

ANA FRIEDLANDER

**ELEMENTOS DE  
PROGRAMAÇÃO  
NÃO-LINEAR**



# Sumário

1	O PROBLEMA DE PROGRAMAÇÃO NÃO-LINEAR	7
2	CONDIÇÕES DE OTIMALIDADE PARA MINIMIZAÇÃO SEM RESTRIÇÕES	11
3	CONVEXIDADE	17
4	MODELO DE ALGORITMO COM BUSCAS DIRECIONAIS	21
5	ORDEM DE CONVERGÊNCIA	31
6	MÉTODOS CLÁSSICOS DE DESCIDA	33
7	MINIMIZAÇÃO COM RESTRIÇÕES LINEARES DE IGUALDADE	47
8	ALGORITMOS PARA RESTRIÇÕES LINEARES DE IGUALDADE	55
9	MINIMIZAÇÃO COM RESTRIÇÕES LINEARES DE DESIGUALDADE	63
10	MÉTODO DE RESTRIÇÕES ATIVAS	77
11	MINIMIZAÇÃO COM RESTRIÇÕES LINEARES DE IGUALDADE E DESIGUALDADE	81
12	MINIMIZAÇÃO COM RESTRIÇÕES NÃO-LINEARES DE IGUALDADE	85
13	MINIMIZAÇÃO COM RESTRIÇÕES NÃO-LINEARES DE IGUALDADE E DESIGUALDADE	95

---

14	ALGORITMOS PARA RESTRIÇÕES NÃO-LINEARES	105
A	NOTAÇÕES	113
	Referências Bibliográficas	115

# Prefácio

Este livro é resultado da experiência de vários anos ministrando um curso de graduação sobre programação não-linear na Unicamp, para alunos de Matemática, Matemática Aplicada e Computação. Não reflete apenas a vivência da autora, mas também a de outros colegas, especialmente Lúcio Tunes dos Santos e José Mario Martínez.

Nossa convicção é que a aprendizagem é o fruto exclusivo do trabalho ativo do aluno, cabendo ao instrutor as tarefas de propor problemas desafiantes, orientar o estudante na sua resolução, e fornecer os elementos teóricos *essenciais* para possibilitar a atividade deste. Nosso curso de Programação não-linear foi estruturado com essa filosofia. Na sala de aula, o professor ocupa, como expositor, uma pequena parte do tempo que, na sua maioria, está dedicado a que os próprios alunos resolvam problemas, e consultem suas dúvidas com o instrutor. Com este esquema, o instrutor deve-se colocar freqüentemente no contexto dos argumentos dos estudantes, e não apenas expor seus conhecimentos usando o próprio marco conceitual.

O papel do livro-texto nesta metodologia é condensar a teoria necessária para a resolução dos problemas. Fundamentalmente, o livro é para ser *lido* pelos estudantes, mais do que *exposto* pelo instrutor. Imaginamos que seja lido da maneira, às vezes ordenada, às vezes caótica, de quem procura elementos para resolver um problema pelo qual está apaixonado.

Do ponto de vista de conteúdo, encaramos com realismo o fato de que os conhecimentos e a capacidade operativa em Álgebra Linear e Cálculo de nossos estudantes são, geralmente, pobres. Em conseqüência, o texto se desvia às vezes da Programação não-linear, e parece um texto de aplicações de Álgebra Linear e Cálculo. Esse desvio é proposital. Parece-nos que o tempo usado neste curso estará muito bem-justificado se dele resultar um conhecimento mais profundo e dinâmico daquelas duas matérias básicas, cujo poder multiplicativo, em termos de aproveitamento em outras áreas da matemática aplicada, é, obviamente, enorme.

A lista de exercícios é essencial neste curso. Ela foi elaborada ao longo destes anos não apenas por mim, mas também por Lúcio e Martínez, usando problemas clássicos da literatura (Mc Cormick, Luenberger, Fletcher etc.) e inventando novos exercícios para a estrutura peculiar do nosso ensino. Ao Lúcio coube a tarefa de colecionar as diferentes listas que circularam nos últimos anos, juntando problemas de provas e, em geral, organizando racionalmente o material. Esses colegas merecem todo meu agradecimento, assim como Sandra Santos, que fez os desenhos, e as várias turmas de alunos que, ao longo destes anos, enriqueceram nossa proposta.

# Capítulo 1

## O PROBLEMA DE PROGRAMAÇÃO NÃO-LINEAR

Neste livro nos ocuparemos de problemas da forma

$$\begin{aligned} &\text{Minimizar } f(x) \\ &\text{sujeita a } x \in S, \end{aligned} \tag{1.1}$$

onde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  e  $S \subset \mathbb{R}^n$ .  $S$  é chamado *conjunto factível* e (1.1) é a forma genérica dos problemas de *programação não-linear* ou *otimização*.

Consideramos dois tipos de *soluções* deste problema:

### Definição 1.1

Um ponto  $x^* \in S$  é um *minimizador local* de  $f$  em  $S$  se e somente se existe  $\varepsilon > 0$  tal que  $f(x) \geq f(x^*)$  para todo  $x \in S$  tal que  $\|x - x^*\| < \varepsilon$ . Se  $f(x) > f(x^*)$  para todo  $x \in S$  tal que  $x \neq x^*$  e  $\|x - x^*\| < \varepsilon$ , diremos que se trata de um *minimizador local estrito* em  $S$ .

### Definição 1.2

Um ponto  $x^* \in S$  é um *minimizador global* de  $f$  em  $S$  se e somente se  $f(x) \geq f(x^*)$  para todo  $x \in S$ . Se  $f(x) > f(x^*)$  para todo  $x \in S$  tal que  $x \neq x^*$ , diremos que se trata de um *minimizador global estrito* em  $S$ .

Em forma análoga, definimos maximizadores locais e globais, o que fica como exercício para o leitor. Observemos que “Maximizar  $f$ ” é equivalente a “Minimizar  $-f$ ”, razão pela qual podemos, sem perda de generalidade, falar apenas de “Minimização” ao longo do texto.

O seguinte é um resultado fundamental relacionado com o problema de otimização.

**Teorema 1.1 (Bolzano-Weierstrass)**

Uma função real contínua  $f$ , definida em um conjunto fechado e limitado  $S \subset \mathbb{R}^n$ , admite um minimizador global em  $S$ .

**Prova:** Ver, por exemplo Rey Pastor et al. [14].■

**Exercícios (Revisão de Álgebra Linear e Cálculo)**

1.1 Sejam  $A \in \mathbb{R}^{n \times n}$  e  $x \in \mathbb{R}^n$ . Quais das seguintes afirmações são verdadeiras? Prove ou dê um contra-exemplo:

- (a) Existe  $x^* \neq 0$  tal que  $Ax^* = 0$  se  $\det(A) = 0$ ;
- (b) Existe  $x^* \neq 0$  tal que  $Ax^* = 0$  somente se  $\det(A) = 0$ ;
- (c) Existe  $x^* \neq 0$  tal que  $Ax^* = 0$  se e somente se  $\det(A) \neq 0$ .

1.2 Seja  $A \in \mathbb{R}^{m \times n}$ ,  $m \geq n$  e posto  $A = n$ . Prove que  $A^t A$  é não-singular.

1.3 Seja  $A \in \mathbb{R}^{m \times n}$ ,  $m \leq n$  e posto  $A = k$ . Definimos os subespaços:

Núcleo de  $A$ :  $Nu(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$ ;

Imagem de  $A$ :  $Im(A) = \{y \in \mathbb{R}^m \mid \exists x \in \mathbb{R}^n \mid y = Ax\}$ ;

Prove que: (a)  $Nu(A) \perp Im(A^t)$ ; (b)  $\dim(Nu(A)) = n - k$ ; (c)  $\mathbb{R}^n = Nu(A) \oplus Im(A^t)$ .

1.4 Considere as equações

$$\sum_{j=1}^n a_{ij}x_j = b_i, \quad i = 1, \dots, n-1,$$

ou equivalentemente,  $Ax = b$  com  $A \in \mathbb{R}^{(n-1) \times n}$ ,  $b \in \mathbb{R}^{n-1}$  e  $x \in \mathbb{R}^n$ , correspondendo a  $n-1$  hiperplanos “linearmente independentes”. A intersecção desses hiperplanos determina uma reta em  $\mathbb{R}^n$ . Podemos representar essa reta na forma

$$y = x + \lambda d$$

com  $\lambda \in \mathbb{R}$  e  $x, d \in \mathbb{R}^n$ . Discuta como escolher  $x$  e  $d$ .

1.5 Encontre os autovalores e autovetores da matriz  $A = uu^t$ , onde  $u \in \mathbb{R}^n$ .

1.6 Prove que os autovetores de uma matriz associados a autovalores distintos são linearmente independentes e que se a matriz é simétrica eles são ortogonais.



1.7 Prove que os autovalores de uma matriz simétrica são positivos se e somente se a matriz é definida positiva.

1.8 Prove que se  $\lambda$  é um autovalor de uma matriz  $A$  não-singular, então  $1/\lambda$  é um autovalor de  $A^{-1}$ .

1.9 Prove que  $A \in \mathbb{R}^{n \times n}$  é singular se e somente se  $0$  é um autovalor.

1.10 Suponha que  $\lim_{k \rightarrow \infty} x^k = \alpha$ . Prove que se  $\alpha > \beta$ , então existe  $M > 0$  tal que para qualquer  $k \geq M$  se verifica que  $x^k > \beta$ .

1.11 Prove que se  $\lim_{k \rightarrow \infty} x^k = \alpha$  e para todo  $k \geq 0$ ,  $x^k \geq \beta$ , então  $\alpha \geq \beta$ . Trocando o sinal de  $\geq$  por  $>$ , a afirmação continua válida? Prove ou dê um contra-exemplo.

1.12 Se  $\{x^k\}$  é uma seqüência convergente, então essa seqüência é limitada? A recíproca é verdadeira?

1.13 É possível ter uma seqüência convergente tal que  $x^{2k} > 0$  e  $x^{2k+1} < 0$  para todo  $k$ ?

1.14 Prove que as funções abaixo são normas:

(a)  $\|\cdot\|_\infty : \mathbb{R}^n \rightarrow \mathbb{R}, \|x\|_\infty = \text{Máximo } |x_i|;$   
 $1 \leq i \leq n$

(b)  $\|\cdot\|_1 : C(a, b) \rightarrow \mathbb{R}, \|f\|_1 = \int_a^b |f(x)| dx$ . ( $C(a, b)$  é o conjunto das funções contínuas  $[a, b] \rightarrow \mathbb{R}$ .)

1.15 Considere as funções  $f : \mathbb{R}^m \rightarrow \mathbb{R}^p$  e  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  com jacobianos  $J_f \in \mathbb{R}^{p \times m}$  e  $J_g \in \mathbb{R}^{m \times n}$ , respectivamente. Encontre o jacobiano da função composta  $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ , dada por  $h(x) = f(g(x))$ .

1.16 Calcule o gradiente e o hessiano das funções  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  abaixo:

(a)  $f(x) = a^t x;$

(b)  $f(x) = \frac{1}{2} x^t A x + b^t x + c$ , onde  $A \in \mathbb{R}^{n \times n}$ ,  $b \in \mathbb{R}^n$ ,  $c \in \mathbb{R};$

(c)  $f(x) = g^t(x)g(x) = \|g(x)\|_2^2$ , onde  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ .

1.17 Sejam  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ . Para  $x \in \mathbb{R}^n$ , definimos  $q(x) = f(Ax + b)$  com  $f : \mathbb{R}^m \rightarrow \mathbb{R}$ . Calcule o gradiente e o hessiano da função  $q$ .

1.18 Desenhe as curvas de nível das seguintes quadráticas:

- (a)  $x^2 - y^2 - x + y - 1$ ;
- (b)  $x^2 + y^2 + 2xy$ ;
- (c)  $x^2 + y^2 - xy$ ;
- (d)  $xy$ .

1.19 Escreva a expansão em série de Taylor em torno do ponto  $x = 0$  para as seguintes funções:

- (a)  $\cos(x)$ ;
- (b)  $\ln(x + 1)$ ;
- (c)  $\exp(x)$ .

1.20 Discuta a geometria das curvas de nível de uma função quadrática

$f(x) = \frac{1}{2}x^t Ax + b^t x + c$ , onde  $A \in \mathbb{R}^{2 \times 2}$  simétrica,  $b \in \mathbb{R}^2$  e  $c \in \mathbb{R}$ , nos seguintes casos:

- (a)  $A > 0$ ;
- (b)  $A \geq 0$  e existe  $x$  tal que  $Ax + b = 0$ ;
- (c)  $A \geq 0$  e não existe  $x$  tal que  $Ax + b = 0$ ;
- (d)  $A$  indefinida e não-singular.

1.21 Considere a função  $f(x, y) = x \cos y + y \sin x$ . Determine a aproximação linear de  $f$  em torno do ponto  $(0, 0)$ . Determine um limitante para o erro na região  $[-1, 1] \times [-1, 1]$ .

## Capítulo 2

# CONDIÇÕES DE OTIMALIDADE PARA MINIMIZAÇÃO SEM RESTRIÇÕES

Analisaremos inicialmente o caso em que o conjunto factível é  $\mathbb{R}^n$ . Neste caso, (1.1) é chamado *problema de minimização irrestrita*.

### 2.1 CONDIÇÕES DE OTIMALIDADE

Supomos conhecidos os seguintes resultados para funções de uma variável.

**R1** - Seja  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f \in C^1$ . Se  $x^*$  é um minimizador local de  $f$  em  $\mathbb{R}$ , então  $f'(x^*) = 0$ .

**R2** - Seja  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f \in C^2$ . Se  $x^*$  é um minimizador local de  $f$  em  $\mathbb{R}$ , então

- (i)  $f'(x^*) = 0$ ;
- (ii)  $f''(x^*) \geq 0$ .

#### Proposição 2.1 (Condições necessárias de primeira ordem)

Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^1$ . Se  $x^*$  é um minimizador local de  $f$  em  $\mathbb{R}^n$ , então  $\nabla f(x^*) = 0$ .

**Prova:** Fixamos  $d \in \mathbb{R}^n$  arbitrário e consideramos a função  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  definida por:

$$\phi(\lambda) = f(x^* + \lambda d).$$

Como  $x^*$  é um minimizador local de  $f$ , resulta que  $\lambda \equiv 0$  é um minimizador local de  $\phi$ . Neste caso, por R1, concluímos que  $\phi'(0) = 0$ .

Utilizando a regra da cadeia obtemos  $\phi'(\lambda) = \nabla^t f(x^* + \lambda d)d$ .

Substituindo para  $\lambda = 0$ , resulta  $0 = \phi'(0) = \nabla^t f(x^*)d$ .

Como  $d \in \mathbb{R}^n$  é arbitrário, esta igualdade significa que  $\nabla f(x^*)$  é um vetor

ortogonal a todos os vetores do espaço, portanto  $\nabla f(x^*) = 0$ .■

**Proposição 2.2 (Condições necessárias de segunda ordem)**

Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^2$ . Se  $x^*$  é um minimizador local de  $f$  em  $\mathbb{R}^n$ , então

- (i)  $\nabla f(x^*) = 0$ ;
- (ii)  $\nabla^2 f(x^*)$  é semidefinida positiva.

