \leq f(x^{k_1}) - \delta \leq f(x^{k_0}) - 2\delta \\
&\;\;\vdots \\
f(x^{k_i}) &\leq f_{k_i-1}^{lev} = \min_{0\leq j\leq k_i-1} f(x^j) - \delta_{k_i-1} \leq f(x^{k_i}) - \delta \leq f(x^{k_0}) - i\delta,
\end{aligned}$$

que contradiz nossa hipótese (3-33) quando $i \to \infty$. Assim, $f(x^k) \leq f_{k-1}^{lev}$ não poderá ocorrer para infinitos k. Logo, existe um índice k_0 tal que $f(x^k) > f_{k-1}^{lev}$ para todo $k \geq k_0$. Por (3-31) sabemos que se $f(x^k) > f_{k-1}^{lev}$ então $\delta_k = \max\{\beta\delta_{k-1}, \delta\}$. Como $\beta < 1$, vemos que existe um índice $\bar{k} \geq k_0$ tal que

$$\delta_k = \delta \qquad \forall\, k \geq \bar{k}. \tag{3-34}$$

Em vista de nossa hipótese (3-33), e a continuidade de f, existe $\bar{y} \in X$ tal que

$$\inf_{k\geq 0} f(x^k) - \delta \geq f(\bar{y}).$$

Daí, usando a definição em (3-30) e a última desigualdade, concluimos por (3-34) que para todo $k \geq \bar{k}$

$$f_k^{lev} = \min_{0\leq j\leq k} f(x^j) - \delta \geq \inf_{k\geq 0} f(x^k) - \delta \geq f(\bar{y}),$$

a qual, combinada com a definição de α_k em (3-29) implica que

$$\alpha_k \left(f(x^k) - f(\bar{y})\right) \geq \alpha_k \left(f(x^k) - f_k^{lev}\right) = \gamma_k \left(\frac{f(x^k) - f_k^{lev}}{C}\right)^2 \qquad \forall\, k \geq \bar{k}. \tag{3-35}$$

Sabemos, pelo Lema 3.3 com $y = \bar{y} \in X$, que para todo $k \geq 0$

$$\|x^{k+1} - \bar{y}\|^2 \leq \|x^k - \bar{y}\|^2 - 2\alpha_k \left(f(x^k) - f(\bar{y})\right) + \alpha_k^2 C^2. \tag{3-36}$$

Então, das expressões (3-35), (3-36) e definição de α_k dado em (3-29), resulta que para todo $k \geq \bar{k}$

$$\begin{aligned}
\|x^{k+1} - \bar{y}\|^2 &\leq \|x^k - \bar{y}\|^2 - 2\gamma_k \left(\frac{f(x^k) - f_k^{lev}}{C}\right)^2 + \gamma_k^2 \left(\frac{f(x^k) - f_k^{lev}}{C}\right)^2 \\
&= \|x^k - \bar{y}\|^2 - \gamma_k (2 - \gamma_k) \left(\frac{f(x^k) - f_k^{lev}}{C}\right)^2 \\
&\leq \|x^k - \bar{y}\|^2 - \underline{\gamma}(2 - \bar{\gamma}) \frac{\delta^2}{C^2},
\end{aligned}$$

onde usamos também a limitação de γ_k e a desigualdade (3-32). Portanto, podemos concluir que para todo $k > \bar{k}$

$$\begin{aligned}
\|x^k - \bar{y}\|^2 &\leq \|x^{k-1} - \bar{y}\|^2 - \underline{\gamma}(2 - \bar{\gamma}) \frac{\delta^2}{C^2} \leq \|x^{k-2} - \bar{y}\|^2 - 2\underline{\gamma}(2 - \bar{\gamma}) \frac{\delta^2}{C^2} \\
&\leq \cdots \leq \|x^{\bar{k}} - \bar{y}\|^2 - \left(k - \bar{k}\right) \underline{\gamma}(2 - \bar{\gamma}) \frac{\delta^2}{C^2},
\end{aligned}$$

onde fazendo $k \to \infty$ chega-se a uma contradição. Então, concluimos que as afirmações (a) e (b) são verdadeiras. $\square$

Apresentaremos agora o segundo processo de ajustamento de f_k^{lev}. Neste procedimento reduzimos δ_k sempre que o método se afasta a uma longa distância não satisfazendo $f(x^k) \leq f_{k-1}^{lev}$.

Algorithm 3.2: Algoritmo comprimento de passo dinâmico

Passo 0 : **inicio**

escolha $x^0 \in X,\ \delta_0 > 0,\ B > 0$;
seja $\sigma_0 = 0,\ f_{-1}^{rec} = \infty$;
defina $k = 0,\ l = 0,\ k(l) = 0$ ($k(l)$ denota o número de iterações l em que ocorre as atualizações de f_k^{lev}).

Passo 1 : calcule $f(x^k)$, **se** $f(x^k) < f_{k-1}^{rec}$ **então**

$f_k^{rec} = f(x^k)$

senão

$f_k^{rec} = f_{k-1}^{rec}$ (assim f_k^{rec} se mantem no menor valor atingido pelas iterações geradas até o momento, ou seja, $f_k^{rec} = \min_{0 \leq j \leq k} f(x^j)$).

Passo 2 : **se** $f(x^k) \leq f_{k(l)}^{rec} - \delta_l/2$ **então**

defina $k(l+1) = k$, $\sigma_k = 0$, $\delta_{l+1} = \delta_l$, aumente l por 1 e siga ao **Passo 4**.

Passo 3 : **se** $\sigma_k > B$ **então**

defina $k(l+1) = k$, $\sigma_k = 0$, $\delta_{l+1} = \delta_l/2$, aumente l por 1.

Passo 4 : seja $f_k^{lev} = f_{k(l)}^{rec} - \delta_l$ escolha γ_k satisfazendo $\underline{\gamma} \leq \gamma_k \leq \bar{\gamma}$ e calcule x^{k+1} pelo Algoritmo incremental 3.1 com comprimento de passo (3-29).

Passo 5 : seja $\sigma_{k+1} = \sigma_k + \alpha_k C$ (onde C é a constante em (3-26)); aumente k por 1 e siga ao **Passo 1**.

O Algoritmo 3.2 usa o mesmo iterando $f_k^{lev} = f_{k(l)}^{rec} - \delta_l$ para os índices $k = k(l), k(l)+1, \ldots, k(l+1)-1$. Esse valor é atualizado somente quando ocorre uma descida suficiente da função, ou σ_k extrapola o valor B, **Passo 2** ou **Passo 3** respectivamente. Pode ser mostrado que o valor σ_k é a soma do comprimento dos passos α_j com $k(l) \leq j \leq k$ para $k < k(l+1)$. Sempre que σ_k excede o prescrito limite superior B, o parâmetro δ_l é decrescido, o que aumenta o valor f_k^{lev}.

Lema 3.10 *Assuma a Hipótese 3.2, e seja $\{x^k\}$ uma sequência gerada pelo Algoritmo comprimento de passo dinâmico 3.2. Então,* $l \to \infty$ *e*

$$\inf_{k \geq 0} f(x^k) = -\infty \qquad \text{ou} \qquad \lim_{l \to \infty} \delta_l = 0.$$

Prova. Suponha que l tenha somente um número finito de valores, digamos $l = 0, 1, \dots, \bar{l}$. Nesse caso vemos pelo **Passo 3** e **Passo 5** que para todo $k \geq k(\bar{l})$

$$\sigma_{k+1} = \sigma_k + \alpha_k C = \sigma_{k-1} + \alpha_{k-1} C + \alpha_k C = \sigma_{k(\bar{l})} + \alpha_{k(\bar{l})} C + \cdots + \alpha_k C \leq B,$$

como isso vale para todo $k \geq k(\bar{l})$ e $\sigma_{k(\bar{l})} = 0$, pois foi a última atualização de l, podemos concluir que

$$C \sum_{j=k(\bar{l})}^{\infty} \alpha_j \leq B \quad \Rightarrow \quad \lim_{k\to\infty} \alpha_k = 0.$$

Mas isso é impossível, pois pelo **Passo 2** vemos que para todo $k \geq k(\bar{l})$

$$f(x^k) > f^{rec}_{k(\bar{l})} - \frac{\delta_{\bar{l}}}{2} = f^{rec}_{k(\bar{l})} - \delta_{\bar{l}} + \frac{\delta_{\bar{l}}}{2} = f^{lev}_k + \frac{\delta_{\bar{l}}}{2},$$

logo

$$\alpha_k = \gamma_k \frac{f(x^k) - f^{lev}_k}{C^2} > \underline{\gamma}\, \frac{\delta_{\bar{l}}}{2C^2} > 0.$$

Portanto, $l \to \infty$. Para mostrar a segunda afirmação suponha que $\lim_{l\to\infty} \delta_l = \delta$. Se $\delta > 0$, então pelo **Passo 2** e **Passo 3** segue que existe $\hat{l} \in \mathbb{N}$, tal que $\delta_l = \delta \quad \forall\, l \geq \hat{l}$, ou seja, as atualizações l, a partir de $\hat{l}$, serão todas feitas pelo **Passo 2**. Mas, se ocorre o **Passo 2** obrigatoriamente o **Passo 1** também deverá ocorrer. Assim, $\forall\, l \geq \hat{l}$ temos

$$f(x^{k(l+1)}) = f^{rec}_{k(l+1)} \leq f^{rec}_{k(l)} - \frac{\delta}{2} \leq f^{rec}_{k(l-1)} - 2\frac{\delta}{2} \leq \cdots \leq f^{rec}_{k(\hat{l})} - \left(l + 1 - \hat{l}\right)\frac{\delta}{2},$$

e quando $l \to \infty$ isso implica que $\inf_{k\geq 0} f(x^k) = -\infty$. □

Para o segundo processo de ajustamento de f^{lev}_k dado no Algoritmo 3.2 mostraremos um resultado de convergência mais satisfatório que no primeiro processo, apresentado no Teorema 3.9.