**Prova:** A primeira parte da tese se segue da Proposição 2.1. Para provar a segunda parte, consideremos  $\phi(\lambda)$ , como na Proposição 2.1. R2 implica que  $\phi''(0) \geq 0$ . Usando a regra da cadeia temos  $\phi''(\lambda) = d^t \nabla^2 f(x^* + \lambda d)d$ , logo,

$$\phi''(0) = d^t \nabla^2 f(x^*)d \geq 0.$$

Como  $d \in \mathbb{R}^n$  é arbitrário obtemos que  $\nabla^2 f(x^*)$  é semidefinida positiva.■

**Proposição 2.3 (Condições suficientes de segunda ordem)**

Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^2$ . Se  $x^* \in \mathbb{R}^n$ ,  $\nabla f(x^*) = 0$ , e  $\nabla^2 f(x^*) > 0$ , então  $x^*$  é um minimizador local estrito de  $f$  em  $\mathbb{R}^n$ .

**Prova:** Seja  $B = \{h \in \mathbb{R}^n \mid \|h\| = 1\}$ . Consideremos a função  $\Gamma : B \rightarrow \mathbb{R}$  dada por

$$\Gamma(h) = h^t \nabla^2 f(x^*)h.$$

$\Gamma$  é uma função contínua e  $B$  é um conjunto fechado e limitado, portanto  $\Gamma$  atinge um valor máximo e um valor mínimo em  $B$ . Chamemos  $a$  ao valor mínimo, então

$$\Gamma(h) \geq a > 0 \text{ para todo } h \in B.$$

Agora, consideremos  $d \in \mathbb{R}^n$ , arbitrário não-nulo. Como  $d / \|d\| \in B$ , temos que

$$d^t \nabla^2 f(x^*)d \geq a \|d\|^2. \quad (2.1)$$

Desenvolvendo  $f$  em série de Taylor em torno de  $x^*$ , temos

$$f(x^* + d) - f(x^*) = \nabla^t f(x^*)d + \frac{1}{2} d^t \nabla^2 f(x^*)d + o(\|d\|^2). \quad (2.2)$$

Desde que, por hipótese,  $\nabla f(x^*) = 0$ , (2.2) implica que

$$f(x^* + d) - f(x^*) \geq \frac{a}{2} \|d\|^2 + o(\|d\|^2).$$

Então, para todo  $d$  tal que  $\|d\|$  é suficientemente pequeno, o primeiro termo do membro direito da desigualdade define o sinal deste lado. Mas

$$\frac{a}{2} \|d\|^2 > 0.$$

Portanto, para  $\|d\|$  suficientemente pequeno não-nulo (digamos  $0 < \|d\| < \varepsilon$ )

$$f(x^* + d) - f(x^*) > 0,$$

ou seja,  $f(x^* + d) > f(x^*)$ . Então, para todo  $x \in B(x^*, \varepsilon)$ ,  $x \neq x^*$ , temos que  $f(x) > f(x^*)$ . Logo,  $x^*$  é um minimizador local estrito de  $f$ . ■

**Observação:** A argumentação utilizada na prova da Proposição 2.3 é essencialmente diferente e mais complicada que a usada nas provas das Proposições 2.1 e 2.2. O Exercício 2.6 mostra por que o argumento mais simples não é válido para provar a Proposição 2.3.

### Exercícios

2.1 Sejam  $g : \mathbb{R} \rightarrow \mathbb{R}$  uma função estritamente crescente e  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . Prove que minimizar  $f(x)$  é equivalente a minimizar  $g(f(x))$ .

2.2 Resolva o problema de minimizar  $\|Ax - b\|$ , onde  $A \in \mathbb{R}^{m \times n}$  e  $b \in \mathbb{R}^m$ . Considere todos os casos possíveis e interprete geometricamente.

2.3 Considere os números reais  $a_1 \leq a_2 \leq \dots \leq a_n$ . Encontre a solução dos seguintes problemas:

(a) Minimizar  $\sum_{i=1}^n |x - a_i|$ ;

(b) Minimizar  $\text{Máximo} \{|x - a_i|, i = 1, \dots, n\}$ ;

(c) Minimizar  $\sum_{i=1}^n |x - a_i|^2$ ;

(d) Maximizar  $\prod_{i=1}^n |x - a_i|$ .

2.4 Obtenha expressões para as derivadas primeiras e segundas da função de Rosenbrock  $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$ . Verifique que  $\tilde{x} = (1, 1)^t$  é um minimizador local. Prove que  $\nabla^2 f(\tilde{x})$  é singular se e somente se  $x_2 - x_1^2 = 0.005$ .

2.5 Encontre os pontos estacionários de

$$f(x) = 2x_1^3 - 3x_1^2 - 6x_1x_2(x_1 - x_2 - 1).$$

Quais desses pontos são minimizadores ou maximizadores, locais ou globais?

## 14Capítulo 2. CONDIÇÕES DE OTIMALIDADE PARA MINIMIZAÇÃO SEM RESTRIÇÕES

2.6 Seja  $f(x) = (x_1 - x_2^2)(x_1 - \frac{1}{2}x_2^2)$ . Verifique que  $\bar{x} = (0, 0)^t$  é um minimizador local de  $\phi(\lambda) \equiv f(\bar{x} + \lambda d)$  para todo  $d \in \mathbb{R}^2$ , mas  $\bar{x}$  não é minimizador local de  $f$ .

2.7 Prove que a função  $f(x) = (x_2 - x_1^2)^2 + x_1^5$  tem um único ponto estacionário que não é minimizador nem maximizador local.

2.8 Encontre funções  $f : \mathbb{R}^n \rightarrow \mathbb{R}, n \geq 2$ , tais que  $\nabla f(\tilde{x}) = 0$  e  $\tilde{x}$  é:

- (a) maximizador local, não global;
- (b) ponto de sela;
- (c) minimizador global.

2.9 Para aproximar uma função  $g$  no intervalo  $[0, 1]$  por um polinômio de grau  $\leq n$ , minimizamos a função critério:

$$f(a) = \int_0^1 [g(x) - p(x)]^2 dx,$$

onde  $p(x) = a_0 + a_1x + \dots + a_nx^n$ . Encontre as equações a serem satisfeitas pelos coeficientes ótimos.

2.10 Considere o problema irrestrito

$$\text{Minimizar } f(x) = x_1^2 - x_1x_2 + 2x_2^2 - 2x_1 + \exp[x_1 + x_2]$$

(a) Escreva as condições necessárias de primeira ordem. São suficientes? Por quê?

(b) O ponto  $\bar{x} = (0, 0)^t$  é ótimo?

(c) Ache uma direção  $d \in \mathbb{R}^2$  tal que  $\nabla^t f(\bar{x})d < 0$ ;

(d) Minimizar a função a partir de  $\bar{x}$  na direção obtida em (c).

2.11 Seja  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  com derivadas contínuas. Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  dada por  $f(x) = \|F(x)\|^2$ . Seja  $\tilde{x}$  minimizador local de  $f$  tal que  $J_F(\tilde{x})$  é não-singular. Prove que  $\tilde{x}$  é solução do sistema  $F(x) = 0$ .

2.12 Considere  $f : \mathbb{R}^2 \rightarrow \mathbb{R}, f(x) = (x_1^3 + x_2)^2 + 2(x_2 - x_1 - 4)^4$ . Dado um ponto  $x \in \mathbb{R}^2$  e uma direção  $0 \neq d \in \mathbb{R}^2$ , construímos a função

$$g(\lambda) = f(x + \lambda d)$$

(a) Obtenha uma expressão explícita para  $g(\lambda)$ ;

(b) Para  $x = (0, 0)^t$  e  $d = (1, 1)^t$  encontre o minimizador de  $g$ .

2.13 Considere a função  $f(x) = (x_1 - 1)^2 x_2$ . Considere os pontos de  $\mathbb{R}^2$  da forma  $\hat{x} = (1, x_2)^t$ .

(a) Analise as condições de otimalidade de primeira e segunda ordem para esses pontos;

(b) O que se pode afirmar sobre  $\hat{x}$  utilizando essas informações?

(c) Use a expressão da função para obter afirmações mais conclusivas sobre as características de  $\hat{x}$ .

2.14 Sejam  $f(x) = \frac{1}{2}x^t Q x - b^t x$ ,  $Q \in \mathbb{R}^{n \times n}$  simétrica definida positiva e  $b \in \mathbb{R}^n$ . Sejam  $x^0, x^1, \dots, x^n \in \mathbb{R}^n$  e definimos  $\delta^j = x^j - x^0$ ,  $\gamma^j = \nabla f(x^j) - \nabla f(x^0)$ ,  $j = 0, 1, \dots, n$ . Prove que se os vetores  $\{\delta^j\}_{j=1}^n$  são linearmente independentes, então

$$\tilde{x} = x^n - [\delta^1 \dots \delta^n] \cdot [\gamma^1 \dots \gamma^n]^{-1} \cdot \nabla f(x^n)$$

é minimizador global de  $f$ .

2.15 Definimos a *norma de Frobenius* de uma matriz  $A \in \mathbb{R}^{m \times n}$  como

$$\|A\|_F = \left( \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 \right)^{1/2}.$$

Dada uma matriz  $A \in \mathbb{R}^{n \times n}$ , encontre a matriz simétrica mais próxima de  $A$  na norma de Frobenius, isto é, encontre a matriz  $B \in \mathbb{R}^{n \times n}$ , simétrica tal que  $\|A - B\|_F$  é mínima.

2.16 Seja  $f : \mathbb{R} \rightarrow \mathbb{R}$  e suponha que  $f^{(j)}(a) = 0$ ,  $j = 0, \dots, n - 1$  e  $f^{(n)}(a) \neq 0$ . Sobre que condições o ponto  $x = a$  poderá ser um minimizador de  $f$ ? Baseado em sua resposta:  $f(x) = x^{13}$  tem um mínimo em  $x = 0$ ? E  $f(x) = x^{16}$ ?

2.17 Se for possível determine  $a$  e  $b$  de modo que  $f(x) = x^3 + ax^2 + bx$  tenha um máximo local em  $x = 0$  e um mínimo local em  $x = 1$ .





## Capítulo 3

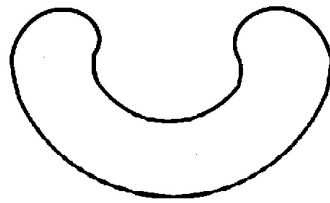
# CONVEXIDADE

As proposições enunciadas no Capítulo 2 são úteis para caracterizar minimizadores locais. Reconhecer se um minimizador local também é global não é fácil, a menos que a função objetivo tenha características especiais. O caso mais simples é o de funções convexas.

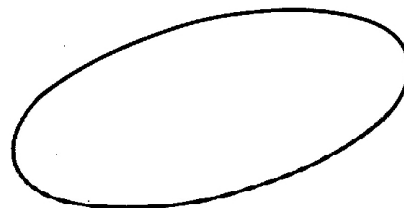
### 3.1 CONCEITOS FUNDAMENTAIS

#### Definição 3.1

Um subconjunto  $S \subset \mathbb{R}^n$  é convexo se e somente se para todo  $x, y \in S$ ,  $\lambda \in [0, 1]$  se verifica que  $\lambda x + (1 - \lambda)y \in S$ . Ver Figura 3.1.



NÃO-CONVEXO



CONVEXO

Figura 3.1

#### Definição 3.2

Uma função  $f$  definida em um convexo  $S$  é convexa se e somente se para todo

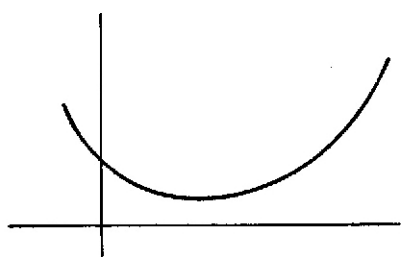
$x, y \in S, \lambda \in [0, 1]$  se verifica que

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

Se para todo  $\lambda \in (0, 1)$  e  $x \neq y$  vale que

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

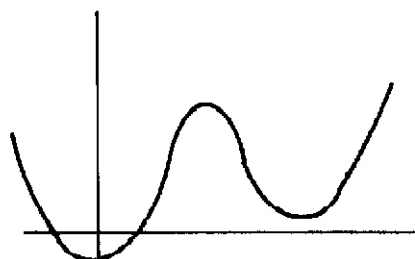
diremos que  $f$  é estritamente convexa. Ver Figura 3.2.



ESTRITAMENTE CONVEXA



CONVEXA



NÃO-CONVEXA

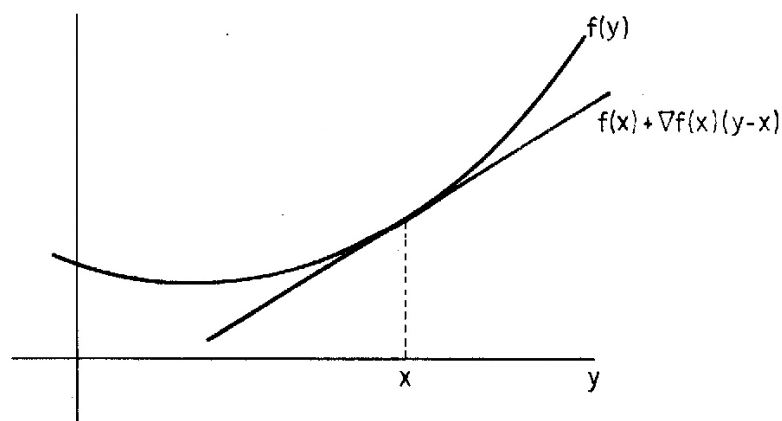
Figura 3.2

### 3.2 FUNÇÕES CONVEXAS DIFERENCIÁVEIS

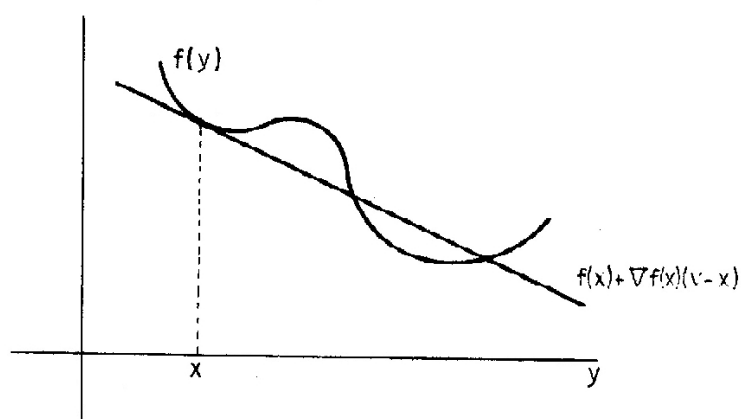
#### Proposição 3.1

Seja  $f \in C^1$ . Então,  $f$  é convexa em  $S$  convexo se e somente se para todo  $x, y \in S$  se verifica

$$f(y) \geq f(x) + \nabla^t f(x)(y - x).$$



FUNÇÃO CONVEXA



FUNÇÃO NÃO-CONVEXA

Figura 3.3 - Ilustração da Proposição 3.1

**Proposição 3.2**

Seja  $f \in C^2$ . Seja  $S \subset \mathbb{R}^n$  convexo tal que  $\overset{\circ}{S}$  não é vazio. Então,  $f$  é convexa se e somente se  $\nabla^2 f(x) \geq 0$  para todo  $x \in S$ .

**Proposição 3.3**

Seja  $f$  uma função convexa definida em  $S$  convexo. Então:

- (i) O conjunto  $\Gamma \subset S$  onde  $f$  toma seu valor mínimo é convexo;
- (ii) Qualquer minimizador local de  $f$  é um minimizador global de  $f$ .

**Proposição 3.4**

Seja  $f \in C^1$  convexa definida em  $S$  convexo. Se existe  $x^* \in S$  tal que para todo  $y \in S$  se verifica que

$$\nabla^t f(x^*)(y - x^*) \geq 0,$$

então  $x^*$  é um minimizador global de  $f$  em  $S$ .

As provas das proposições desta seção podem ser encontradas em Luenberger [11].

**Exercícios**

3.1 Prove que a intersecção de conjuntos convexos é convexa.

3.2 Prove que  $S = \{x \in \mathbb{R}^n \mid \|x\| \leq c, c > 0\}$ , onde  $\|\cdot\|$  é uma norma qualquer em  $\mathbb{R}^n$ , é um conjunto convexo.

3.3 Verifique se as funções abaixo são convexas:

- (a)  $f(x) = \text{máximo} \{g(x), h(x)\}$  onde  $g$  e  $h$  são funções convexas;
- (b)  $t(x) = \sum_{i=1}^n x_i^2$ ;
- (c)  $s(x) = \exp[f(x)]$ ,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ .

3.4 Desenhe as curvas de nível de uma função convexa. Justifique!

3.5 Seja  $S$  um conjunto convexo não vazio em  $\mathbb{R}^n$ . Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  a função definida por

$$f(y) = \text{Mínimo} \{ \|y - x\| \mid x \in S \}.$$

Esta função é convexa. Prove esta afirmação quando

$$S = \{x \in \mathbb{R}^2 \mid ax_1 + bx_2 = c\}.$$

Interprete geometricamente.

## Capítulo 4

# MODELO DE ALGORITMO COM BUSCAS DIRECIONAIS

### 4.1 DIREÇÕES DE DESCIDA

Dado  $x \in \mathbb{R}^n$ , se  $\nabla f(x) \neq 0$ , sabemos, pela Proposição 2.1, que  $x$  não é um minimizador local de  $f$  em  $\mathbb{R}^n$ . Portanto, em toda vizinhança de  $x$  existe  $z \in \mathbb{R}^n$  tal que  $f(z) < f(x)$ .

Interessa-nos caracterizar as direções a partir de  $x$ , nas quais é possível achar um ponto  $z \in \mathbb{R}^n$  que verifique  $f(z) < f(x)$ .

#### Proposição 4.1

Sejam  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^1$ ,  $x \in \mathbb{R}^n$  tal que  $\nabla f(x) \neq 0$ ,  $d \in \mathbb{R}^n$  tal que  $\nabla^t f(x)d < 0$ . Então existe  $\bar{\alpha} > 0$  tal que  $f(x + \alpha d) < f(x)$  para todo  $\alpha \in (0, \bar{\alpha}]$ .

**Prova:** Consideramos a função  $\phi(\alpha) \equiv f(x + \alpha d)$ . Então  $\phi(0) = f(x)$ , e aplicando a regra da cadeia temos  $\phi'(0) = \nabla^t f(x)d$ .

Como  $\phi'(0) = \lim_{\alpha \rightarrow 0} \frac{\phi(\alpha) - \phi(0)}{\alpha}$ , então para  $0 < \alpha < \bar{\alpha}$ , com  $\bar{\alpha}$  suficientemente pequeno, o sinal de  $\phi'(0)$  e o sinal de  $\phi(\alpha) - \phi(0)$  deve ser o mesmo.

Como  $\nabla^t f(x)d < 0$  temos que  $\phi'(0) < 0$  e  $\phi(\alpha) - \phi(0) < 0$  para  $0 < \alpha < \bar{\alpha}$ , portanto  $f(x + \alpha d) < f(x)$ . ■

A Proposição 4.1 diz que, dado  $d \in \mathbb{R}^n$  tal que  $\nabla^t f(x)d < 0$ , certamente podemos encontrar nessa direção pontos onde o valor da função seja estritamente menor que  $f(x)$ .

As direções  $d \in \mathbb{R}^n$ , tais que  $\nabla^t f(x)d < 0$ , são chamadas *direções de descida* a partir de  $x$ . A existência dessas direções sugere um modelo geral de algoritmo

para minimizar uma função sem restrições.

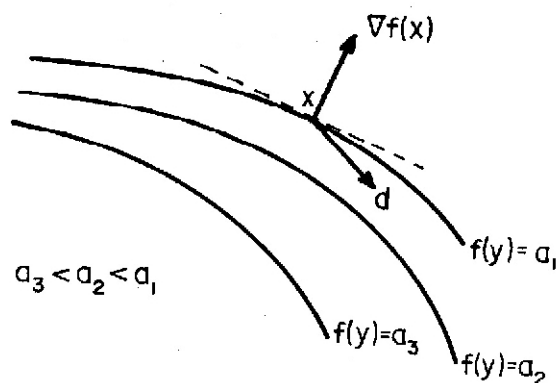


Figura 4.1 - Ilustração da Proposição 4.1

## 4.2 MODELO DE ALGORITMO

Se  $x^*$  é uma solução de

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

e  $x^k$  é uma estimativa de  $x^*$ , tal que  $\nabla f(x^k) \neq 0$ ; os passos para definir uma nova estimativa  $x^{k+1}$  são dados pelo seguinte algoritmo.

### Algoritmo 4.1

**Passo 1:** Escolher  $d_k \in \mathbb{R}^n$  tal que  $\nabla^t f(x^k)d_k < 0$ .

**Passo 2:** (Determinação do tamanho do passo)

Calcular  $\lambda_k > 0$  tal que  $f(x^k + \lambda_k d_k) < f(x^k)$ .

(Este subproblema é chamado de busca linear.)

**Passo 3:** Fazer  $x^{k+1} = x^k + \lambda_k d_k$ .

O processo termina se para algum valor de  $k$ , digamos  $k_0$ , resulta  $\nabla f(x^{k_0}) = 0$ . Neste caso  $x^{k_0}$  é um ponto estacionário e o Passo 1 não é mais possível. A condição  $\nabla f(x^k) = 0$  é necessária mas não é suficiente para deduzir

que  $x^k$  é uma solução do problema. Na verdade, este processo nos leva a detectar “candidatos” à solução.

Porém, é mais provável que o processo continue indefinidamente sem verificar a condição  $\nabla f(x^k) = 0$  para nenhum valor de  $k$ . Neste caso, mediante este algoritmo, estamos gerando uma seqüência infinita  $\{x^k\}$  de pontos em  $\mathbb{R}^n$ . Fazem sentido então as seguintes perguntas:

1. Existe  $\lim_{k \rightarrow \infty} x^k$  ?
2. Se  $\lim_{k \rightarrow \infty} x^k = x^*$ , é possível garantir alguma das seguintes afirmações?
  - a)  $x^*$  é uma solução do problema;
  - b)  $x^*$  é um ponto estacionário.

Daremos alguns passos na direção de responder essas perguntas. Claramente, o Algoritmo 4.1 gera uma seqüência de pontos  $\{x^k\}$  tal que a seqüência de números reais associada  $\{f(x^k)\}$  é monótona decrescente.

Agora consideremos a função de uma variável  $f(x) = x^2$ . O único minimizador desta função é  $x^* = 0$ . A seqüência definida por

$$x^k = 1 + 1/k, \text{ para } k \geq 1$$

pode ser gerada pelo algoritmo porque

$$f(x^{k+1}) = (1 + 1/(k+1))^2 < (1 + 1/k)^2 = f(x^k).$$

No entanto,

$$\lim_{k \rightarrow \infty} x^k = 1.$$

Este exemplo mostra que a resposta à pergunta (2) é negativa.

Portanto, o método deve ser modificado para evitar situações como esta. No exemplo, o que parece estar acontecendo é que, apesar de haver sempre decréscimo da função, este decréscimo é pequeno demais devido à distância entre  $x^{k+1}$  e  $x^k$  que se aproxima de zero muito rapidamente.

O decréscimo pode ser muito pequeno também com distâncias grandes entre  $x^{k+1}$  e  $x^k$ , como vemos na Figura 4.2.

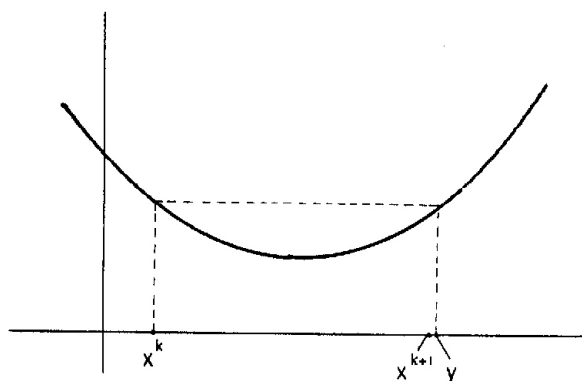


Figura 4.2

No caso da Figura 4.2,  $f(y) = f(x^k)$  e tomando  $x^{k+1}$  arbitrariamente próximo de  $y$  teremos  $f(x^{k+1}) < f(x^k)$ . Mas a diferença entre estes valores será arbitrariamente pequena.

Há uma terceira situação que pode levar-nos a obter decréscimos excessivamente pequenos do valor da função. Com efeito, consideremos o conjunto de nível que passa por  $x^k$ :

$$\Gamma = \{x \mid f(x) = f(x^k)\}.$$

Se nos limitássemos a andar sobre  $\Gamma$ , o decréscimo da função seria nulo. Assim, se a direção  $d_k$  é “quase” perpendicular a  $\nabla f(x^k)$ , essa direção é “quase” tangente a  $\Gamma$  em  $x^k$ . Neste caso também podemos ter pouco decréscimo do valor da função na direção  $d_k$ . Ilustramos na Figura 4.3 a situação.

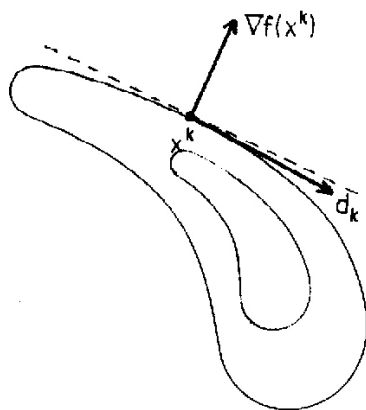


Figura 4.3



### 4.3 ALGORITMO COM CONVERGÊNCIA GLOBAL

Para impedir passos que se aproximem muito rapidamente de zero pediremos que

$$\|d_k\| \geq \sigma \|\nabla f(x^k)\|, \text{ para todo } k \in \mathbb{N},$$

onde  $\sigma > 0$  é uma constante.

Para impedir passos grandes com pouco decréscimo, na busca linear pediremos que  $\lambda_k$  verifique

$$f(x^k + \lambda_k d_k) < f(x^k) + \alpha \nabla^t f(x^k) \lambda_k d_k, \text{ para todo } k \in \mathbb{N},$$

onde  $\alpha \in (0, 1)$  é uma constante. Esta condição exige que o decréscimo seja em certo sentido proporcional ao tamanho do passo.

Observemos que, como  $d_k$  é uma direção de descida, resulta

$$\alpha \nabla^t f(x^k) \lambda_k d_k < 0$$

e, portanto, essa condição significa que queremos algo mais que simplesmente

um decréscimo no valor da função. Chamamos essa condição de *decréscimo suficiente*, também conhecida como *condição de Armijo*.

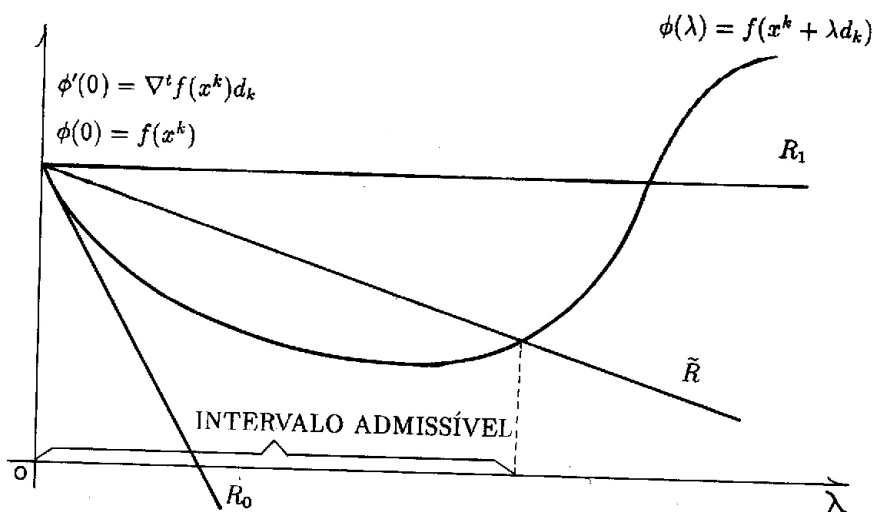


Figura 4.4 – Condição de Armijo

Na Figura 4.4  $R_0$  é a reta que passa pelo ponto  $(0, \phi(0))^t$  e tem coeficiente angular  $\phi'(0)$ . A equação de  $R_0$  é

$$z = \phi(0) + \phi'(0)\lambda,$$

$R_1$  é a reta que passa pelo mesmo ponto e tem coeficiente angular 0.  $\tilde{R}$  é uma reta que passa pelo mesmo ponto com coeficiente angular entre  $\phi'(0)$  e 0. Portanto, o coeficiente angular de  $\tilde{R}$  pode ser escrito da forma  $\alpha\phi'(0)$  com  $\alpha \in (0, 1)$ . Logo a equação de  $\tilde{R}$  é:

$$z = \phi(0) + \alpha\phi'(0)\lambda.$$

Substituindo nesta equação  $\phi(0)$  por  $f(x^k)$  e  $\phi'(0)$  por  $\nabla^t f(x^k)d_k$  obtemos

$$z = f(x^k) + \alpha\lambda\nabla^t f(x^k)d_k.$$

Então, os valores de  $\lambda$  que verificam a condição de Armijo são os que estão na região admissível na Figura 4.4.