Teorema 3.11 *Assuma a Hipótese 3.2. Então para uma sequência $\{x^k\}$ gerada pelo Algoritmo comprimento de passo dinâmico 3.2, temos que*

$$\inf_{k\geq 0} f(x^k) = f^*.$$

Prova. Se $\lim_{l\to\infty} \delta_l > 0$, então de acordo com o Lema 3.10, $\inf_{k\geq 0} f(x^k) = -\infty$, mas como $f^* \leq \inf_{k\geq 0} f(x^k)$, neste caso, podemos conclir que $\inf_{k\geq 0} f(x^k) = f^*$. Suponha agora que

$\lim_{l\to\infty} \delta_l = 0$, e considere o conjunto

$$L = \left\{ l \in \mathbb{N} \mid \delta_l = \frac{\delta_{l-1}}{2} \right\}.$$

Vemos que L é o conjunto dos índices $l \geq 1$ em que ocorre o **Passo 3**. Então, para $k > k(l)$, usando o **Passo 3** juntamente com o **Passo 5**, obtemos

$$\sigma_k = \sigma_{k-1} + \alpha_{k-1}C = \sigma_{k-2} + \alpha_{k-2}C + \alpha_{k-1}C = \cdots = \sum_{j=k(l)}^{k-1} C\alpha_j.$$

Assim, novamente pelo **Passo 3**, sempre que $\sum_{j=k(l)}^{k-1} C\alpha_j > B$ temos $k(l+1) = k$ e $l+1 \in L$. Então, podemos concluir que

$$\sum_{j=k(l-1)}^{k(l)-1} \alpha_j > \frac{B}{C} \qquad \forall\, l \in L.$$

Sendo $\lim_{l\to\infty} \delta_l = 0$ a cardinalidade de L é infinita, portanto

$$\sum_{j=0}^{\infty} \alpha_j \geq \sum_{l\in L} \sum_{j=k(l-1)}^{k(l)-1} \alpha_j > \sum_{l\in L} \frac{B}{C} = \infty. \tag{3-37}$$

Agora, suponha por contradição que $\inf_{k\geq 0} f(x^k) > f^*$, assim existe $\hat{y} \in X$ e algum $\varepsilon > 0$ tais que

$$\inf_{k\geq 0} f(x^k) - \varepsilon \geq f(\hat{y}). \tag{3-38}$$

Como $\delta_l \to 0$, existe um índice $\hat{l}$ suficientemente grande tal que $\delta_l \leq \varepsilon$ para todo $l \geq \hat{l}$, assim para todo $k \geq k(\hat{l})$

$$f_k^{lev} = f_{k(l)}^{rec} - \delta_l \geq \inf_{k\geq 0} f(x^k) - \varepsilon \geq f(\hat{y}) \;\Rightarrow\; f(x^k) - f(\hat{y}) \geq f(x^k) - f_k^{lev}.$$

Usando essa relação no Lema 3.3 para o caso $y = \hat{y}$ e a definição de α_k, temos

$$\begin{aligned}
\|x^{k+1} - \hat{y}\|^2 &\leq \|x^k - \hat{y}\|^2 - 2\alpha_k \left(f(x^k) - f_k^{lev} \right) + \alpha_k^2 C^2 \\
&= \|x^k - \hat{y}\|^2 - \gamma_k (2 - \gamma_k) \left(\frac{f(x^k) - f_k^{lev}}{C} \right)^2 \\
&\leq \|x^k - \hat{y}\|^2 - \underline{\gamma}(2 - \overline{\gamma}) \left(\frac{f(x^k) - f_k^{lev}}{C} \right)^2 \qquad \forall\, k \geq k(\hat{l}).
\end{aligned}$$

Somando estas desigualdades sobre $k \geq k(\hat{l})$, obtemos

$$\frac{\gamma(2-\bar{\gamma})}{C^2} \sum_{k=k(\hat{l})}^{\infty} \left(f(x^k) - f_k^{lev} \right)^2 \leq \sum_{k=k(\hat{l})}^{\infty} \left(\|x^k - \hat{y}\|^2 - \|x^{k+1} - \hat{y}\|^2 \right) \leq \|x^{k(\hat{l})} - \hat{y}\|^2,$$

e consequentemente $\sum_{k=k(\hat{l})}^{\infty} \alpha_k^2 < \infty$ (ver a definição de α_k em (3-29)). Assim, o termo geral desta série tende a zero, isto é, $\lim_{k\to\infty} \alpha_k^2 = 0 \;\Rightarrow\; \lim_{k\to\infty} \alpha_k = 0$. Além disso, por (3-37) a série dos termos α_k é infinita, logo as hipóteses do Teorema 3.5 estão satisfeitas, então

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

Portanto $\inf_{k\geq 0} f(x^k) = f^*$, o que contradiz (3-38). □

3.2 Experimentos numéricos

Nesta seção relataremos alguns resultados numéricos com um certo tipo de problema dado por

$$\min f(x) = \sum_{i=1}^{m} \left(\frac{1}{2} \|A_i x - b_i\|^2 + \lambda \|x - \bar{x}\|_1 \right), \quad \text{sujeito a} \quad x \geq 0, \tag{3-39}$$

onde λ é um número positivo, $x \in \mathbb{R}^n$ e para todo $i = 1, \ldots, m$, A_i são matrizes $p \times n$ e b_i vetores em $\mathbb{R}^p$. Os experimentos foram feitos no programa MATLAB e escolhemos $n = 4$, $p = 3$, $\lambda = 1/m$ e m igual aos valores, 100 e 1000.

No problema (3-39) as matrizes A_i são geradas aleatóriamente, ou seja, cada uma de suas entradas são valores tomados aleatóriamente no intervalo $[-10, 10]$. Também construimos este problema de tal forma que $f(\bar{x}) = 0$, ou seja, de modo que $\bar{x} \in \mathbb{R}^n$ é uma solução.

Além de verificar a eficiência dos algoritmos estudados, com os experimentos númericos, desejamos também, comparar a performace do método do subgradiente clássico (1-2) e o método do subgradiente incremental (Algoritmo 3.1) aplicados para solucionar o problema (3-39) usando diferentes escolhas de comprimento de passo. Escolhemos duas regras comprimento de passo.

- Regra comprimento de passo pré-determinado definido por

$$\alpha_k = \frac{D}{k+1} \qquad \forall\, k \geq 0,$$

onde D é uma constante positiva, e neste caso, adotamos regra de parada como um certo número máximo de iterações k, ou erro absoluto $\|x^k - \bar{x}\| \leq 10^{-3}$.

- Regra comprimento de passo dinâmico dado pelo Algoritmo 3.2, onde definimos a constante $B = 100$ e usamos regra de parada como um certo número máximo de iterações k, ou erro $\left|f(x^k) - f^*\right| \leq 10^{-3}$.

Os resultandos dos experimentos estão resumidos nas tabelas seguintes e mostram, em cada método avaliado, o número de iterações k feitas até que ocorra o critério de parada, o valor funcional $f(x^k)$, onde x^k é o último ponto gerado, e o tempo T, em segundos, gasto para processar essas iterações. Definimos o ponto inicial x^0, em todos os experimentos, como sendo a origem.

Tabela 3.1: $n = 4,\ m = 100,\ x^0 = 0,\ f^* = 0.$ *Critério de parada,* $k = 5000$ *iterações ou* $\|x^k - \bar{x}\| \leq 10^{-3}$.

Regra comprimento de passo pré-determinado $\alpha_k = D/(k+1)$						
	Subgradiente clássico			Subgradiente incremental		
D	$f(x^k)$	k	T	$f(x^k)$	k	T
1	7.6×10^7	5000	52.9	2.1×10^8	5000	69.3
0.05	7.9×10^7	5000	51.41	0.31	470	7.92
0.007	0.53	4150	57.28	0.14	66	0.85
0.001	0.43	642	9.28	7.2×10^{-6}	10	0.15
0.0005	0.42	321	4.49	5.3×10^{-6}	6	0.06

Tabela 3.2: $n = 4,\ m = 1000,\ x^0 = 0,\ f^* = 0.$ *Critério de parada,* $k = 5000$ *iterações ou* $\|x^k - \bar{x}\| \leq 10^{-3}$.