Para impedir que as direções sejam “quase” ortogonais à  $\nabla f(x^k)$  pediremos que dada uma constante  $\theta \in (0, 1)$ ,

$$\nabla^t f(x^k)d_k \leq -\theta \|\nabla f(x^k)\| \|d_k\|, \text{ para todo } k \in \mathbb{N},$$

Se  $\beta$  é o ângulo entre  $\nabla f(x^k)$  e  $d_k$ ,

$$\cos \beta = \nabla^t f(x^k)d_k / (\|\nabla f(x^k)\| \|d_k\|)$$

e, conseqüentemente,

$$\cos \beta \leq -\theta.$$

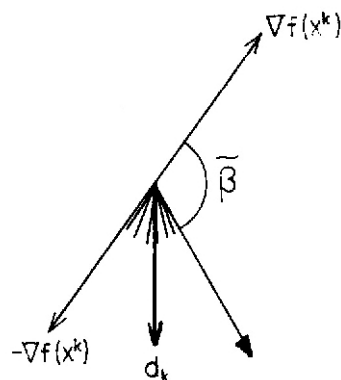


Figura 4.5

Na Figura 4.5, se  $\tilde{\beta}$  é um ângulo tal que  $\cos \tilde{\beta} = -\theta$ ,  $d_k$  deve formar um ângulo maior que  $\tilde{\beta}$  com  $\nabla f(x^k)$ . Vamos definir um algoritmo para minimizar

funções sem restrições, que seja o mais geral possível e que incorpore essas condições.

### Algoritmo 4.2

Sejam  $\sigma > 0$ ,  $\alpha$  e  $\theta \in (0, 1)$  constantes dadas. Se  $x^k \in \mathbb{R}^n$  é tal que  $\nabla f(x^k) \neq 0$ , os passos para determinar  $x^{k+1}$  são:

**Passo 1:** Escolher  $d_k \in \mathbb{R}^n$ , tal que

- (i)  $\|d_k\| \geq \sigma \|\nabla f(x^k)\|$ ;
- (ii)  $\nabla^t f(x^k) d_k \leq -\theta \|\nabla f(x^k)\| \|d_k\|$ .

**Passo 2:** (Busca linear)

- (i)  $\lambda = 1$ ;
- (ii) Se  $f(x^k + \lambda d_k) < f(x^k) + \alpha \lambda \nabla^t f(x^k) d_k$ , ir a (iv);
- (iii) Escolher  $\tilde{\lambda} \in [0.1\lambda, 0.9\lambda]$ . Fazer  $\lambda = \tilde{\lambda}$  e ir a (ii);
- (iv) Fazer  $\lambda_k = \lambda$ , e  $x^{k+1} = x^k + \lambda_k d_k$ .

### Lema 4.1

O Algoritmo 4.2 está bem-definido. (É possível completar a busca linear com um número finito de tentativas para  $\lambda$ ).

**Prova:** Fica como exercício para o leitor. ■

Enunciaremos um teorema que responde as perguntas (1) e (2), feitas em 4.2.

### Teorema 4.1 (Convergência Global)

O Algoritmo 4.2 pára com algum valor  $k$  tal que  $\nabla f(x^k) = 0$ , ou gera uma seqüência infinita  $\{x^k\}$  tal que qualquer ponto de acumulação dela é um ponto estacionário de  $f$ .

**Prova:** Trata-se de um caso particular do teorema demonstrado em Friedlander et al.[6]. ■

Observemos que neste teorema não é garantida a convergência da seqüência  $\{x^k\}$ . No entanto, ele afirma que se existe  $\lim_{k \rightarrow \infty} x^k$ , então este limite é um ponto estacionário. Finalmente, se a seqüência é limitada existe um ponto de acumulação e este deve ser um ponto estacionário.

**Exercícios**

4.1 Considere a função quadrática  $f(x) = \frac{1}{2}x^tAx + b^tx + c$ , onde  $A \in \mathbb{R}^{n \times n}$  simétrica,  $b \in \mathbb{R}^n$  e  $c \in \mathbb{R}$ . Seja  $\tilde{x}$  minimizador local de  $f$ . Prove que  $\tilde{x}$  é minimizador global.

4.2 Através de um desenho mostre que se  $d$  é uma direção tal que  $\nabla^t f(x)d = 0$  então  $d$  pode ser de descida, subida ou nenhuma das duas coisas.

4.3 Considere o sistema não-linear

$$f_i(x) = 0, \quad f_i : \mathbb{R}^n \rightarrow \mathbb{R}, \quad i = 1, \dots, m.$$

Como resolveria o sistema com técnicas de minimização irrestrita?

4.4 Seja  $f(x) = \frac{1}{2}\|F(x)\|^2$ , onde  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n, F \in C^1$ . Considere o método iterativo definido por

$$x^{k+1} = x^k - \lambda_k (J_F(x^k))^{-1} F(x^k).$$

Suponha que  $J_F(x)$  é não-singular para todo  $x$ . Prove que se na condição de Armijo usamos  $\alpha = 0.5$ , resulta

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

4.5 Seja  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f \in C^2$ ,  $f'(0) < 0$  e  $f''(x) < 0$  para todo  $x \in \mathbb{R}$ . Seja  $\alpha \in (0, 1)$ . Prove que, para todo  $x > 0$ ,

$$f(x) \leq f(0) + \alpha x f'(0).$$

4.6 Se um método de direções de descida com busca linear exata é utilizado para minimizar uma função quadrática  $q : \mathbb{R}^n \rightarrow \mathbb{R}$ , mostre que o passo ótimo é dado por

$$\lambda = -\frac{d^t \nabla q(x)}{d^t \nabla^2 q(x) d},$$

onde  $d$  é a direção utilizada a partir do ponto  $x$ .

4.7 O critério de decréscimo suficiente (condição de Armijo) exige  $\lambda \in \mathbb{R}$  tal que

$$\varphi(\lambda) = f(x + \lambda d) < f(x) + \alpha \lambda \nabla^t f(x) d = \varphi(0) + \alpha \lambda \varphi'(0), \quad (*)$$

---

com  $\alpha \in (0, 1)$ . Se  $f$  é uma função quadrática, então  $\varphi$  é uma parábola. Prove que se o minimizador  $\tilde{\lambda}$  dessa parábola é admissível em (\*) devemos ter  $\alpha \in (0, 1/2)$ .

4.8 Sejam  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $x, d \in \mathbb{R}^n$  e  $\lambda > 0$  tal que  $x + \lambda d$  satisfaz a condição de Armijo. Seja  $0 < \mu < \lambda$ .  $\mu$  satisfaz a condição de Armijo? Prove ou dê um contra-exemplo.

4.9 Sejam  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^2$  e  $\tilde{x} \in \mathbb{R}^n$  tal que  $\nabla f(\tilde{x}) = 0$  e  $\nabla^2 f(\tilde{x})$  não é semidefinida positiva. Prove que existe uma direção de descida  $d$  em  $\tilde{x}$ .

4.10 No processo de minimizar uma função  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^1$ , a iteração  $x^k$  foi obtida fazendo uma busca linear ao longo da direção  $d^{k-1}$ . Determine uma direção  $d^k$  ortogonal a  $d^{k-1}$ , de descida a partir de  $x^k$  e que seja uma combinação linear de  $d^{k-1}$  e  $\nabla f(x^k)$ .

4.11 Sejam  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $\tilde{x} \in \mathbb{R}^n$  com  $\nabla f(\tilde{x}) \neq 0$ . Seja  $M \in \mathbb{R}^{n \times n}$  definida positiva. Prove que  $d = -M \nabla f(\tilde{x})$  é uma direção de descida em  $\tilde{x}$ .



## Capítulo 5

# ORDEM DE CONVERGÊNCIA

Se a seqüência  $\{x^k\}$  gerada pelo Algoritmo 4.2 converge, podemos nos perguntar sobre a rapidez da convergência. Para analisar este aspecto introduzimos o conceito de ordem de convergência. Claramente, se  $\lim_{k \rightarrow \infty} x^k = x^*$ , então  $\lim_{k \rightarrow \infty} \|x^k - x^*\| = 0$  e podemos considerar que  $\|x^k - x^*\|$  é o erro cometido na aproximação  $x^k$ . Quanto mais “rápido” o erro se aproximar de zero, melhor. Uma forma de medir este progresso é comparar os erros cometidos em duas aproximações sucessivas

$$e_{k+1} = \|x^{k+1} - x^*\| \text{ e } e_k = \|x^k - x^*\|.$$

Obviamente é desejável que a partir de algum índice  $k_0$ , seja verdade que

$$e_{k+1} \leq r e_k \tag{5.1}$$

para algum  $r \in [0, 1)$ .

A inequação (5.1) significa que o erro na aproximação  $x^{k+1}$  não pode superar uma fração do erro na aproximação  $x^k$ , determinada pela constante  $r$ .

A condição  $r < 1$  exclui a possibilidade de que  $e_{k+1}/e_k$  se aproxime arbitrariamente de 1, situação na qual o progresso seria lento demais. Quanto menor for  $r$ , mais rápida será a convergência da seqüência  $\{x^k\}$ .

### Definição 5.1

*Se (5.1) se verifica para algum  $r \in (0, 1)$ , diremos que a seqüência  $\{x^k\}$  converge com ordem linear e taxa não-superior a  $r$ .*

### Definição 5.2

*Se*

$$\lim_{k \rightarrow \infty} e_{k+1}/e_k = 0, \tag{5.2}$$

diremos que a seqüência  $\{x^k\}$  converge com ordem superlinear.

A convergência superlinear significa que, assintoticamente, a redução do erro é maior que qualquer fração fixa.

Podemos ainda caracterizar a convergência com “ordem melhor que superlinear”.

**Definição 5.3**

Se  $e_{k+1} \leq a (e_k)^p$ , onde  $a > 0$  e  $p > 1$ , diremos que a seqüência  $\{x^k\}$  converge a  $x^*$  com ordem não-inferior a  $p$ . Se  $p = 2$ , diremos que a convergência é quadrática.

**Exercícios**

5.1 Prove que convergência superlinear implica linear.

5.2 Prove que convergência quadrática implica superlinear.

5.3 Mostre que uma seqüência pode convergir linearmente com uma norma mas não com outra. No entanto, a convergência superlinear é independente da norma.



## Capítulo 6

# MÉTODOS CLÁSSICOS DE DESCIDA

### 6.1 MÉTODO DO GRADIENTE

No contexto do Algoritmo 4.2, este método corresponde a escolher  $d_k$  na direção de  $-\nabla f(x^k)$ .

Se, no Passo 1 do Algoritmo 4.2,  $d_k = -\sigma \nabla f(x^k)$ , as condições (i) e (ii) são verificadas trivialmente. Consideremos o seguinte algoritmo para minimizar uma função  $f$  definida em  $\mathbb{R}^n$ .

#### Algoritmo 6.1

*Se  $x^k \in \mathbb{R}^n$  é tal que  $\nabla f(x^k) \neq 0$ , os passos para determinar  $x^{k+1}$  são:*

**Passo 1:** Calcular  $d_k = -\nabla f(x^k)$ .

**Passo 2:** (Busca linear exata)

*Determinar  $\lambda_k$ , minimizador de  $f(x^k + \lambda d_k)$  sujeita a  $\lambda \geq 0$ .*

**Passo 3:** Fazer  $x^{k+1} = x^k + \lambda_k d_k$ .

#### Observações:

No Passo 1 as condições (i) e (ii) do Algoritmo 4.2 são omitidas.

No Passo 2 a busca linear é mais exigente que a do Algoritmo 4.2, porque  $\lambda_k$  é o minimizador de  $f$  na direção  $d_k$ . Chamamos a este processo de *busca linear exata*. É importante notar que este subproblema pode não ter solução e portanto o Algoritmo 6.1 nem sempre está bem-definido.

**Caso 1: Função objetivo quadrática**

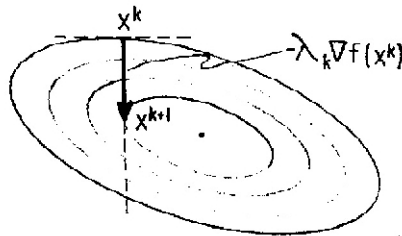
Se

$$f(x) = \frac{1}{2}x^t G x + b^t x + c$$

com  $G$  definida positiva, então existe um único  $x^* \in \mathbb{R}^n$  que é minimizador global de  $f$ . Ver Figura 6.1.

Neste caso a busca linear exata determina

$$\lambda_k = \nabla^t f(x^k) \nabla f(x^k) / \nabla^t f(x^k) G \nabla f(x^k).$$



**Figura 6.1**

O seguinte teorema garante a convergência da seqüência gerada pelo Algoritmo 6.1, para qualquer aproximação inicial e que a ordem de convergência da seqüência associada  $\{f(x^k)\}$  é linear.

**Teorema 6.1**

Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  uma função quadrática com matriz hessiana  $G$  definida positiva. Seja  $x^*$  o minimizador global de  $f$ .

Dado  $x^0 \in \mathbb{R}^n$ , arbitrário, o Algoritmo 6.1 gera uma seqüência  $\{x^k\}$  tal que:

(i)  $\lim_{k \rightarrow \infty} x^k = x^*$

(ii)  $\lim_{k \rightarrow \infty} f(x^k) = f(x^*)$

e

$$f(x^{k+1}) - f(x^*) \leq ((A - a)/(A + a))^2 (f(x^k) - f(x^*)),$$

onde  $A$  e  $a$  são o maior e o menor autovalor de  $G$ , respectivamente.

**Prova:** Ver Luenberger [11].■

## Caso 2: Função objetivo não quadrática

Enunciaremos um teorema que não garante convergência mas que fala da ordem quando a convergência ocorre.

### Teorema 6.2

Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^2$ . Seja  $x^* \in \mathbb{R}^n$  um minimizador local de  $f$ , tal que a matriz  $\nabla^2 f(x^*)$  é definida positiva. Se o Algoritmo 6.1 está bem-definido para todo  $k \in \mathbb{N}$  e a seqüência  $\{x^k\}$  gerada por ele converge a  $x^*$ , então a seqüência  $\{f(x^k)\}$  converge linearmente a  $f(x^*)$  com taxa não superior a  $((A - a)/(A + a))^2$ , onde  $A$  e  $a$  são o maior e o menor autovalor de  $\nabla^2 f(x^*)$ , respectivamente.

**Prova:** ver Luenberger [11].■

## 6.2 MÉTODO DE NEWTON

### Proposição 6.1

Se  $f$  é uma função quadrática com matriz hessiana  $G$  definida positiva, dado  $x^0 \in \mathbb{R}^n$  arbitrário, a direção  $d \in \mathbb{R}^n$  dada por:

$$d = -G^{-1}(G x^0 + b) \quad (6.1)$$

verifica que

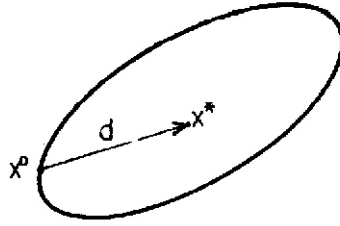
$$x^* \equiv x^0 + d \quad (6.2)$$

é o minimizador global de  $f$  em  $\mathbb{R}^n$ . Ver Figura 6.2.

**Prova:** Seja  $f(x) = \frac{1}{2}x^t G x + b^t x + c$ . Temos, por (6.2), que  $\nabla f(x^*) = G(x^0 + d) + b$ . Logo, usando (6.1), obtemos que

$$\nabla f(x^*) = G(x^0 - G^{-1}(G x^0 + b)) + b.$$

Portanto,  $\nabla f(x^*) = G x^0 - G x^0 - b + b = 0$ , o que prova a proposição.■



**Figura 6.2**

A direção  $d$  é a solução do sistema linear

$$Gd = -(Gx^0 + b) = -\nabla f(x^0).$$

Portanto, minimizar uma função quadrática com hessiana definida positiva é um problema equivalente a resolver um sistema linear com matriz simétrica e definida positiva.

Se a função não é quadrática e temos uma aproximação  $x^k$  da solução de

$$\text{Minimizar } f(x), \quad x \in \mathbb{R}^n,$$

podemos utilizar o resultado anterior na função quadrática que resulta da consideração dos três primeiros termos do desenvolvimento em série de Taylor de  $f$  em torno de  $x^k$ :

$$q(d) = f(x^k) + \nabla^t f(x^k)d + \frac{1}{2} d^t \nabla^2 f(x^k)d.$$

Chamamos  $c = q(0) = f(x^k)$ ,  $b = \nabla q(0) = \nabla f(x^k)$ ,  $G = \nabla^2 q(0) = \nabla^2 f(x^k)$ .

Se escrevemos  $q(d) = \frac{1}{2} d^t Gd + b^t d + c$  e se  $\nabla^2 f(x^k)$  é definida positiva podemos calcular o minimizador global desta quadrática a partir de  $d_0 = 0$ . Assim, obtemos

$$d^* = -G^{-1}(Gd_0 + b) = -G^{-1}b = -(\nabla^2 f(x^k))^{-1} \nabla f(x^k).$$

Isto sugere a escolha  $d_k = -(\nabla^2 f(x^k))^{-1} \nabla f(x^k)$  no Passo 1 do Algoritmo 4.2.

As seguintes perguntas são pertinentes:

$d_k$  é sempre uma direção de descida?

Se  $d_k$  é uma direção de descida, as condições (i) e (ii) do Passo 1 do Algoritmo 4.2 serão verificadas?

Infelizmente,  $d_k$  pode não ser uma direção de descida se  $\nabla^2 f(x^k)$  não

for definida positiva. Por exemplo, a função  $f(x, y) = (1/2)(x^2 - y^2)$  no ponto  $x^0 = (0, 1)^t$  verifica que:

$$\nabla f(x^0) = (0, -1)^t, \text{ e } \nabla^2 f(x^0) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Neste caso a direção de Newton é

$$d_0 = (0, -1)^t,$$

e

$$\nabla^t f(x^0)d_0 = 1 > 0.$$

Apesar de  $d_0$  ser uma direção de subida, pode-se argumentar que basta escolher  $\tilde{d} = -d_0$  para obter uma direção de descida. Mas o seguinte exemplo devido a Powell mostra que a situação pode não ter conserto:

A função  $f(x, y) = x^4 + xy + (1 + y)^2$  em  $x^0 = (0, 0)^t$  verifica

$$\nabla f(x^0) = (0, 2)^t \text{ e } \nabla^2 f(x^0) = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}.$$

A solução de  $\nabla^2 f(x^0)d = -(0, 2)^t$  é  $d_0 = (-2, 0)^t$  e  $\nabla^t f(x^0)d_0 = 0$ .

No caso em que  $d_k$  é uma direção de descida, a verificação de (i) e (ii) no Passo 1 do Algoritmo 4.2 depende de propriedades da função objetivo. Uma hipótese para garantir estas condições é que os autovalores das matrizes  $\nabla^2 f(x)$  estejam uniformemente incluídos em algum intervalo  $(a, b) \subset \mathbb{R}$ , com  $a > 0$ .

Consideremos agora o seguinte algoritmo:

### Algoritmo 6.2 (Método de Newton)

Se  $x^k$  é tal que  $\nabla f(x^k) \neq 0$ , os passos para determinar  $x^{k+1}$  são:

**Passo 1:** Determinar  $d_k$  tal que

$$\nabla^2 f(x^k)d_k = -\nabla f(x^k),$$

(ou seja, resolver este sistema linear. Notemos que este passo pode não estar bem-definido se  $\nabla^2 f(x^k)$  for singular.)

**Passo 2:** Fazer  $x^{k+1} = x^k + \lambda_k d_k$ , onde  $\lambda_k$  é determinado como no Passo 2 do Algoritmo 4.2.

Para o Algoritmo 6.2 temos o seguinte resultado:

**Teorema 6.3**

Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^3$ . Seja  $x^*$  um minimizador local de  $f$  em  $\mathbb{R}^n$ , tal que  $\nabla^2 f(x^*)$  é definida positiva. Então, existe  $\varepsilon > 0$  tal que se  $x^0 \in \mathcal{B}(x^*, \varepsilon)$ , e  $\lambda_k = 1$  para todo  $k \in \mathbb{N}$ , a seqüência  $\{x^k\}$  gerada pelo Algoritmo 6.2 verifica:

- (i)  $\nabla^2 f(x^k)$  é definida positiva para todo  $k \in \mathbb{N}$ ;
- (ii)  $\lim_{k \rightarrow \infty} x^k = x^*$ ;
- (iii) Existe  $c > 0$  tal que  $\|x^{k+1} - x^*\| \leq c \|x^k - x^*\|^2$  para todo  $k \in \mathbb{N}$ .

**Prova:** ver Luenberger [11].■

Este é um resultado de convergência local que diz que se escolhermos  $x^0$  suficientemente perto de  $x^*$ ,

- (i) os sistemas lineares do Passo 1 têm solução única e portanto  $d_k$  está bem-definido para todo  $k \in \mathbb{N}$ ;
- (ii) a seqüência converge a  $x^*$ ;
- (iii) a ordem de convergência é pelo menos quadrática.

Uma pergunta óbvia que surge aqui é: como sabemos se  $x^0$  está suficientemente próximo de  $x^*$ ? Em geral, não sabemos. Mas, se usarmos o Algoritmo 4.2 com

$$d_k = -(\nabla^2 f(x^k))^{-1} \nabla f(x^k)$$

para uma função na qual as condições (i) e (ii), do Passo 1 do Algoritmo 4.2, possam ser obtidas e a seqüência gerada pelo algoritmo converge a um minimizador  $x^*$  de  $f$  com  $\nabla^2 f(x^*)$  definida positiva, podemos afirmar que a partir de algum índice  $k_0$  os termos da seqüência estarão tão próximos de  $x^*$  quanto é preciso para obter o resultado anterior.

Frisamos aqui o fato de que a ordem de convergência quadrática é obtida devido ao uso das derivadas segundas (a matriz  $\nabla^2 f(x)$ ). É bom lembrar que considerar essa informação envolve avaliar  $\nabla^2 f(x^k)$  e resolver  $\nabla^2 f(x^k) d_k = -\nabla f(x^k)$ . Portanto, o processo é caro em termos de trabalho computacional (tempo). Se o número de variáveis for muito grande a memória necessária para armazenar esta informação pode ser insuficiente e este processo torna-se inviável.

**6.3 MÉTODOS QUASE-NEWTON**

No método do gradiente escolhemos

$$d_k = -I \nabla f(x^k),$$

e, no método de Newton,

$$d_k = -(\nabla^2 f(x^k))^{-1} \nabla f(x^k).$$

Outros métodos podem ser definidos fazendo

$$d_k = -H_k \nabla f(x^k),$$

onde  $H_k \in \mathbb{R}^{n \times n}$  é uma matriz simétrica. Se  $H_k$  for definida positiva,  $d_k$  é uma direção de descida.

É desejável determinar matrizes  $H_k$  de modo que o trabalho computacional do método resultante seja menor que o do método de Newton e tais que a seqüência  $\{x^k\}$  gerada por ele, quando converge, tenha ordem pelo menos superlinear.

Se quisermos obter um comportamento melhor do que o do método do gradiente, precisaremos utilizar alguma informação de segunda ordem.

Outra vez a análise específica das funções quadráticas é pertinente.

Se  $x^*$  é o minimizador global de uma quadrática com matriz hessiana definida positiva, o método de Newton encontra  $x^*$  numa única iteração a partir de qualquer  $x^0 \in \mathbb{R}^n$ . O método do gradiente converge a  $x^*$ , mas não necessariamente num número finito de iterações.

Um método intermediário para funções quadráticas encontraria  $x^*$  num número finito de iterações sem estar baseado no conhecimento completo da matriz hessiana.

Se

$$f(x) = \frac{1}{2} x^t G x + b^t x + c,$$

temos que

$$\nabla f(x) = Gx + b$$

e

$$\nabla f(x + d) - \nabla f(x) = G(x + d) - Gx = Gd \text{ para todo } d \in \mathbb{R}^n.$$

Temos então as seguintes equações:

$$\nabla f(x + d) - \nabla f(x) = Gd$$

ou

$$G^{-1}(\nabla f(x + d) - \nabla f(x)) = d.$$

Observemos que estas equações fornecem informação sobre  $G$  ou  $G^{-1}$  utilizando  $\nabla f$  em dois pontos. Dados  $n$  pares de pontos  $\{x^i, x^i + d_i\}$ , de modo que o conjunto de vetores  $\{d_1, d_2, \dots, d_n\}$  é linearmente independente, as  $n$  diferenças

$$\nabla f(x^i + d_i) - \nabla f(x^i)$$

determinam completamente  $G$  e  $G^{-1}$ . Isto significa que a informação contida nessas  $n$  diferenças equivale à informação completa de  $G$  e  $G^{-1}$ .

Estas observações sugerem o seguinte algoritmo.

### Algoritmo 6.3

Sejam  $x^0 \in \mathbb{R}^n$  arbitrário,  $H_0 \in \mathbb{R}^{n \times n}$  simétrica e definida positiva. Se  $\nabla f(x^k) \neq 0$ , os passos para obter  $x^{k+1}$  são:

**Passo 1:** Calcular  $d_k = -H_k \nabla f(x^k)$ .

**Passo 2:** Determinar  $\lambda_k$  através de uma busca linear e definir  $x^{k+1} = x^k + \lambda_k d_k$ .

**Passo 3:** Determinar  $H_{k+1}$  simétrica e definida positiva tal que

$$H_{k+1}(\nabla f(x^{j+1}) - \nabla f(x^j)) = x^{j+1} - x^j \text{ para todo } j \leq k.$$

Se a função objetivo é quadrática e o conjunto  $\{d_0, d_1, \dots, d_{n-1}\}$  é linearmente independente, pelas observações anteriores teremos que

$$H_n = G^{-1}.$$

Portanto,

$$d_n = -G^{-1}(\nabla f(x^n))$$

e

$$x^{n+1} = x^*.$$

É possível construir um algoritmo com estas propriedades. O primeiro método deste tipo foi proposto por Davidon, Fletcher e Powell e consiste no seguinte:

### Algoritmo 6.4 (DFP)

Sejam  $x^0 \in \mathbb{R}^n$  arbitrário e  $H_0 \in \mathbb{R}^{n \times n}$  uma matriz simétrica e definida positiva. Se  $\nabla f(x^k) \neq 0$ , os passos para obter  $x^{k+1}$  são:

**Passo 1:** Calcular  $d_k = -H_k \nabla f(x^k)$ .

**Passo 2:** Determinar  $\lambda_k$  através de uma busca linear e definir  $x^{k+1} = x^k + \lambda_k d_k$ .



**Passo 3:** Definir  $p_k = \lambda_k d_k = x^{k+1} - x^k$ ,  $q_k = \nabla f(x^{k+1}) - \nabla f(x^k)$  e calcular

$$H_{k+1} = H_k + (p_k p_k^t) / (p_k^t q_k) - (H_k q_k q_k^t H_k) / (q_k^t H_k q_k).$$

### Observações:

O que caracteriza o método DFP é a fórmula recursiva do Passo 3 para atualizar  $H_k$ .

Notemos que  $H_{k+1}$  é obtida a partir de uma correção de  $H_k$  que consiste em somar duas matrizes simétricas da forma  $vv^t$ , onde  $v \in \mathbb{R}^n$ . Cada uma dessas matrizes tem posto 1.

A vantagem em termos de trabalho computacional é que o número de operações para determinar  $d_k$  é da ordem de  $n^2$ , em lugar de  $n^3$  como no método de Newton.

### Teorema 6.4

Se o método DFP é usado para minimizar uma função quadrática com hessiana definida positiva fazendo busca linear exata, então:

- (i) Se  $H_k$  é definida positiva então  $H_{k+1}$  também é;
- (ii)  $\{d_0, d_1, \dots, d_{n-1}\}$  é linearmente independente;
- (iii)  $H_k q_j = p_j$  para todo  $j \leq k$ ;
- (iv)  $x^n = x^*$ ;
- (v)  $H_n = G^{-1}$ .

**Prova:** Ver Bazaraa e Shetty [2].■

Outra fórmula com estas propriedades, muito popular devido a seu bom desempenho numérico, é devida a Broyden, Fletcher, Goldfarb, Shanno (BFGS):

$$H_{k+1}^{BFGS} = H_k + \left( \frac{1 + q_k^t H_k q_k}{q_k^t p_k} \right) \frac{p_k p_k^t}{p_k^t q_k} - \frac{p_k q_k^t H_k + H_k q_k p_k^t}{q_k^t p_k}.$$

Usando esta fórmula no Passo 3 do Algoritmo 6.4 resulta o método BFGS.

Para estes métodos temos o seguinte teorema de convergência local:

### Teorema 6.5

Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^3$  tal que existe  $x^*$  minimizador local de  $f$  com  $\nabla^2 f(x^*)$  definida positiva. Existem  $\delta > 0$ ,  $\varepsilon > 0$  tais que se  $x^0 \in \mathcal{B}(x^*, \delta)$  e  $\|H_0 - \nabla^2 f(x^*)\| < \varepsilon$ , as seqüências  $\{x^k\}$  e  $\{H_k\}$  geradas pelos métodos DFP e BFGS, usando  $\lambda_k = 1$  para todo  $k \in \mathbb{N}$  no Passo 2, verificam

- (i)  $H_k$  é definida positiva para todo  $k \in \mathbb{N}$ ;  
(ii)  $\lim_{k \rightarrow \infty} x^k = x^*$ ;  
(iii) a ordem de convergência é pelo menos superlinear.

**Prova:** Ver Dennis e Schnabel [4].■

### Exercícios

6.1 Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , diferenciável em  $\tilde{x}$  e sejam  $d^1, \dots, d^n \in \mathbb{R}^n$  vetores linearmente independentes. Suponha que o mínimo de  $f(\tilde{x} + \lambda d^j)$  com  $\lambda \in \mathbb{R}$  ocorra em  $\lambda = 0$  para  $j = 1, \dots, n$ . Prove que  $\nabla f(\tilde{x}) = 0$ . Isso implica que  $f$  tem um mínimo local em  $\tilde{x}$ ?

6.2 Seja  $f(x) = \frac{1}{2}x^t A x + b^t x + c$ , onde  $A \in \mathbb{R}^{n \times n}$  é simétrica e definida positiva,  $b \in \mathbb{R}^n$  e  $c \in \mathbb{R}$ . Sejam  $L_1$  e  $L_2$  duas retas diferentes e paralelas em  $\mathbb{R}^n$ , cujo vetor diretor é  $d$ . Sejam  $x^1$  e  $x^2$  minimizadores de  $f$  em  $L_1$  e  $L_2$ , respectivamente. Prove que  $(x^2 - x^1)^t A d = 0$ .