Regra comprimento de passo pré-determinado $\alpha_k = D/(k+1)$						
	Subgradiente clássico			Subgradiente incremental		
D	$f(x^k)$	k	T	$f(x^k)$	k	T
1	7.2×10^8	5000	604.47	2.9×10^{10}	5000	588.50
0.05	6.0×10^8	5000	557.98	0.13	469	60.24
0.007	6.5×10^8	5000	567.75	0.20	67	14.07
0.001	1.0×10^9	5000	573.41	9.0×10^{-5}	10	2.25
0.0005	5.15	2725	296.67	1.0×10^{-6}	5	0.76

Tabela 3.3: $n = 4,\ m = 100,\ x^0 = 0,\ f^* = 0$. *Critério de parada, $k = 5000$ iterações ou $|f(x^k) - f^*| \leq 10^{-3}$.*

Regra comprimento de passo dinâmico dado pelo Algoritmo 3.2						
	Subgradiente clássico			Subgradiente incremental		
δ_0	$f(x^k)$	k	T	$f(x^k)$	k	T
6×10^6	9.9×10^{-4}	143	1.95	9.8×10^{-4}	84	3.03
7×10^6	9.3×10^{-4}	65	0.95	7.2×10^{-4}	16	0.34
8×10^6	9.0×10^{-4}	30	0.37	4.5×10^{-4}	29	0.69
9×10^6	5.7×10^{-4}	26	0.23	5.9×10^{-4}	15	0.26
1×10^7	6.9×10^{-4}	23	0.24	3.9×10^{-4}	9	0.23

Tabela 3.4: $n = 4,\ m = 1000,\ x^0 = 0,\ f^* = 0$. *Critério de parada, $k = 5000$ iterações ou $|f(x^k) - f^*| \leq 10^{-3}$.*

Regra comprimento de passo dinâmico dado pelo Algoritmo 3.2						
	Subgradiente clássico			Subgradiente incremental		
δ_0	$f(x^k)$	k	T	$f(x^k)$	k	T
6×10^6	9.8×10^{-4}	901	218.51	9.3×10^{-4}	182	44.51
7×10^6	9.9×10^{-4}	1082	176.07	9.9×10^{-4}	192	26.75
8×10^6	9.9×10^{-4}	640	80.82	8.9×10^{-4}	116	28.86
9×10^6	9.9×10^{-4}	481	86.76	9.8×10^{-4}	95	14.82
1×10^7	9.5×10^{-4}	339	56.64	9.4×10^{-4}	119	29.88

CAPÍTULO 4

Método Proximal-Subgradiente Incremental

Visando ampliar sob novos aspectos o método incremental, discutiremos nesse capítulo alguns algoritmos cujas iterações são combinações do método do ponto proximal e subgradiente. Estes resultados foram estudados especialmente no artigo [2].

Nesse capítulo continuaremos considerando problemas da forma

$$\min f(x) := \sum_{i=1}^{m} f_i(x), \quad \text{sujeito a} \quad x \in X, \tag{4-1}$$

onde $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$ são funções convexas não necessariamente diferenciáveis, e $X \subseteq \mathbb{R}^n$ é um conjunto não vazio, fechado e convexo.

As iterações do método do subgradiente incremental, estudado neste capítulo, são dadas da seguinte forma

$$x^{k+1} = P_X\left(x^k - \alpha_k \tilde{\nabla} f_{i_k}(x^k)\right), \tag{4-2}$$

onde α_k é um passo positivo, $P_X(.)$ é a projeção em X, $\tilde{\nabla} f_{i_k}(x^k)$ é um subgradiente de f_{i_k} em x^k, e $i_k \in \{1, \ldots, m\}$ representa o índice da função componente a ser considerada na k-ésima iteração. Embora, existam várias formas de escolha do índice i_k, consideraremos a seguinte regra de atualização

$$i_k = (k \text{ modulo } m) + 1, \quad \text{ou seja,} \quad i_k = (\text{resto da divisão de } k \text{ por } m) + 1.$$

Também iremos supor que α_k é constante em um ciclo, ou seja, para todo $l = 0, 1, \ldots$, definimos $\alpha_{lm} = \alpha_{lm+1} = \cdots = \alpha_{lm+m-1}$. Desta forma, se k é o início de um cíclo, i. e., $i_k = 1$, comparando a sequência gerada no método subgradiente incremental do Capítulo 3 vemos que $\psi^{m,k}$ corresponde a iterada x^{k+m} no método (4-2).

4.1 Método Proximal Incremental

Nesta seção faremos uma breve introdução do método proximal e em seguida apresentaremos o modelo proximal subgradiente, que é o objetivo central de estudo neste capítulo. Quando $f : \mathbb{R}^n \to \mathbb{R}$ é um função convexa, o trabalho de Iusem em [8] mostra que um sequência gerada pelo método proximal

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

com $\lambda_k \leq \bar{\lambda}$, onde $\bar{\lambda} > 0$ converge para um ponto ótimo, caso exista. Essa convergência é obtida para qualquer escolha de λ_k. É conhecido que o método do ponto proximal é geralmente mais estável que o método do subgradiente, devido ao termo de regularização quadrática.

Agora combinaremos a metodologia incremental do método descrito em (4-2) com o método proximal acima. Tentando solucionar o problema (4-1), uma alternativa simples seria gerar uma sequência $\{x^k\}$ em $\mathbb{R}^n$ da forma

$$x^{k+1} = \arg\min_{x \in X} \left\{ f_{i_k}(x) + \frac{1}{2\alpha_k} \|x - x^k\|^2 \right\}, \tag{4-3}$$

onde assumimos que $\{\alpha_k\}$ é uma sequência de números reais positivos, cada $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$ é convexa e $X \subseteq \mathbb{R}^n$ é um conjunto não vazio, fechado e convexo.

Se algumas funções componentes f_i possuirem uma estrutura favorável, a iteração proximal (4-3) pode ser obtida de forma relativamente simples, e nesse caso pode ser conveniente adotar o método incremental proximal ao invéz do subgradiente (4-2). No entanto, para outras funções componentes, pode ser inconveniente a iteração (4-3). Então, podemos considerar combinações de iterações subgradiente e proximal.

Consideraremos agora um problema de otimização da forma

$$\min F(x) := \sum_{i=1}^{m} F_i(x), \quad \text{sujeito a} \quad x \in X, \tag{4-4}$$

onde para todo $i = 1, \ldots, m$

$$F_i(x) := f_i(x) + h_i(x), \tag{4-5}$$

$f_i : \mathbb{R}^n \to \mathbb{R}$ e $h_i : \mathbb{R}^n \to \mathbb{R}$ são funções convexas a valores reais e $X \subseteq \mathbb{R}^n$ é um conjunto não vazio fechado e convexo.

Na próxima seção apresentaremos, entre outras coisas, alguns métodos incrementais que operam, na componente f_i com uma iteração proximal e na componente h_i com uma iteração subgradiente. Em particular, escolhendo todas as componentes f_i ou todas as h_i como sendo identicamente nulas, obtem-se a iteração proximal ou subgradiente, (4-3) ou (4-2) respectivamente. Na seção 4.3 apresentaremos os resultados de convergência desses métodos.

4.2 Método Proximal-Subgradiente Incremental

Nesta seção vamos apresentar alguns métodos incrementais que envolvem combinações de uma iteração proximal e uma subgradiente. Na tentativa de solucionar o problema (4-4)-(4-5) um desses métodos possui a forma

$$z^k = \arg\min_{x\in X}\left\{f_{i_k}(x)+\frac{1}{2\alpha_k}\|x-x^k\|^2\right\} \tag{4-6}$$

$$x^{k+1} = P_X\left(z^k-\alpha_k\tilde{\nabla}h_{i_k}(z^k)\right), \tag{4-7}$$

onde $\tilde{\nabla}h_{i_k}(z^k)$ é um subgradiente qualquer de h_{i_k} em z^k. Veja que o mínimo na equação (4-6) é atingido e é único, pois todas as f_i são contínuas e convexas e $\|\cdot-x^k\|^2$ é uma função real fortemente convexa, logo a soma de f_i com $\|\cdot-x^k\|^2$ será fortemente convexa e portanto possui um único minimizador em X (pois é fechado e convexo), portanto z^k está bem definido. Agora, o subdiferencial $\partial h_i(z^k)$ é não vazio para todo $i=1,\dots,m$ pois h_i é uma função real convexa. Isso mostra que as iterações (4-6)-(4-7) estão bem definidas.

Como já foi dito, note que se todos as f_i são identicamente nulas, então $z^k=x^k$ e a iteração (4-7) será a iteração subgradiente incremental (4-2) (com h_{i_k} no lugar de f_{i_k}). Agora se todas as h_i são identicamente nulas claro que o subdiferencial $\partial h_{i_k}(z^k)$ é 0, então $z^k=x^{k+1}$ e a iteração (4-6) será a mesma iteração (4-3), ou seja, uma iteração proximal.