6.3 Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^1$ . Para  $k = 0, 1, 2, \dots$ , definimos  $x^{k+1} = x^k - \lambda_k \nabla f(x^k)$  onde  $\lambda_k \geq \bar{\lambda} > 0$  para todo  $k \geq 0$ . Suponha que  $\{x^k\}_{k=0}^\infty$  converja para  $\tilde{x}$ . Prove que  $\nabla f(\tilde{x}) = 0$ .

6.4 Prove que no método do gradiente com busca linear exata temos que  $\nabla^t f(x^k) \nabla f(x^{k+1}) = 0$ .

6.5 Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^1$ . Seja  $y$  o resultado de aplicarmos uma iteração do método do gradiente com busca linear exata a partir de  $x$ . Seja  $z$  o resultado de aplicarmos uma iteração do método do gradiente a partir de  $y$ . Prove que  $z - x$  é uma direção de descida a partir de  $x$ .

6.6 Desenhe as curvas de nível da função  $f(x) = x_1^2 + 4x_2^2 - 4x_1 - 8x_2$ . Encontre o ponto  $\tilde{x}$  que minimiza  $f$ . Prove que o método do gradiente, aplicado a partir de  $x^0 = (0, 0)^t$  não pode convergir para  $\tilde{x}$  em um número finito de passos, se usarmos busca linear exata. Há algum ponto  $x^0$  para o qual o método converge em um número finito de passos?

6.7 Considere o método do gradiente aplicado à minimização de uma função quadrática  $q(x)$  com hessiana definida positiva  $G$ . Seja  $\tilde{x}$  a solução e suponha que  $x^0$  possa ser escrito como  $x^0 = \tilde{x} + \mu v$ , onde  $v$  é um autovetor de  $G$  associado ao autovalor  $\lambda$  e  $\mu$  é um número real. Prove que  $\nabla q(x^0) = \mu \lambda v$  e que se for feita uma

busca linear exata a partir de  $x^0$  haverá convergência em uma iteração. A partir daí, mostre que o método do gradiente converge em uma iteração para qualquer  $x^0$  sempre que  $G$  for da forma  $\alpha I$  com  $\alpha \in \mathbb{R}$ .

6.8 Seja  $f$  uma função quadrática com hessiana definida positiva. Prove que se ao aplicarmos o método do gradiente a partir de um certo  $x^0$ ,  $\nabla f(x^0) \neq 0$ , encontramos a solução em uma iteração, então  $d = x^1 - x^0$  é um autovetor da hessiana.

6.9 Seja  $f(x) = \frac{1}{2}(x_1^2 - x_2)^2 + \frac{1}{2}(1 - x_1)^2$ . Qual é o minimizador de  $f$ ? Faça uma iteração do método de Newton para minimizar  $f$  a partir de  $x^0 = (2, 2)^t$ . É um bom passo? Antes de decidir, calcule  $f(x^0)$  e  $f(x^1)$ .

6.10 Considere o método de Newton aplicado para achar o minimizador de  $f(x) = \sin x$  a partir de  $x_0 \in [-\pi, \pi]$ . A resposta desejada é  $\tilde{x} = -\pi/2$ . Seja  $\varepsilon > 0$  suficientemente pequeno. Prove que se  $x_0 = -\varepsilon$  então  $x_1 \simeq -1/\varepsilon$ . Analogamente, o que acontece se  $x_0 = \varepsilon$ , mas  $f''(x_0)$  é substituída por um número positivo pequeno?

6.11 O método de Newton pode convergir para um maximizador local! Para verificar esta afirmação, use o método de Newton para minimizar a função  $f(x) = -x^4/4 + x^3/3 + x^2$  a partir de  $x_0 = 1$  e tomando  $\lambda_0 = 1$ . O que acontece com o método de Newton quando aplicado à minimização de  $f(x) = x^3/3 + x$  (equivalente a calcular os zeros de  $f'(x) = x^2 + 1$ )?

6.12 Seja  $f(x) = x_1^4 + x_1x_2 + (1 + x_2)^2$ . Para  $x^0 = (0, 0)^t$ , por que o método de Newton não pode ser aplicado satisfatoriamente? Se a direção  $d^0 = -(\nabla^2 f(x^0))^{-1}\nabla f(x^0)$  é usada, mostre que nem  $d^0$  nem  $-d^0$  são direções de descida.

6.13 No método de Newton é necessário que a matriz hessiana seja definida positiva. Na prática devemos modificar o método quando falha essa hipótese. Uma idéia é tomar

$$M^k = (\nabla^2 f(x^k) + \mu_k I)^{-1}, \quad \mu_k > 0,$$

$$d^k = -M^k \nabla f(x^k).$$

- (a) Quais são os valores aceitáveis de  $\mu_k$  para garantir que o método gere direções de descida?  
 (b) Que método é esse quando  $\mu \rightarrow \infty$ ?

6.14 Seja  $f(x) = \sum_{i=1}^n (a_i x_i^2 + b_i x_i)$  com  $a_1, \dots, a_n$  e  $b_1, \dots, b_n$  constantes reais. Encontre condições suficientes para que a direção utilizada pelo método de Newton esteja bem-definida e seja de descida para qualquer  $x$  tal que  $\nabla f(x) \neq 0$ .

6.15 Prove que  $A = vv^t$  onde  $0 \neq v \in \mathbb{R}^n$  tem posto 1.

6.16 Seja  $0 \neq s \in \mathbb{R}^n$ . Prove que  $\|I - ss^t/s^t s\| = 1$ .

6.17 Sejam  $u, v \in \mathbb{R}^n$  e suponha que  $A \in \mathbb{R}^{n \times n}$  é não-singular. Seja  $B = A + uv^t$ . Se  $\sigma = 1 + v^t A^{-1} u \neq 0$  verifique a fórmula de Sherman-Morrison:

$$B^{-1} = A^{-1} - \frac{1}{\sigma} A^{-1} u v^t A^{-1}.$$

6.18 Seja  $H \in \mathbb{R}^{n \times n}$  simétrica definida positiva e seja  $\{v^1, \dots, v^n\}$  uma base ortonormal de autovetores de  $H$  com autovalores associados  $\{\lambda_1, \dots, \lambda_n\}$ . Prove que para  $g = \sum_{i=1}^m \alpha_i v^i$  e  $\mu \geq 0$  temos

$$(H + \mu I)^{-1} g = \sum_{i=1}^n \left( \frac{\alpha_i}{\lambda_i + \mu} \right) v^i.$$

6.19 Considere a formula DFP. Se  $H^k$  é definida positiva mostre que  $H^{k+1}$  será definida positiva se o passo  $\lambda_k > 0$  é tal que  $(x^{k+1} - x^k)^t (\nabla f(x^{k+1}) - \nabla f(x^k)) > 0$ . Prove que para uma função quadrática qualquer  $\lambda_k \neq 0$  garante a positividade de  $H^{k+1}$ .

6.20 Considere o problema de minimizar uma função  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^2$ , cuja matriz hessiana tem a forma  $\nabla^2 f(x^k) = I + F^k$ , onde  $I$  é a matriz identidade e  $F^k$  é uma matriz esparsa com  $\|F^k\| < 1$ . Sabe-se que para  $\|A\| < 1$  vale a igualdade

$$(I + A)^{-1} = I - A + A^2 - A^3 + \dots$$

- (a) Verifique a afirmação acima;
- (b) Descreva como utilizar um método quase-Newton de maneira eficiente.

6.21 Aplique o método DFP com busca linear exata para minimizar a função  $f(x) = 10x_1^2 + x_2^2$  a partir de  $x^0 = (0.1, 1)^t$  com  $H^0 = I$ . Verifique a propriedade de terminação em  $n$  passos para funções quadráticas, onde  $n$  é a dimensão do problema.

6.22 Considere o método quase-Newton com correção de posto 1

$$H^{k+1} = H^k + \frac{(p - H^k q)(p - H^k q)^t}{q^t(p - H^k q)},$$

onde  $p = x^{k+1} - x^k$  e  $q = \nabla f(x^{k+1}) - \nabla f(x^k)$ . Sobre que condições a correção acima pode ser utilizada?

6.23 Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^1$ . Considere o método quase-Newton definido por  $x^{k+1} = x^k - B_k^{-1} \nabla f(x^k)$ , onde a fórmula de recorrência para as  $B_k$  é

$$B_{k+1} = B_k + yy^t/y^t p,$$

$y = q - B_k p$ ,  $q = \nabla f(x^{k+1}) - \nabla f(x^k)$  e  $p = x^{k+1} - x^k$ . Se  $z = p - B_k^{-1} q$ , mostre que se  $B_{k+1}$  é inversível, então

$$B_{k+1}^{-1} = B_k^{-1} + zz^t/z^t q.$$

6.24 Considere o espaço  $Q(u, v) = \{A \in \mathbb{R}^{n \times n} | Au = v\}$ . Prove que  $Q(u, v)$  é uma variedade afim. Qual é a sua dimensão? Idem para  $\tilde{Q}(u, v) = \{A \in Q(u, v) | A = A^t\}$ . Seja  $F(x) = Gx + b$  com  $G \in \mathbb{R}^{n \times n}$  e  $b \in \mathbb{R}^n$ . Prove que para quaisquer  $x, y \in \mathbb{R}^n$ ,  $G \in Q(y - x, F(y) - F(x))$ .



## Capítulo 7

# MINIMIZAÇÃO COM RESTRICÇÕES LINEARES DE IGUALDADE

A partir deste capítulo analisaremos casos em que o conjunto factível  $S$  não é necessariamente  $\mathbb{R}^n$ . A dificuldade dos problemas de *minimização com restrições* depende fortemente da complexidade destas. O caso mais geral que será tratado neste livro é

Minimizar  $f(x)$

sujeita a  $h(x) = 0, g(x) \leq 0$ ,

onde  $f, h, g \in C^2, f: \mathbb{R}^n \rightarrow \mathbb{R}, h: \mathbb{R}^n \rightarrow \mathbb{R}^m$  com  $m < n$  e  $g: \mathbb{R}^n \rightarrow \mathbb{R}^p$ .  
Ou seja,  $S = \{x \in \mathbb{R}^n \mid h(x) = 0 \text{ e } g(x) \leq 0\}$ .

Nesta seção consideramos a situação mais simples:

Minimizar  $f(x)$

sujeita a  $Ax = b$ , (7.1)

onde  $A \in \mathbb{R}^{m \times n}, 1 \leq m < n$  e posto  $A = m$ .

### 7.1 A REGIÃO DE FACTIBILIDADE

$S \equiv \{x \in \mathbb{R}^n \mid Ax = b\}$  é chamado conjunto de factibilidade de (7.1). Este conjunto é a variedade afim de soluções do sistema linear

$$Ax = b. \tag{7.2}$$

Se  $n = 2$ ,  $S$  é uma reta. Para  $n = 3$ ,  $S$  é um plano se  $m = 1$  ou uma reta se  $m = 2$ . Em geral,  $S$  é uma reta se  $m = n - 1$ , um plano se  $m = n - 2$  e uma variedade de dimensão  $n - m$  para  $m$  genérico. Se  $n > 3$  e  $m = 1$

falaremos em hiperplanos.

Associado a  $S$ , temos o conjunto de soluções do sistema homogêneo  $Ax = 0$  que é chamado Núcleo de  $A$  e denotado  $Nu(A)$ . Este é um subespaço de  $\mathbb{R}^n$  de dimensão  $n - m$ , já que posto de  $A = m$ . Claramente,  $Nu(A)$  é paralelo a  $S$  e passa pela origem. Ver Figura 7.1.

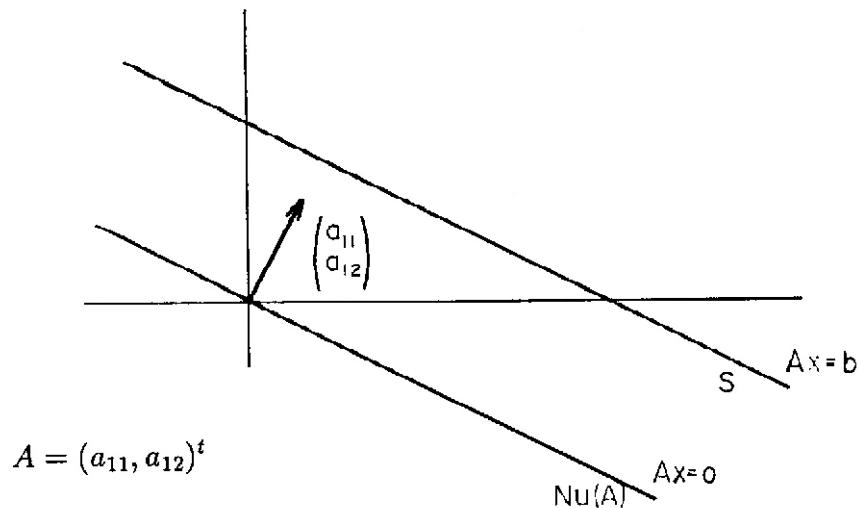


Figura 7.1

Pela sua própria definição, as linhas de  $A$  são ortogonais a  $Nu(A)$ . Mais ainda, como posto  $A = m$ , temos que as  $m$  linhas de  $A$  formam um conjunto de vetores linearmente independentes e geram um subespaço de dimensão  $m$  ortogonal a  $Nu(A)$ , que denotamos  $Im(A^t)$  (Imagem de  $A^t$ ).

Os subespaços  $Nu(A)$  e  $Im(A^t)$  verificam

$$\mathbb{R}^n = Nu(A) + Im(A^t)$$

e

$$Nu(A) \cap Im(A^t) = \{0\}.$$

Se  $d \in Nu(A)$  e  $\tilde{x}$  é uma solução de (7.2), então  $x \equiv \tilde{x} + \alpha d$  também é uma solução de (7.2). Em outras palavras, qualquer  $d \in Nu(A)$  é uma direção no espaço na qual podemos nos deslocar a partir de uma solução factível sem correr o risco de abandonar a região de factibilidade. A afirmação recíproca também é válida. Se a partir de uma solução factível  $\tilde{x}$ , andando numa direção  $d \in \mathbb{R}^n$  obtemos

$$x = \tilde{x} + \alpha d \text{ e } Ax = b,$$



então, necessariamente  $Ad = 0$  e, portanto,  $d \in Nu(A)$ . Diremos que  $Nu(A)$  é o conjunto de direções factíveis em  $S$ .

Se  $\{z^1, z^2, \dots, z^{n-m}\}$  é uma base de  $Nu(A)$  e denotamos  $Z$  a matriz de  $n \times (n - m)$  cujas colunas são os vetores  $z^i$ , resulta que para todo  $d \in Nu(A)$ , existe  $\gamma \in \mathbb{R}^{n-m}$  tal que  $d = Z\gamma$ . Se  $\tilde{x}$  é uma solução de (7.2), então

$$S = \{x \in \mathbb{R}^n \mid x = \tilde{x} + Z\gamma, \gamma \in \mathbb{R}^{n-m}\}. \quad (7.3)$$

## 7.2 CONDIÇÕES NECESSÁRIAS DE PRIMEIRA ORDEM

A caracterização de  $S$  dada em (7.3) sugere a definição da seguinte função  $\varphi : \mathbb{R}^{n-m} \rightarrow \mathbb{R}$

$$\varphi(\gamma) = f(\tilde{x} + Z\gamma). \quad (7.4)$$

Consideremos o problema irrestrito

$$\text{Minimizar } \varphi(\gamma). \quad (7.5)$$

### Proposição 7.1

$\gamma^*$  é um minimizador local (global) de  $\varphi$  em  $\mathbb{R}^{n-m}$  se e somente se  $x^* \equiv \tilde{x} + Z\gamma^*$  é um minimizador local (global) de (7.1).