Da forma como as iterações (4-6) e (4-7) foram definidas, as sequências $\{z^k\}$ e $\{x^k\}$ estão no conjunto X. Pode ser conveniente efetuar primeiro a iteração z^k na componente f_i pelo método proximal em todo o espaço $\mathbb{R}^n$ e logo após, atualizar x^{k+1}, para a componente h_i, no método subgradiente restrito ao conjunto X. Deste modo, o método proximal-subgradiente incremental fica da forma

$$z^k = \arg\min_{x\in\mathbb{R}^n}\left\{f_{i_k}(x)+\frac{1}{2\alpha_k}\|x-x^k\|^2\right\} \tag{4-8}$$

$$x^{k+1} = P_X\left(z^k-\alpha_k\tilde{\nabla}h_{i_k}(z^k)\right). \tag{4-9}$$

Com raciocínio análogo, podemos fazer a iteração de z^k pelo método subgradiente, na componente h_i, em todo espaço $\mathbb{R}^n$ e só depois efetuar a iteração de x^{k+1} pelo método proximal restrito ao conjunto X. Este algoritmo que é também uma combinação proximal incremental com subgradiente incremental é apresentado do seguinte modo

$$z^k = x^k - \alpha_k \tilde{\nabla} h_{i_k}(x^k) \tag{4-10}$$

$$x^{k+1} = \arg\min_{x \in X} \left\{ f_{i_k}(x) + \frac{1}{2\alpha_k} \|x - z^k\|^2 \right\}. \tag{4-11}$$

Certamente é possível usar diferentes sequências comprimento de passo $\{\alpha_k\}$ no método proximal-subgradiente incremental, no entanto, este assunto não será abordado neste capítulo.

O próximo resultado apresenta uma caracterização da iteração do ponto proximal em termos do subdiferencial. Além disso mostra uma estimativa utilizada em resultados de convergência.

Proposição 4.1 *Seja $X \subseteq \mathbb{R}^n$ um conjunto não vazio, fechado e convexo. Suponha que $f : \mathbb{R}^n \to \mathbb{R}$ é uma função convexa tal que $\mathrm{int}(X) \neq \emptyset$. Para qualquer $x^k \in \mathbb{R}^n$ e $\alpha_k > 0$, considere a iteração proximal*

$$x^{k+1} = \arg\min_{x \in X} \left\{ f(x) + \frac{1}{2\alpha_k} \|x - x^k\|^2 \right\} \tag{4-12}$$

(a) *A iteração (4-12) pode ser escrita como*

$$x^{k+1} = P_X\left(x^k - \alpha_k \tilde{\nabla} f(x^{k+1}) \right),$$

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

(b) *Para todo $y \in X$, temos*

$$\begin{aligned} \|x^{k+1} - y\|^2 &\leq \|x^k - y\|^2 - 2\alpha_k \left(f(x^{k+1}) - f(y) \right) - \|x^k - x^{k+1}\|^2 \\ &\leq \|x^k - y\|^2 - 2\alpha_k \left(f(x^{k+1}) - f(y) \right). \end{aligned} \tag{4-13}$$

Prova. Mostraremos inicialmente que x^{k+1} em (4-12) está bem definido. Para isto escreva

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

Como f é uma função convexa e $1/2\alpha_k \| \cdot - x^k \|^2$ é fortemente convexa, segue que f_k é fortemente convexa e contínua, logo possui um único minimizador em X, ou seja, x^{k+1} está bem definido. Agora, sabemos que o subdiferencial da soma de funções convexas é a soma dos subdiferenciais destas funções. Assim,

$$\partial f_k(x) = \partial f(x) + \partial\left(\frac{1}{2\alpha_k}\|x - x^k\|^2\right) = \partial f(x) + \frac{1}{\alpha_k}\left(x - x^k\right).$$

Como x^{k+1} é o minimizador de f_k, pela condição de otimalidade para minimização de função convexa em conjunto convexo (Teorema 2.26), temos

$$0 \in \partial f_k(x^{k+1}) + N_X(x^{k+1}) = \partial f(x^{k+1}) + \frac{1}{\alpha_k}\left(x^{k+1} - x^k\right) + N_X(x^{k+1}),$$

onde $N_X(x^{k+1}) = \left\{y \in \mathbb{R}^n \mid \left\langle y, x - x^{k+1}\right\rangle \leq 0, \quad \forall\, x \in X\right\}$ é o cone normal de X em x^{k+1}. A relação acima equivale a

$$\frac{1}{\alpha_k}\left(x^k - x^{k+1}\right) \in \partial f(x^{k+1}) + N_X(x^{k+1}) \tag{4-14}$$

Assim, temos que (4-14) é verdadeiro se e somente se

$$x^k - x^{k+1} - \alpha_k \tilde{\nabla} f(x^{k+1}) = \left(x^k - \alpha_k \tilde{\nabla} f(x^{k+1})\right) - x^{k+1} \in N_X(x^{k+1})$$

para algum $\tilde{\nabla} f(x^{k+1}) \in \partial f(x^{k+1})$. Sendo $x^{k+1} \in X$, pelo Teorema da projeção 2.20 resulta que

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

Para mostrar o item **(b)** observe inicialmente que para qualquer $y \in X$, temos

$$\|x^k - y\|^2 = \|x^k - x^{k+1}\|^2 + 2\langle x^k - x^{k+1}, x^{k+1} - y\rangle + \|x^{k+1} - y\|^2. \tag{4-15}$$

Por (4-14) existem vetores $d_1 \in \partial f(x^{k+1})$ e $d_2 \in N_X(x^{k+1})$ tais que $1/\alpha_k(x^k - x^{k+1}) = d_1 + d_2$. Sendo $y \in X$, pela definição de subgradiente e cone normal temos

$$f(y) \geq f(x^{k+1}) + \langle d_1, y - x^{k+1}\rangle \quad \text{e} \quad 0 \geq \langle d_2, y - x^{k+1}\rangle,$$

somando membro a membro essas duas desigualdades obtemos

$$f(y) \geq f(x^{k+1}) + \frac{1}{\alpha_k}\langle x^k - x^{k+1}, y - x^{k+1}\rangle,$$

após simples manipulações algébricas a desigualdade acima pode ser escrita como

$$\langle x^k - x^{k+1}, x^{k+1} - y\rangle \geq \alpha_k\left(f(x^{k+1}) - f(y)\right). \tag{4-16}$$

Logo, combinando as desigualdades (4-15) e (4-16) conclui-se que

$$\begin{aligned}\|x^{k+1}-y\|^2 &= \|x^k-y\|^2-2\langle x^k-x^{k+1},x^{k+1}-y\rangle-\|x^{k+1}-x^k\|^2\\ &\leq \|x^k-y\|^2-2\alpha_k\left(f(x^{k+1})-f(y)\right)-\|x^{k+1}-x^k\|^2.\end{aligned}$$

□

Como resultado da Proposição 4.1 **(a)** vemos que todos os algoritmos descritos nesta seção podem ser reescritos no formato subgradiente incremental:

1. As iterações no método (4-6)-(4-7) podem ser reescritas na forma

$$z^k = P_X\left(x^k-\alpha_k\tilde{\nabla}f_{i_k}(z^k)\right), \qquad x^{k+1}=P_X\left(z^k-\alpha_k\tilde{\nabla}h_{i_k}(z^k)\right); \tag{4-17}$$

2. As iterações no método (4-8)-(4-9) podem ser reescritas na forma

$$z^k = x^k-\alpha_k\tilde{\nabla}f_{i_k}(z^k), \qquad x^{k+1}=P_X\left(z^k-\alpha_k\tilde{\nabla}h_{i_k}(z^k)\right); \tag{4-18}$$

3. As iterações no método (4-10)-(4-11) podem ser reescritas na forma

$$z^k = x^k-\alpha_k\tilde{\nabla}h_{i_k}(x^k), \qquad x^{k+1}=P_X\left(z^k-\alpha_k\tilde{\nabla}f_{i_k}(x^{k+1})\right). \tag{4-19}$$

Observe que em todas as atualizações anteriores o subgradiente $\tilde{\nabla}h_{i_k}$ é calculado em um ponto que não depende da iteração que está sendo realizada, assim pode ser qualquer vetor no subdiferencial ∂h_{i_k}. Contudo, o subgradiente $\tilde{\nabla}f_{i_k}$ depende da própria iteração que está sendo processada, mais precisamente, $\tilde{\nabla}f_{i_k}$ é um vetor específico no subdiferencial ∂f_{i_k}, fornecido de acordo com a Proposição 4.1. Lembrando também que $F_i(x)=f_i(x)+h_i(x)$, se na iteração (4-18) subtituirmos o vetor z^k na atualização de x^{k+1}, podemos ainda escrever (4-18) como

$$x^{k+1}=P_X\left(x^k-\alpha_k\tilde{\nabla}F_{i_k}(z^k)\right)$$

que é semelhante ao método do subgradiente incremental (4-2) com a única diferença que o subgradiente $\tilde{\nabla}F_{i_k}$ é tomado no ponto z^k ao invéz de ser em x^k.