**Prova:** A deixamos como exercício para o leitor. ■

A condição necessária de primeira ordem para (7.5) é:

$$\nabla\varphi(\gamma^*) = 0. \quad (7.6)$$

Por (7.4),  $\varphi(\gamma) = f(g(\gamma))$ , onde  $g : \mathbb{R}^{n-m} \rightarrow \mathbb{R}^n$  está definida por  $g(\gamma) = \tilde{x} + Z\gamma$ . Logo, aplicando a regra da cadeia, obtemos

$$J_\varphi(\gamma) = J_f(g(\gamma)) J_g(\gamma) = \nabla^t f(g(\gamma)) Z.$$

Portanto,

$$\nabla\varphi(\gamma) = Z^t \nabla f(g(\gamma)). \quad (7.7)$$

Assim, da condição de primeira ordem (7.6), resulta que

$$\nabla\varphi(\gamma^*) = Z^t \nabla f(\tilde{x} + Z\gamma^*) = Z^t \nabla f(x^*) = 0.$$

Ou seja, uma condição necessária para que  $x^*$  seja minimizador local de (7.1) é que

$$Z^t \nabla f(x^*) = 0, \quad (7.8)$$

isto é, que  $\nabla f(x^*)$  seja ortogonal a  $Nu(A)$ . Ver Figura 7.2. Nesta figura, temos que  $\nabla f(x^*) \perp z^1$  e que  $\nabla f(x^*) = \begin{pmatrix} a_{11} \\ a_{12} \end{pmatrix} \lambda$ .

Pelas considerações feitas na seção anterior, temos que  $\nabla f(x^*) \in Im(A^t)$ , ou seja,  $\nabla f(x^*)$  deve ser uma combinação linear das linhas de  $A$ . Portanto, existe  $\lambda^* \in \mathbb{R}^m$  tal que

$$\nabla f(x^*) = A^t \lambda^*. \quad (7.9)$$

Claramente, (7.8) e (7.9) são equivalentes.

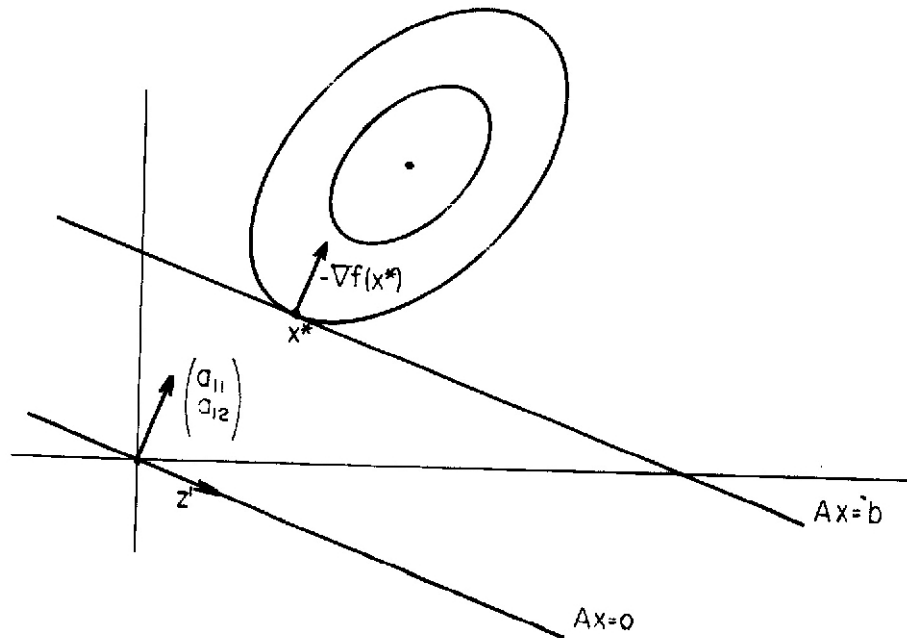


Figura 7.2

Observemos que se  $x^*$  é um minimizador local de (7.1), então, por (7.9), existe  $\lambda^* \in \mathbb{R}^m$  tal que  $(x^*, \lambda^*)$  é solução do seguinte sistema de  $(n + m)$  equações:

$$\begin{aligned}\nabla f(x^*) &= A^t \lambda^* \\ Ax^* &= b\end{aligned}\tag{7.10}$$

Toda solução de (7.1) é necessariamente solução de (7.10). A afirmação recíproca não é verdadeira. Com efeito, precisa-se informação de segunda ordem para dizer se uma solução de (7.10) é solução de (7.1).

O vetor  $\lambda^* \in \mathbb{R}^m$  é chamado vetor de multiplicadores de Lagrange associado a  $x^*$ .

### 7.3 CONDIÇÕES DE SEGUNDA ORDEM

A condição necessária de segunda ordem para uma solução de (7.5) é:

$$\nabla^2 \varphi(\gamma^*) \geq 0 \text{ (semidefinida positiva).}\tag{7.11}$$

Temos que  $\nabla \varphi(\gamma) = Z^t \nabla f(\tilde{x} + Z\gamma)$ , logo, aplicando a regra da cadeia, obtemos

$$\nabla^2 \varphi(\gamma) = Z^t \nabla^2 f(\tilde{x} + Z\gamma) Z.\tag{7.12}$$

Assim, a condição  $\nabla^2 \varphi(\gamma^*) \geq 0$  implica

$$Z^t \nabla^2 f(x^*) Z \geq 0.$$

Notemos que  $Z^t \nabla^2 f(x^*) Z$  é uma matriz de  $(n - m) \times (n - m)$ . O fato de ser semidefinida positiva significa que

$$y^t \nabla^2 f(x^*) y \geq 0 \text{ para todo } y \in Nu(A).$$

Analogamente, obtemos as seguintes condições suficientes de segunda ordem:

Se  $x^* \in \mathbb{R}^n$  verifica  $Ax^* = b$  e

- (i)  $Z^t \nabla f(x^*) = 0$
- (ii)  $Z^t \nabla^2 f(x^*) Z > 0$  (definida positiva),

então  $x^*$  é um minimizador local de (7.1).

### Exercícios

7.1 Os problemas abaixo consistem em minimizar  $f$  sujeita a  $Ax = b$  onde  $A \in \mathbb{R}^{m \times n}$  e  $b \in \mathbb{R}^m$ . Para cada um deles:

- (i) Encontre uma base de  $Nu(A)$ ;
- (ii) Construa uma parametrização que caracterize o conjunto factível;
- (iii) Transforme o problema em outro equivalente sem restrições;

(iv) Escreva as condições de primeira e segunda ordem para os dois problemas equivalentes.

(a) Minimizar  $x_1^2 + x_2^2 + x_3^2 - 2x_1x_2$  s.a.  $2x_1 + x_2 = 4$ ,  $5x_1 - x_3 = 8$ ;

(b) Minimizar  $x_1^2 + 2x_2^2 - 2x_1 - 2x_1x_2$  s.a.  $2x_1 + x_2 = 1$ .

7.2 Considere a função  $f(x, y) = xy$ .

(a) Analise os pontos estacionários do problema: Minimizar  $f(x, y)$  sem restrições;

(b) Acrescente a restrição  $x + y = 0$ . Analise as condições de otimalidade de primeira e segunda ordem;

(c) Resolva (b) para a restrição  $x - y = 0$ ;

(d) Analise (a), (b) e (c). Que conclusões podem ser tiradas?

7.3 Encontre o ponto sobre o plano  $x + 2y + 2z = 4$ , cuja distância à origem é mínima.

7.4 Seja  $f(x) = \|x\|$ ,  $x \in \mathbb{R}^n$ . Considere o problema de minimizar  $f$  sujeita a  $Ax = b$  com  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $m < n$  e posto  $A = m$ . Prove que a solução  $\tilde{x}$  desse problema pode ser escrita como  $\tilde{x} = \tilde{A}b$  onde  $\tilde{A} \in \mathbb{R}^{n \times m}$  e  $A\tilde{A} = I$ .

7.5 Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^2$ . Seja  $\tilde{x} \in \mathbb{R}^n$  tal que  $A\tilde{x} = b$  ( $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ) e tal que existe  $\lambda \in \mathbb{R}^m$  com  $\nabla f(\tilde{x}) = A^t\lambda$  e  $\nabla^2 f(\tilde{x})$  definida positiva. O ponto  $\tilde{x}$  é um minimizador local de  $f$  sujeita a  $Ax = b$ ? Prove ou dê um contra-exemplo.

7.6 Considere o problema

$$\text{Minimizar } \frac{1}{2}x^t Qx + p^t x + q$$

$$\text{s.a. } Ax = b,$$

onde  $Q \in \mathbb{R}^{n \times n}$  é simétrica,  $x, p \in \mathbb{R}^n$ ,  $q \in \mathbb{R}$ ,  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ . Seja  $Z$  uma base de  $Nu(A)$  e suponha que  $Z^t Q Z$  é definida positiva. Seja  $x^0$  tal que  $Ax^0 = b$ . Prove que a solução  $\tilde{x}$  é dada por

$$\tilde{x} = x^0 - Z(Z^t Q Z)^{-1} Z^t (Qx^0 + p).$$

7.7 Considere o problema

$$\text{Minimizar } f(x)$$

$$\text{s.a. } Ax = b,$$

---

onde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^1$ ,  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $m < n$  e posto  $A = m$ .  
Seja  $\tilde{p}$  a solução de

$$\text{Minimizar } \|\nabla f(x) - p\|$$

$$\text{s.a. } Ap = 0.$$

Encontre  $\tilde{p}$  e interprete geometricamente.

7.8 Dadas as variedades afins em  $\mathbb{R}^n$ ,  $S = \{x \in \mathbb{R}^n \mid Ax = b\}$  e  $U = \{x \in \mathbb{R}^n \mid Cx = d\}$ , onde  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $C \in \mathbb{R}^{p \times n}$  e  $d \in \mathbb{R}^p$ , considere o problema de encontrar o ponto de  $S$  mais próximo de  $U$ . Formule esse problema como um problema de otimização e escreva as condições de otimalidade.



## Capítulo 8

# ALGORITMOS PARA RESTRICÇÕES LINEARES DE IGUALDADE

### 8.1 MÉTODOS BÁSICOS DE DESCIDA

Seja  $x^k \in \mathbb{R}^n$  tal que  $Ax^k = b$  e  $Z^t \nabla f(x^k) \neq 0$ . Equivalentemente, para todo  $\lambda \in \mathbb{R}^m$

$$\nabla f(x^k) \neq A^t \lambda.$$

Ou seja,  $x^k$  não verifica as condições necessárias de primeira ordem (7.10). Desejamos determinar, a partir de  $x^k$ , um novo ponto factível  $x^{k+1}$  tal que

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

Sabemos que, se  $x^{k+1} = x^k + \alpha d$ , para manter a factibilidade ( $Ax^{k+1} = b$ ) é preciso que  $d \in \text{Nu}(A)$ .

Para garantir que, para algum  $\alpha > 0$ ,  $f(x^{k+1}) < f(x^k)$ , precisamos que  $d$  seja ademais uma direção de descida, ou seja

$$\nabla^t f(x^k) d < 0.$$

Então, precisamos encontrar  $d \in \text{Nu}(A)$  tal que

$$\nabla^t f(x^k) d < 0.$$

Se olharmos para o problema irrestrito associado em  $\mathbb{R}^{n-m}$  onde a função objetivo  $\varphi$  é dada por

$$\varphi(\gamma) = f(x^k + Z\gamma)$$

temos

$$\nabla \varphi(\gamma) = Z^t \nabla f(x^k + Z\gamma),$$

$$\varphi(0) = f(x^k),$$

e

$$\nabla\varphi(0) = Z^t\nabla f(x^k) \neq 0. \quad (8.1)$$

Então,  $\omega \equiv -\nabla\varphi(0) \equiv -Z^t\nabla f(x^k)$  é uma direção de descida para  $\varphi$  em  $\gamma = 0$ . Mas  $\omega \in \mathbb{R}^{n-m}$ , e queremos  $d \in \mathbb{R}^n$  e  $d \in Nu(A)$ .

Como  $d \in Nu(A)$  se e somente se  $d = Z\mu$  para algum  $\mu \in \mathbb{R}^{n-m}$  e  $\omega \in \mathbb{R}^{n-m}$  é uma direção de descida para  $\varphi$  em  $\gamma = 0$ , é bastante natural usar

$$d = Z\omega.$$

Com efeito, por (8.1),

$$\nabla^t f(x^k)d = \nabla^t f(x^k)Z\omega = \nabla^t\varphi(0)\omega < 0$$

Assim, resulta que

$$d = -ZZ^t\nabla f(x^k)$$

é uma direção factível de descida para  $f$  em  $x^k$ .

Agora estamos em condições de propor um algoritmo para o problema (7.1).

### Algoritmo 8.1

Seja  $\alpha \in (0, 1)$  dado. Seja  $x^k$  uma aproximação à solução de (7.1) tal que  $Ax^k = b$ . Seja  $Z \in \mathbb{R}^{n \times (n-m)}$  uma matriz cujas colunas formam uma base de  $Nu(A)$ .

Os passos para definir  $x^{k+1}$  são:

**Passo 1:** Se  $Z^t\nabla f(x^k) = 0$  parar. ( $x^k$  é um ponto estacionário). Caso contrário ir ao Passo 2.

**Passo 2:** Calcular  $d_k = -ZZ^t\nabla f(x^k)$ .

**Passo 3:** (Busca linear)

(i) Fazer  $\lambda = 1$ ;  
(ii) Se  $f(x^k + \lambda d_k) < f(x^k) + \alpha \lambda \nabla^t f(x^k)d_k$ , ir a (iv);

(iii) Escolher  $\bar{\lambda} \in [0.1\lambda, 0.9\lambda]$ . Fazer  $\lambda = \bar{\lambda}$ , ir a (ii);  
(iv) Fazer  $\lambda_k = \lambda$  e  $x^{k+1} = x^k + \lambda_k d_k$ .



Notemos que este processo exige a determinação de uma solução inicial factível e a determinação de uma base de  $Nu(A)$ .

Em geral, se  $\omega \in \mathbb{R}^{n-m}$  é uma direção de descida para  $\varphi(\gamma)$  em  $\gamma = 0$ , ( $\nabla^t \varphi(0)\omega < 0$ ), obtemos  $(Z^t \nabla f(x^k))^t \omega < 0$ , ou seja,  $\nabla^t f(x^k)Z\omega < 0$ .