Outra questão importante é como escolher as componentes em cada iteração. Neste capítulo, assumiremos sempre a ordem cíclica, onde $\{f_i,h_i\}$ serão escolhidas na ordem $1,\ldots,m$ e terminando esse processo retoma um novo ciclo, assim definimos $i_k=(k \text{ modulo } m)+1$. Um ciclo é formado por iterações envolvendo $\{f_1,h_1\},\cdots,\{f_m,h_m\}$, nesta ordem e exatamente uma vez. Consideramos também que o comprimento de passo α_k é constante ao longo de um ciclo, ou seja, para todo k que inicia um ciclo, i. e., $i_k=1$

temos $\alpha_k = \alpha_{k+1} = \cdots = \alpha_{k+m-1}$.

Com notação semelhante ao Capítulo 3, denotaremos por F^* o valor ótimo do problema (4-4)

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

por X^* o conjunto das soluções ótimas (que pode ser vazio)

$$X^* = \{x^* \in X \mid F(x^*) = F^*\},$$

e sendo X um conjunto convexo, fechado e não vazio, denotamos por $\text{dist}(\cdot, X)$ a função distância definida em $\mathbb{R}^n$ dada por

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

4.3 Resultados para Métodos com Ordem Cíclica

Reservamos esta última seção para discutirmos a convergência dos métodos apresentados adotando a ordem cíclica. As propriedades que serão mostradas se referem a sequência $\{x^k\}$ pois a sequência $\{z^k\}$ não necessariamente se encontra no conjunto X, no caso das iterações (4-18) e (4-19) quando $X \neq \mathbb{R}^n$. Como será visto, esta análise de convergência é semelhante a do subgradiente incremental apresentada no Capítulo 3 com algumas mudanças nas hipóteses requeridas e nos resultados obtidos. No restante dessa seção usaremos as seguintes hipóteses:

Hipótese 4.2 (Para as iterações (4-17) e (4-18)) *Existe uma constante $c \in \mathbb{R}$ tal que para todo k*

$$\max\left\{\|\tilde{\nabla} f_{i_k}(z^k)\|, \|\tilde{\nabla} h_{i_k}(z^k)\|\right\} \leq c.$$

Além disso, para todo k que inicia um ciclo (i. e. $k \geq 1$ com $i_k = 1$) temos

$$\max\left\{f_j(x^k) - f_j(z^{k+j-1}), h_j(x^k) - h_j(z^{k+j-1})\right\} \leq c\|x^k - z^{k+j-1}\|, \quad \forall j = 1, \dots, m.$$

Hipótese 4.3 (Para a iteração (4-19)) *Existe uma constante $c \in \mathbb{R}$ tal que para todo k*

$$\max\left\{\|\tilde{\nabla} f_{i_k}(x^{k+1})\|, \|\tilde{\nabla} h_{i_k}(x^k)\|\right\} \leq c.$$

Além disso, para todo k que inicia um ciclo (i. e. $k \geq 1$ com $i_k = 1$) temos

$$\max\left\{f_j(x^k) - f_j(x^{k+j-1}), h_j(x^k) - h_j(x^{k+j-1})\right\} \leq c\|x^k - x^{k+j-1}\|, \quad \forall j = 1, \dots, m$$

$$f_j(x^{k+j-1}) - f_j(x^{k+j}) \leq c\|x^{k+j-1} - x^{k+j}\|, \quad \forall j = 1, \dots, m.$$

Se para todo $i = 1, \ldots, m$ e todo k, os subgradientes de f_i em x^k e os subgradientes de h_i em x^k, são, em norma, limitados por c, então a Hipótese 4.2 é satisfeita. Basta usar a definição de subgradiente e a desigualdade de Cauchy-Shwarz para obter

$$f_i(x^k) - f_i(z^{k+j-1}) \leq \langle \tilde{\nabla} f_i(x^k), x^k - z^{k+j-1} \rangle \leq c\|x^k - z^{k+i-1}\|, \quad j = 1, \ldots, m,$$

para a componente f_i, e de forma análoga para a componente h_i.

Como estamos sempre considerando que as componentes f_i e h_i, $i = 1, \ldots, m$ são funções convexas, verifica-se facilmente que:

1. Para o método iterativo (4-17), se todas as componentes f_i e h_i, $i = 1, \ldots, m$ são Lipschitz contínuas sobre o conjunto X então a Hipótese 4.2 será satisfeita;
2. Para os métodos iterativos (4-18) e (4-19), se todas as componentes f_i e h_i, $i = 1, \ldots, m$ são Lipschitz contínuas sobre o espaço inteiro $\mathbb{R}^n$, então as Hipóteses 4.2 e 4.3 serão satisfeitas;
3. Para todos os métodos iterativos (4-17), (4-18) e (4-19), se as sequências $\{x^k\}$ e $\{z^k\}$ são limitadas, então as Hipóteses 4.2 e 4.3 serão satisfeitas (pois nesse caso, as componentes f_i e h_i, que são funções convexas a valores reais, serão Lipschitz contínuas sobre qualquer conjunto limitado que contém $\{x^k\}$ e $\{z^k\}$).

Provaremos a seguir uma estimativa que fornece a ferramenta necessária para a análise de convergência do método incremental subgradiente-proximal. Esse resultado é semelhante ao Lema 3.3 do Capítulo 3.

Lema 4.4 *Seja $\{x^k\}$ uma sequência gerada por qualquer um dos algoritmos (4-17), (4-18) ou (4-19), com ordem cíclica de seleção das componentes. Então, para todo $y \in X$ e todo indice k que inicia um ciclo (i. e., $k \geq 1$ com $i_k = 1$), temos*

$$\|x^{k+m} - y\|^2 \leq \|x^k - y\|^2 - 2\alpha_k \left(F(x^k) - F(y) \right) + \beta \alpha_k^2 m^2 c^2.$$

onde $\beta = \frac{1}{m} + 4$.

Prova. A demonstração será feita em duas etapas. Na primeira parte provaremos o resultado para os algoritmos (4-17) e (4-18). Na segunda parte faremos as modificações necessárias para adaptar ao algoritmo (4-19). Usando a Proposição 4.1 item **(b)** temos para todo $y \in X$ e todo k

$$\|z^k - y\|^2 \leq \|x^k - y\|^2 - 2\alpha_k \left(f_{i_k}(z^k) - f_{i_k}(y) \right). \tag{4-20}$$

Temos também, pela propriedade não expansiva do operador projeção, a definição de subgradiente (onde, $\tilde{\nabla} h_{i_k}(z^k) \in \partial h_{i_k}(z^k) \Rightarrow \langle \tilde{\nabla} h_{i_k}(z^k), z^k - y \rangle \geq h_{i_k}(z^k) - h_{i_k}(y)$), e a Hipótese

4.2 que para todo $y \in X$ e todo k,

$$
\begin{aligned}
\left\|x^{k+1}-y\right\|^2 &= \left\|P_X\left(z^k-\alpha_k\tilde{\nabla}h_{i_k}(z^k)\right)-P_X(y)\right\|^2 \leq \left\|z^k-\alpha_k\tilde{\nabla}h_{i_k}(z^k)-y\right\|^2 \\
&\leq \left\|z^k-y\right\|^2-2\alpha_k\left\langle\tilde{\nabla}h_{i_k}(z^k),z^k-y\right\rangle+\alpha_k^2\left\|\tilde{\nabla}h_{i_k}(z^k)\right\|^2 \\
&\leq \left\|z^k-y\right\|^2-2\alpha_k\left(h_{i_k}(z^k)-h_{i_k}(y)\right)+\alpha_k^2c^2. \qquad (4\text{-}21)
\end{aligned}
$$

Combinando as relações (4-20) e (4-21) e usando a definição $F_j = f_j + h_j$, temos

$$
\begin{aligned}
\|x^{k+1}-y\|^2 &\leq \|x^k-y\|^2-2\alpha_k\left(f_{i_k}(z^k)+h_{i_k}(z^k)-f_{i_k}(y)-h_{i_k}(y)\right)+\alpha_k^2c^2 \\
&= \|x^k-y\|^2-2\alpha_k\left(F_{i_k}(z^k)-F_{i_k}(y)\right)+\alpha_k^2c^2. \qquad (4\text{-}22)
\end{aligned}
$$

Seja agora k o início de um ciclo (i.e. $i_k = 1$). Em vista da ordem cíclica de seleção das componentes, para as iterações $k+j-1$, onde $j = 1,\dots,m$, as componentes selecionadas são $\{f_j,h_j\}$, $j = 1,\dots,m$ nesta ordem. Podemos então repetir a desigualdade anterior em cadeia para os índices $k+1,\dots,k+m-1$ obtendo

$$
\|x^{k+m}-y\|^2 \leq \|x^k-y\|^2-2\alpha_k\sum_{j=1}^{m}\left(F_j(z^{k+j-1})-F_j(y)\right)+m\alpha_k^2c^2.
$$

Lembrando que $F(x) = \sum_{j=1}^m F_j(x)$ podemos ainda reescrever essa desigualdade da forma

$$
\begin{aligned}
\|x^{k+m}-y\|^2 &\leq \|x^k-y\|^2-2\alpha_k\sum_{j=1}^{m}\left(F_j(x^k)-F_j(y)-F_j(x^k)+F_j(z^{k+j-1})\right)+m\alpha_k^2c^2 \\
&= \|x^k-y\|^2-2\alpha_k\left(F(x^k)-F(y)\right)+m\alpha_k^2c^2+ \\
&\quad 2\alpha_k\sum_{j=1}^{m}\left(F_j(x^k)-F_j(z^{k+j-1})\right). \qquad (4\text{-}23)
\end{aligned}
$$

Mostraremos que o último termo da desigualdade (4-23) é limitado. Pela Hipótese 4.2 para todo $j = 1,\dots,m$ temos

$$
f_j(x^k)-f_j(z^{k+j-1}) \leq c\|x^k-z^{k+j-1}\| \qquad \text{e} \qquad h_j(x^k)-h_j(z^{k+j-1}) \leq c\|x^k-z^{k+j-1}\|.
$$

Logo, somando membro a membro as duas expressões, pode-se assegurar que, para todo $j = 1,\dots,m$

$$
F_j(x^k)-F_j(z^{k+j-1}) \leq 2c\|x^k-z^{k+j-1}\|. \qquad (4\text{-}24)
$$

Usando a desigualdade triangular, podemos concluir ainda que

$$\begin{aligned}\|x^k - z^{k+j-1}\| &\leq \|x^k - x^{k+1}\| + \|x^{k+1} - z^{k+j-1}\| \\ &\vdots \\ &\leq \|x^k - x^{k+1}\| + \cdots + \|x^{k+j-1} - z^{k+j-1}\|. \end{aligned} \tag{4-25}$$

Pela definição dos algoritmos (4-17) e (4-18), a propriedade não expansiva da projeção, e a Hipótese 4.2, para $l = k, \ldots, k+j-2$ temos

$$\begin{aligned}\left\|x^l - x^{l+1}\right\| &= \left\|x^l - P_X(z^l - \alpha_l \tilde{\nabla} h_{i_l}(z^l))\right\| \leq \left\|x^l - z^l + \alpha_l \tilde{\nabla} h_{i_l}(z^l)\right\| \\ &\leq \left\|x^l - z^l\right\| + \alpha_l \left\|\tilde{\nabla} h_{i_l}(z^l)\right\| \\ &\leq \left\|x^l - x^l - \alpha_l \tilde{\nabla} f_{i_l}(z^l)\right\| + \alpha_l c \leq 2\alpha_l c, \end{aligned} \tag{4-26}$$

e para $l = k+j-1$, temos $\|x^l - z^l\| \leq \alpha_l \|\tilde{\nabla} f_{i_l}(z^l)\| \leq \alpha_l c$. Então, de (4-25) e sendo $\alpha_k = \alpha_{k+1} = \cdots = \alpha_{k+m-1}$, resulta que

$$\|x^k - z^{k+j-1}\| \leq (j-1)2\alpha_k c + \alpha_k c = \alpha_k c(2j-1).$$

Combinando esta relação com a obtida em (4-24), temos

$$F_j(x^k) - F_j(z^{k+j-1}) \leq 2\alpha_k c^2 (2j-1). \tag{4-27}$$

Somando, sobre o índice $j = 1$ até $j = m$ a desigualdade (4-27) acima, e relacionando com (4-23) conclui-se que

$$\|x^{k+m} - y\|^2 \leq \|x^k - y\|^2 - 2\alpha_k \left(F(x^k) - F(y)\right) + m\alpha_k^2 c^2 + 4\alpha_k^2 c^2 \sum_{j=1}^{m} (2j-1).$$

Observe que $\sum_{j=1}^{m} (2j-1) = (1+2m-1)m/2 = m^2$, e portanto podemos escrever

$$m\alpha_k^2 c^2 + 4\alpha_k^2 c^2 \sum_{j=1}^{m} (2j-1) = m\alpha_k^2 c^2 + 4\alpha_k^2 c^2 m^2 = \alpha_k^2 c^2 m^2 \left(\frac{1}{m} + 4\right)$$

e sendo $\beta = \frac{1}{m} + 4$ isso conclui a primeira parte do resultado desejado. Para a segunda parte da demonstração, vemos pela definição do algoritmo (4-19), que para todo $y \in X$ e todo $k \geq 0$

$$\|z^k - y\|^2 = \|x^k - y - \alpha_k \tilde{\nabla} h_{i_k}(x^k)\|^2 \leq \|x^k - y\|^2 - 2\alpha_k \langle \tilde{\nabla} h_{i_k}(x^k), x^k - y\rangle + \alpha_k^2 \|\tilde{\nabla} h_{i_k}(x^k)\|^2.$$

Da definição de subgradiente $\tilde{\nabla} h_{i_k}(x^k) \in \partial h_{i_k}(x^k) \Rightarrow \langle \tilde{\nabla} h_{i_k}(x^k), x^k - y \rangle \geq h_{i_k}(x^k) - h_{i_k}(y)$, e pela Hipótese 4.3 segue que

$$\|z^k - y\|^2 \leq \|x^k - y\|^2 - 2\alpha_k \left(h_{i_k}(x^k) - h_{i_k}(y) \right) + \alpha_k^2 c^2. \tag{4-28}$$

Usando a Proposição 4.1 ítem **(b)** temos para todo $y \in X$ e todo $k \geq 0$

$$\|x^{k+1} - y\|^2 \leq \|z^k - y\|^2 - 2\alpha_k \left(f_{i_k}(x^{k+1}) - f_{i_k}(y) \right). \tag{4-29}$$

As desigualdades (4-28) e (4-29) combinadas resultam em uma relação semelhante a (4-22), pois

$$\|x^{k+1} - y\|^2 \leq \|x^k - y\|^2 - 2\alpha_k \left(h_{i_k}(x^k) - f_{i_k}(y) - h_{i_k}(y) + f_{i_k}(x^{k+1}) \right) + \alpha_k^2 c^2$$

então, somando $0 = f_{i_k}(x^k) - f_{i_k}(x^k)$ dentro dos parenteses do lado direito, e sendo $F_{i_k} = f_{i_k} + h_{i_k}$ obtemos

$$\|x^{k+1} - y\|^2 \leq \|x^k - y\|^2 - 2\alpha_k \left(F_{i_k}(x^k) - F_{i_k}(y) \right) + \alpha_k^2 c^2 + 2\alpha_k \left(f_{i_k}(x^k) - f_{i_k}(x^{k+1}) \right).$$

De forma análoga à primeira parte, suponha que k marca o inicio de um ciclo (i.e. $i_k = 1$). Repetindo a desigualdade anterior, com k substituido por $k+1, \ldots, k+m-1$ e combinando as relações obtidas segue que

$$\begin{aligned} \|x^{k+m} - y\|^2 \leq \;& \|x^k - y\|^2 - 2\alpha_k \sum_{j=1}^{m} \left(F_j(x^{k+j-1}) - F_j(y) \right) + \\ & m\alpha_k^2 c^2 + 2\alpha_k \sum_{j=1}^{m} \left(f_j(x^{k+j-1}) - f_j(x^{k+j}) \right). \end{aligned}$$

Somando $0 = F_j(x^k) - F_j(x^k)$ dentro do primeiro parenteses do lado direito da relação acima, e lembrando que $F = \sum_{j=1}^{m} F_j$, obtemos

$$\begin{aligned} \|x^{k+m} - y\|^2 \leq \;& \|x^k - y\|^2 - 2\alpha_k \left(F(x^k) - F(y) \right) + 2\alpha_k \sum_{j=1}^{m} \left(F_j(x^k) - F_j(x^{k+j-1}) \right) \\ & + m\alpha_k^2 c^2 + 2\alpha_k \sum_{j=1}^{m} \left(f_j(x^{k+j-1}) - f_j(x^{k+j}) \right). \end{aligned} \tag{4-30}$$

Agora, usando a Hipótese 4.3 obtemos uma desigualdade semelhante a (4-24), apenas com x^{k+j-1} no lugar de z^{k+j-1}

$$\begin{aligned} F_j(x^k) - F_j(x^{k+j-1}) &\leq 2c\|x^k - x^{k+j-1}\| \\ &\leq 2c\left(\|x^k - x^{k+1}\| + \cdots + \|x^{k+j-2} - x^{k+j-1}\|\right). \end{aligned} \tag{4-31}$$

Pela definição do algoritmo (4-19), a Hipótese 4.3 e a propriedade não expansiva do operador projeção, temos para cada $l = k, \ldots, k+j-1$

$$\begin{aligned} \left\|x^l - x^{l+1}\right\| &= \left\|x^l - P_X(z^l - \alpha_l \tilde{\nabla} f_{i_l}(x^{l+1}))\right\| \leq \left\|x^l - z^l + \alpha_l \tilde{\nabla} f_{i_l}(x^{l+1})\right\| \\ &= \left\|x^l - x^l + \alpha_l \tilde{\nabla} h_{i_l}(x^l) + \alpha_l \tilde{\nabla} f_{i_l}(x^{l+1})\right\| \leq 2\alpha_l c. \end{aligned} \tag{4-32}$$

Por (4-32), cada termo em norma do lado direito de (4-31) é limitado por $2\alpha_k c$ (pois $\alpha_k = \alpha_{k+1} = \cdots = \alpha_{k+m-1}$), então $F_j(x^k) - F_j(x^{k+j-1}) \leq 4\alpha_k c^2 (j-1)$. Novamente, por (4-32) e pela Hipótese 4.3 temos também $f_j(x^{k+j-1}) - f_j(x^{k+j}) \leq c\|x^{k+j-1} - x^{k+j}\| \leq 2c^2\alpha_k$. Destes resultados, podemos concluir que

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

Portanto, substituindo esse resultado na expressão (4-30) com $\beta = \left(\frac{1}{m} + 4\right)$ concluimos finalmente a segunda parte da demonstração. □

O Lema 4.4 garante que, entre outras coisas, dado um iterando x^k no início de um ciclo, e qualquer ponto $y \in X$ onde $F(y) \leq F(x^k)$ (por exemplo um ponto ótimo), os algoritmos (4-17), (4-18) ou (4-19) fornecem um ponto x^{k+m}, no final do cíclo, que será mais próximo de y que x^k, desde que o comprimento de passo α_k seja menor que

$$2\frac{F(x^k) - F(y)}{\beta m^2 c^2}.$$

Com efeito, nesse caso

$$\alpha_k^2 < 2\alpha_k \frac{F(x^k) - F(y)}{\beta m^2 c^2} \Leftrightarrow -2\alpha_k \left(F(x^k) - F(y)\right) + \alpha_k^2 \beta m^2 c^2 < 0,$$

e portanto, $\|x^{k+m} - y\|^2 < \|x^k - y\|^2$. Em particular, assumindo que o conjunto solução X^* do problema (4-4) seja não vazio, sendo $x^* \in X^*$ então para todo $\varepsilon > 0$ dado, se

$$F(x^k) \leq F(x^*) + \frac{\alpha_k \beta m^2 c^2}{2} + \varepsilon,$$

onde $\{x^k\}$ é uma sequência gerada por um dos algoritmos (4-17), (4-18) ou (4-19) com ordem cíclica, então o iterando x^k está no conjunto de nível de F e poderá estar próximo à solução (desde que α_k seja suficientemente pequeno). Por outro lado, se

$$F(x^k) > F(x^*) + \frac{\alpha_k \beta m^2 c^2}{2} + \varepsilon, \text{ ou seja, } 2\left(F(x^k) - F(x^*)\right) - \alpha_k \beta m^2 c^2 > 2\varepsilon,$$

então $\|x^{k+m} - x^*\|^2 < \|x^k - x^*\|^2 - 2\alpha_k \varepsilon$, i.e., ao final do cíclo, o quadrado da distância do iterando x^{k+m} ao conjunto solução decresce por pelo menos $2\alpha_k \varepsilon$.

Podemos usar o Lema 4.4 para provar resultados de convergência. Iniciamos analisando o comprimento de passo constante ($\alpha_k = \alpha$ para todo k), onde a convergência pode ser estabelecida à uma vizinhança da solução, caso exista. O próximo resultado é semelhante ao Teorema 3.4 do Capítulo 3.

Teorema 4.5 *Seja $\{x^k\}$ uma sequência gerada por um dos algoritmos (4-17), (4-18) ou (4-19), com ordem cíclica de seleção das componentes, e suponha que o comprimento de passo $\alpha_k = \alpha$ para todo k, onde α é uma constante positiva. Temos o seguinte:*

(a) Se $F^* = -\infty$, então

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

(b) Se $F^* > -\infty$, então

$$\liminf_{k \to \infty} F(x^k) \leq F^* + \frac{\alpha \beta m^2 c^2}{2},$$

onde β, e c são as constantes do Lema 4.4.

Prova. Da mesma forma como fizemos na demonstração do Teorema 3.4, provaremos (a) e (b) simultaneamente. Suponha que o resultado não seja verdadeiro, ou seja, que

$$\liminf_{k \to \infty} F(x^{mk}) > F^* + \frac{\alpha \beta m^2 c^2}{2}.$$

Então (em qualquer caso $F^* = -\infty$ ou $F^* > -\infty$), o limite inferior acima é finito, logo existe $\varepsilon > 0$ tal que

$$\liminf_{k \to \infty} F(x^{mk}) - 2\varepsilon > F^* + \frac{\alpha \beta m^2 c^2}{2}.$$

Como F é contínua, pois é a soma das funções convexas, existe também $\hat{y} \in X$ tal que

$$\liminf_{k\to\infty} F(x^{mk}) - 2\varepsilon \geq F(\hat{y}) + \frac{\alpha\beta m^2 c^2}{2}. \tag{4-33}$$

Pela definição de limite inferior, podemos encontrar um índice k_0 suficientemente grande tal que, para todo $k \geq k_0$

$$F(x^{mk}) \geq \liminf_{k\to\infty} F(x^{mk}) - \varepsilon. \tag{4-34}$$

Somando ambos os termos das desigualdades (4-33) e (4-34) temos, para todo $k \geq k_0$

$$F(x^{mk}) - F(\hat{y}) \geq \frac{\alpha\beta m^2 c^2}{2} + \varepsilon. \tag{4-35}$$

Aplicando o Lema 4.4 para o caso $y = \hat{y}$, $\alpha_k = \alpha$, e combinado com a relação (4-35), obtemos para todo $k \geq k_0$

$$\begin{aligned} \|x^{(k+1)m} - \hat{y}\|^2 &\leq \|x^{km} - \hat{y}\|^2 - 2\alpha\left(F(x^{km}) - F(\hat{y})\right) + \beta\alpha^2 m^2 c^2 \\ &\leq \|x^{km} - \hat{y}\|^2 - 2\alpha\varepsilon. \end{aligned}$$

Então, esta última relação implica que para todo $k \geq k_0$ tem-se

$$\|x^{(k+1)m} - \hat{y}\|^2 \leq \|x^{(k-1)m} - \hat{y}\|^2 - 4\alpha\varepsilon \leq \cdots \leq \|x^{k_0} - \hat{y}\|^2 - 2(k+1-k_0)\alpha\varepsilon,$$

que não é verdadeiro quando $k \to \infty$, uma contradição. $\square$

A seguir mostraremos um resultado que relaciona quantidade de iterações e valor mínimo funcional.

Teorema 4.6 *Suponha que $X^* \neq \emptyset$, e seja $\{x^k\}$ uma sequência gerada por um dos algoritmos (4-17), (4-18) ou (4-19), com ordem cíclica de seleção das componentes. Então para todo $\varepsilon > 0$ dado, temos*

$$\min_{0\leq k\leq \bar{k}} F(x^k) \leq F^* + \frac{\alpha\beta m^2 c^2 + \varepsilon}{2} \tag{4-36}$$

onde $\bar{k}$ é o índice dado por

$$\bar{k} = m\frac{\mathrm{dist}^2\left(x^0, X^*\right)}{\alpha\varepsilon} \tag{4-37}$$

Prova. Suponha por contradição que a equação (4-36) não seja verdadeira, assim para todo k com $0 \le km \le \bar{k}$, temos

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

Seja $x^* \in X^*$, usando a relação acima no Lema 4.4 para o caso $y = x^*$ e $\alpha_k = \alpha$, temos para todo k com $0 \le km \le \bar{k}$

$$\begin{aligned} \|x^{(k+1)m} - x^*\|^2 &\le \|x^{km} - x^*\|^2 - 2\alpha\left(F(x^{km}) - F^*\right) + \beta\alpha^2 m^2 c^2 \\ &\le \|x^{km} - x^*\|^2 - \alpha\varepsilon. \end{aligned}$$

Tomando o mínimo sobre todos os $x^* \in X^*$ obtemos

$$\mathrm{dist}^2\left(x^{(k+1)m}, X^*\right) \le \mathrm{dist}^2\left(x^{km}, X^*\right) - \alpha\varepsilon.$$

Como a desigualdade anterior vale para todo $k = 0, 1, \ldots, \frac{\bar{k}}{m}$, podemos concluir que

$$\mathrm{dist}^2\left(x^{\bar{k}+m}, X^*\right) \le \mathrm{dist}^2\left(x^0, X^*\right) - \left(\frac{\bar{k}}{m} + 1\right)\alpha\varepsilon,$$

fazendo algumas manipulações algébricas, esta relação é equivalente a

$$\bar{k} \le m\frac{\mathrm{dist}^2\left(x^0, X^*\right)}{\alpha\varepsilon} - m$$

que contradiz a definição de $\bar{k}$ em (4-37). □

Quando o comprimento de passo α_k tende a zero, é possível obter resultados de convergência. De fato, para o comprimento de passo constante, vimos no Teorema 4.5, que o valor funcional da sequência $\{x^k\}$ poderá estar próximo ao valor ótimo (dependendo de $\alpha\beta m^2 c^2$). Assim, se o comprimento de passo $\alpha_k \to 0$, essa distância deverá ser arbitrariamente menor. Contudo, para a convergência, o passo α_k não pode ser reduzido muito rápido, ou seja, a série dos termos α_k deve ser divergente.

Teorema 4.7 *Seja $\{x^k\}$ uma sequência gerada por um dos algoritmos* (4-17)*,* (4-18) *ou* (4-19)*, com ordem cíclica de seleção das componentes. Suponha que o comprimento de passo α_k satisfaz*

$$\lim_{k\to\infty} \alpha_k = 0, \qquad \sum_{k=0}^{\infty} \alpha_k = \infty.$$

Então

$$\liminf_{k\to\infty} F(x^k) = F^*.$$

Além disso, se X^* é não vazio e

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

então $\{x^k\}$ converge para alguma solução $x^* \in X^*$.

Prova. Para provar a primeira parte, é suficiente mostrar que $\liminf_{k\to\infty} F(x^{km}) = F^*$. Suponha por contradição que $\liminf_{k\to\infty} F(x^{km}) > F^*$, então existe $\varepsilon > 0$ tal que

$$\liminf_{k\to\infty} F(x^{km}) - 2\varepsilon > F^*.$$

Como F é contínua, pois é a soma de funções convexas, segue que existe $\hat{y} \in X$ tal que

$$\liminf_{k\to\infty} F(x^{km}) - 2\varepsilon \geq F(\hat{y}). \tag{4-38}$$

Existe também, pela definição de limite inferior, um k_0 suficientemente grande, tal que para todo $k \geq k_0$

$$F(x^{km}) \geq \liminf_{k\to\infty} F(x^{km}) - \varepsilon. \tag{4-39}$$

Assim, combinando as relações (4-38) e (4-39) segue que, para todo $k \geq k_0$

$$F(x^{km}) - F(\hat{y}) > \varepsilon.$$

Usando esta relação no Lema 4.4 com $y = \hat{y}$ obtemos para todo $k \geq k_0$

$$\begin{aligned} \|x^{(k+1)m} - \hat{y}\|^2 &\leq \|x^{km} - \hat{y}\|^2 - 2\alpha_{km}\left(F(x^{km}) - F(\hat{y})\right) + \beta\alpha_{km}^2 m^2 c^2 \\ &\leq \|x^{km} - \hat{y}\|^2 - 2\alpha_{km}\varepsilon + \beta\alpha_{km}^2 m^2 c^2. \end{aligned}$$

Como $\alpha_k \to 0$ existe $k_1 \geq k_0$ tal que para todo $k \geq k_1$, $\beta\alpha_k m^2 c^2 \leq \varepsilon \Rightarrow \beta\alpha_k^2 m^2 c^2 \leq \alpha_k \varepsilon$. Logo, podemos concluir que para todo $k \geq k_1$

$$\|x^{(k+1)m} - \hat{y}\|^2 \leq \|x^{km} - \hat{y}\|^2 - \alpha_{km}\varepsilon \leq \cdots \leq \|x^{k_1 m} - \hat{y}\|^2 - \varepsilon \sum_{l=k_1}^{k} \alpha_{lm},$$

mas isto é uma contradição quando $k \to \infty$, pois por hipótese $\sum_{k=0}^{\infty} \alpha_k = \infty$. Portanto, temos que $\liminf_{k\to\infty} F(x^{km}) = F^*$.

Para provar a segunda afirmação, seja $x^* \in X^*$, claro que $F(x^k) \geq F(x^*)$ para todo $k \geq 0$. Logo, usando o Lema 4.4 com $y = x^*$ temos para todo $k \geq 0$

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

Como esta desigualdade vale para todo $k \geq 0$ e todo $x^* \in X^*$, e sendo ainda

$$\beta m^2 c^2 \sum_{k=0}^{\infty} \alpha_{km}^2 \leq \beta m^2 c^2 \sum_{k=0}^{\infty} \alpha_k^2 < \infty,$$

vemos pela Definição 2.5 que a sequência $\{x^{km}\}$ é quasi-Fejér convergente, segue da Proposição 2.6 que $\{x^{km}\}$ é limitada, portanto possui um ponto de acumulação $\bar{x}$. Seja $\{x^{k_j m}\}$ a subsequência de $\{x^{km}\}$ tal que $\lim_{j\to\infty} x^{k_j m} = \bar{x}$. Pela primeira parte da demonstração, podemos concluir que $\bar{x} \in X^*$. Novamente pela Proposição 2.6 temos que $\lim_{k\to\infty} x^{km} = \bar{x}$. Para mostrar que toda a sequência $\{x^k\}$ também converge para $\bar{x}$, na demonstração do Lema 4.4, em especial, nas inequações (4-26) e (4-32) vimos que, para qualquer iteração (4-17), (4-18) ou (4-19), $\|x^{k+1} - x^k\| \leq \alpha_k c$. Agora, como $\alpha_k \to 0$ e $x^{km} \to \bar{x}$, podemos concluir que $x^k \to \bar{x}$. $\square$

CAPÍTULO 5

Considerações finais

Neste trabalho descrevemos, sob vários contextos, o método incremental objetivando minimizar uma função definida como soma de funções convexas. Apresentamos o método subgradiente incremental, estudado no artigo [14] por Nedić e Bertsekas, e analisamos suas propriedades de convergência sob várias regras para o comprimento de passo. Consideramos também o método do ponto proximal no modelo incremental, obtido em [2], de forma mais geral consideramos um método do tipo *Forward-Backward* incremental e algumas variações, ou seja, em cada iteração combinamos um método proximal com subgradiente, de forma cíclica, e em cada componente. Estudamos suas propriedades de convergência, em particular estendendo alguns resultados obtidos para o método do subgradiente incremental.

Algumas idéias deste trabalho merecem uma investigação mais aprofundadas. Como por exemplo, poderiamos considerar um valor aproximado para o cálculo do subgradiente, ou seja, um ε-subgradiente. Também, poderíamos variar o tamanho do comprimento do passo dentro do ciclo, considerando diferentes regras para a escolha do comprimento do passo. Outro interesse seria analisar o método incremental subgradiente-proximal utilizando os demais comprimentos de passo estudados no Capítulo 3, como o comprimento de passo dinâmico para f^* desconhecido.

Referências Bibliográficas

[1] Bertsekas, D. P. *A new class of incremental gradient methods for least squares problems*, SIAM J. Optimization, Vol 7, November 1997, pp. 913-926.

[2] Bertsekas, D. P. *Incremental Gradient, Subgradient, and Proximal Methods for Convex Optimization: A Survey*, August 2010 (revised December 2010).

[3] Bertsekas, D. P. *Incremental Proximal Methods for Large Scale Convex Optimization*, Math. Program., Ser. B (2011), pp. 163–195

[4] Bertsekas, D. P. *Nonlinear Programming*, 2nd edition, Athena Scientific, Belmont, Massachusetts 1999.

[5] Correa, R.; Lemaréchal C. *Convergence of some algorithms for convex minimiazation*, Mathematical Programming, 62 (1993), pp. 261-275.

[6] Drezner, Z.; Hamacher, H. W. *Facility Location Applications and Theory*, Springer, New Iork, 2002.

[7] Iusem, A. N.; Svaiter, B. F. *A Proximal Regularization of the Steepest Descent Method*, vol. 29, nº 2, 1995, pp. 123-130.

[8] Iusem, A. N. *Proximal Point Methods in Optimization*, 20º Colóquio Brasileiro de Matemática, IMPA Maio 1995

[9] Izmailov, A.; Solodov, M. *Otimização-Volume 1-Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade*, IMPA, Rio de Janeiro, 2009.

[10] Izmailov, A.; Solodov, M. *Otimização-Volume 2-Métodos computacionais*, IMPA, Rio de Janeiro, 2007.

[11] Kibardin, V. M. *Decomposition into functions in the minimization problem*, Automation and Remote Control, 40, (1980), pp. 1311-1323.

[12] Lima, E.L. *Curso de análise - Volume 1*, IMPA, Rio de Janeiro, 13ª edição, 2011.

[13] Lima, E.L. *Curso de análise - Volume 2*, IMPA, Rio de Janeiro, 11ª edição, 2011.

[14] Nedić, A.; Bertsekas, D. P. *Incremental Subgradient Methods for Non-differentiable Optimization*, SIAM J. on Optimization, Vol 12, 2001, pp. 109-138.

[15] Polyak, B. T. *Mininization of unsmooth functionals*, Zhurnal Vychisdi-tel´noi Matematiki i Matematicheskoi Fiziki 9, pp. 509-521.

[16] Solodov, M. V.; Zavriev, S. K. *Error stability properties of generalized gradient-type algorithms*, J. of Optimization Theroy and Applications, Vol. 98, No. 3, September 1998, pp. 663-680.