Se escolhemos  $d = Z\omega$ , então  $d$  resulta uma direção de descida factível para  $f$  em  $x^k$ . Portanto, associado a cada método de descida para um problema irrestrito definido em  $\mathbb{R}^{n-m}$  temos um método de descida para o problema definido em  $\mathbb{R}^n$  com restrições lineares de igualdade. A cada iteração do método aplicado ao problema irrestrito em  $\mathbb{R}^{n-m}$  corresponde uma iteração do método associado para o problema em  $\mathbb{R}^n$  com restrições e reciprocamente.

Os resultados de convergência discutidos nos Capítulos 4 e 6 para métodos de descida aplicados a funções sem restrições são válidos para os métodos correspondentes para problemas com restrições lineares de igualdade.

Outro enfoque tentador para obter direções factíveis de descida é o seguinte: Se  $Z^t \nabla f(x^k) \neq 0$  (portanto  $\nabla f(x^k)$  não é ortogonal a  $Nu(A)$ ) podemos considerar a projeção de  $-\nabla f(x^k)$  sobre  $Nu(A)$  que denotamos  $P_{Nu(A)}(-\nabla f(x^k))$ .

Para todo  $v \in \mathbb{R}^n$

$$P_{Nu(A)}v = (I - A^t(AA^t)^{-1}A)v \quad (8.2)$$

ou

$$P_{Nu(A)}v = Z(Z^tZ)^{-1}Z^t v. \quad (8.3)$$

Observemos que, se a matriz  $Z$  é ortogonal ( $Z^tZ = I$ ), então

$$P_{Nu(A)}(-\nabla f(x^k)) = -ZZ^t \nabla f(x^k),$$

e a direção coincide com a obtida antes. Se  $Z$  não é ortogonal,  $d \equiv P_{Nu(A)}(-\nabla f(x^k))$  é factível e também é de descida. Esta direção define outro método conhecido com o nome de *método de gradiente projetado*.

A pergunta que segue é pertinente: Dada uma direção de descida  $d \in \mathbb{R}^n$  em  $x^k$  para o problema sem restrições, será que  $P_{Nu(A)}d$  é de descida para o problema com restrições?

A resposta é negativa como ilustra o exemplo na Figura 8.1. Nesta figura,  $d_N$  é a direção do método de Newton.

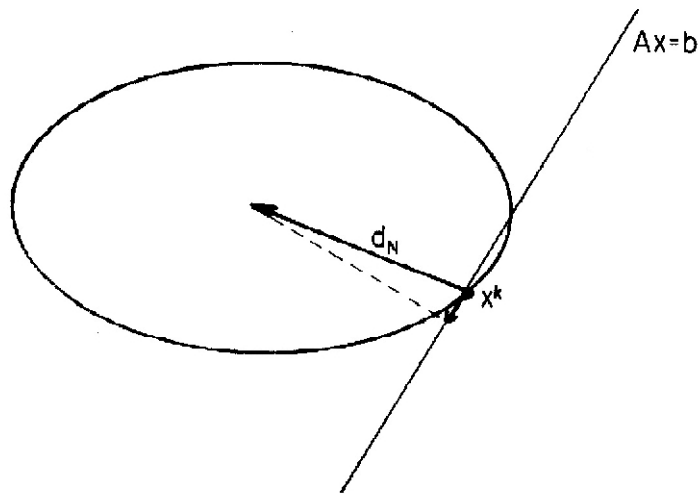


Figura 8.1

Outra forma de reduzir o problema (7.1) a um problema irrestrito definido num espaço de dimensão menor consiste em expressar algumas variáveis em função das outras. Como posto  $A = m$ , existem  $m$  colunas de  $A$  que são linearmente independentes. Chamamos  $B$  a submatriz de  $A$  formada por essas colunas e  $C$  a submatriz que resta uma vez retiradas as colunas de  $B$ . Reordenando as variáveis de forma conveniente, o problema (7.1) pode ser reescrito assim:

$$\begin{aligned} & \text{Minimizar } f(y, \omega) \\ & \text{sujeita a } By + C\omega = b \end{aligned} \quad (8.4)$$

com  $y \in \mathbb{R}^m$  e  $\omega \in \mathbb{R}^{n-m}$ .

Como  $B$  é não-singular temos que

$$y = B^{-1}b - B^{-1}C\omega$$

e o problema (7.1) é equivalente ao problema irrestrito

$$\text{Minimizar } \varphi(\omega) = f(B^{-1}b - B^{-1}C\omega, \omega), \quad \omega \in \mathbb{R}^{n-m}.$$

Aplicando a regra da cadeia, obtemos

$$\nabla\varphi(\omega) = \nabla_{\omega}f(y, \omega) - C^t(B^t)^{-1}\nabla_y f(y, \omega),$$

onde  $\nabla^t f(y, \omega) = (\nabla_y^t f(y, \omega), \nabla_{\omega}^t f(y, \omega))$ . Se  $\nabla\varphi(\omega) \neq 0$ , ao método do gradiente para este problema corresponde um método para o problema com restrições.

Esse método, chamado de *método de gradiente reduzido*, é na verdade um caso particular dos métodos discutidos acima, onde

$$Z = \begin{bmatrix} -B^{-1}C \\ I \end{bmatrix}$$

e  $I$  é a matriz identidade em  $\mathbb{R}^{(n-m) \times (n-m)}$ .

### Exercícios

8.1 Considere o problema de minimizar  $x^2 + 3y^2 + 2z^2$ , sujeita a  $x + 2y + 3z = 6$ .

Seja  $x^0 = (1, 1, 1)^t$ . Resolva o problema aplicando o método de Newton ao problema reduzido e verificando que  $x^1$  satisfaz as condições de otimalidade de primeira e segunda ordem.

8.2 Considere o problema quadrático

$$\text{Minimizar } \frac{1}{2}x^t Q x - c^t x$$

$$\text{s.a. } Ax = b,$$

onde  $Q \in \mathbb{R}^{n \times n}$  é simétrica,  $c \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{m \times n}$  e  $b \in \mathbb{R}^m$ . Prove que  $\tilde{x}$  é um minimizador local se e somente se  $\tilde{x}$  é um minimizador global. (Note que não há nenhuma hipótese sobre  $Q$ .)

8.3 Considere o problema de minimizar  $f$  sujeita a  $Ax = b$  com  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $m < n$  e posto  $A = m$ . Sejam  $\tilde{x} \in \mathbb{R}^n$  tal que  $A\tilde{x} = b$  e  $g = \nabla f(\tilde{x}) \neq 0$ . Seja  $d \in \mathbb{R}^n$  tal que  $\nabla^t f(\tilde{x})d < 0$ . Sejam  $\hat{g}$  e  $\hat{d}$  as projeções de  $g$  e  $d$  sobre  $Nu(A)$ , respectivamente. Considere as seguintes afirmações:

(a)  $\hat{d}^t \hat{g} < 0$ ;

(b) Existem  $\hat{d}$  e  $\hat{g}$  tais que  $\hat{d}^t \hat{g} \geq 0$ ;

Qual das duas afirmações é verdadeira? Prove ou dê um contra-exemplo.

8.4 Considere o seguinte problema

$$\text{Minimizar } x_1^2 + x_2^2$$

$$\text{s.a. } x_1 + x_2 = 1$$

(a) Encontre a solução ótima  $x^*$ ;

(b) Considere o *problema penalizado* Minimizar  $x_1^2 + x_2^2 + \mu(x_1 + x_2 - 1)^2$ . Para cada  $\mu > 0$ , calcule a solução ótima  $\tilde{x}(\mu)$ ;

(c) Verifique que  $\lim_{\mu \rightarrow \infty} \tilde{x}(\mu) = x^*$ ;

- (d) Repita (a), (b) e (c) trocando a função objetivo por  $x_1^3 + x_2^3$ ;  
 (e) Analise os resultados obtidos.

8.5 Seja  $z^1 = (1, -1, 2)^t$ . Escolha  $z^2 \in \mathbb{R}^3$  tal que  $z^1$  e  $z^2$  sejam linearmente independentes. Considere  $Z = [z^1 z^2]$  uma base de  $Nu(A)$  com  $A \in \mathbb{R}^{m \times n}$ .

- (a) Determine  $m$  e  $n$ ;  
 (b) Encontre  $A$ . É única?  
 (c) Ache as equações da variedade afim paralela a  $Nu(A)$  que passa pelo ponto  $(2, 5, 1)^t$ ;  
 (d) Se  $S$  é a variedade em (c) e  $\tilde{x}$  é a solução de minimizar  $f$  sujeita a  $x \in S$ , onde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , qual é a relação entre  $Z$  e  $f$  no ponto  $\tilde{x}$ ?

8.6 Considere o problema de minimizar  $f$  sujeita a  $Ax = b$  com  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^2$ ,  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ . Se  $\tilde{x} \in \mathbb{R}^n$  é uma solução desse problema então existe  $\tilde{\lambda} \in \mathbb{R}^m$  tal que  $\nabla f(\tilde{x}) + A^t \tilde{\lambda} = 0$ . Definimos

$$\begin{aligned} \text{a função lagrangeana:} & \quad L(x, \lambda) = f(x) + \lambda^t(Ax - b); \\ \text{a função dual:} & \quad \phi(\lambda) = \text{Minimizar}_x L(x, \lambda); \end{aligned}$$

para todo  $\lambda$  tal que  $\phi(\lambda)$  esteja bem definida, e

$$\text{o problema dual} \quad \text{Maximizar } \phi(\lambda)$$

- (a) Que tipo de ponto é  $(\tilde{x}, \tilde{\lambda})$  em relação a  $L(x, \lambda)$ ?  
 (b) Prove que  $\phi(\lambda) \leq f(x)$  para todo  $x$  tal que  $Ax = b$ ;  
 (c) Exiba o problema dual para  $f(x) = c^t x$ , onde  $c \in \mathbb{R}^n$ .

8.7 Considere o problema de minimizar  $\frac{1}{2}x^t(x - 2c)$  sujeita a  $Ax = b$ , onde  $c \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $m \leq n$  e posto  $A = m$ . Seja  $P$  a matriz de projeção sobre o Núcleo de  $A$ . Seja  $\tilde{x}$  uma solução do problema. Prove que  $P\tilde{x} = Pc$ . Interprete geometricamente em  $\mathbb{R}^2$ .

8.8 Considere o problema (P) Minimizar  $\frac{1}{2}x^t Bx + c^t x$  sujeita a  $Ax = b$ , onde  $\{x \in \mathbb{R}^n \mid Ax = b\}$  é não vazio e  $B$  é simétrica.

- (a) Prove que se (P) tem solução, então  $z^t Bz \geq 0$  para todo  $z \in Nu(A)$ ;  
 (b) Prove que (P) tem solução única se e somente se  $z^t Bz > 0$  para todo  $z \in Nu(A)$ ,  $z \neq 0$ ;  
 (c) Mostre com um exemplo que (a) é condição necessária de otimalidade mas não é suficiente.

---

8.9 Seja  $B$  uma matriz simétrica. Dizemos que  $B \geq 0$  em  $Nu(A)$  se  $z^t B z \geq 0$  para todo  $z \in Nu(A)$  e que  $B > 0$  em  $Nu(A)$  se  $z^t B z > 0$  para todo  $z \in Nu(A)$ ,  $z \neq 0$ .

- (a) Prove que se existe  $r \in \mathbb{R}$  tal que  $B + rA^t A > 0$ , então  $B > 0$  em  $Nu(A)$ ;
- (b) Prove que se existe  $r \in \mathbb{R}$  tal que  $B + rA^t A \geq 0$ , então  $B \geq 0$  em  $Nu(A)$ ;
- (c) Prove que se  $B > 0$  em  $Nu(A)$ , então existe  $r \in \mathbb{R}$  tal que  $B + rA^t A > 0$ ;
- (d) Através de um exemplo mostre que a recíproca de (b) não é verdadeira.

8.10 Relacione os exercícios 8.8 e 8.9 com a resolução do problema

$$\text{Minimizar } \frac{1}{2}x^t B x + c^t x + r\|Ax - b\|^2.$$

8.11 Considere o problema de minimizar  $\frac{1}{2}x^t L x$  sujeita a  $Ax = 0$ , onde  $L \in \mathbb{R}^{n \times n}$  simétrica,  $A \in \mathbb{R}^{m \times n}$ ,  $m < n$  e posto  $A = m$ .

- (a) Escreva as condições de otimalidade de primeira e segunda ordem;
- (b) Suponha que são válidas as condições suficientes em (a) e encontre a solução.



## Capítulo 9

# MINIMIZAÇÃO COM RESTRICÇÕES LINEARES DE DESIGUALDADE

Neste capítulo, consideramos o problema

Minimizar  $f(x)$

sujeita a  $Ax \leq b$ , (9.1)  
onde  $x \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{m \times n}$ .

### 9.1 A REGIÃO DE FACTIBILIDADE

Neste caso,  $S = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ . Denotamos  $a_i^t = (a_{i1}, a_{i2}, \dots, a_{in})$ , então

$$S = \{x \in \mathbb{R}^n \mid a_i^t x \leq b_i \text{ para todo } i \in \{1, 2, \dots, m\}\}.$$

Cada uma das  $m$  desigualdades

$$a_i^t x \leq b_i$$

define em  $\mathbb{R}^n$  um semi-espaço. O hiperplano divisor é  $a_i^t x = b_i$  e o semi-espaço definido é aquele que está do lado contrário à direção apontada pelo vetor  $a_i$ . Por exemplo, na Figura 9.1, onde  $n = 2$  e  $m = 1$ , temos que  $S = \{x \in \mathbb{R}^2 \mid a_1^t x \leq b_1\}$ .

No problema (9.1), a região  $S$  consiste na intersecção de  $m$  semi-espaços. Portanto,  $S$  é um poliedro em  $\mathbb{R}^n$ . Ver a Figura 9.2, onde  $n = 2$ ,  $m = 5$ .

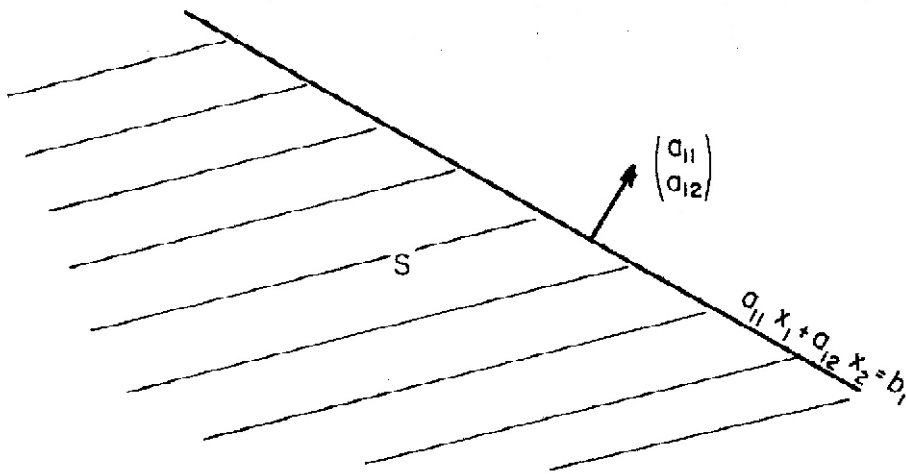


Figura 9.1

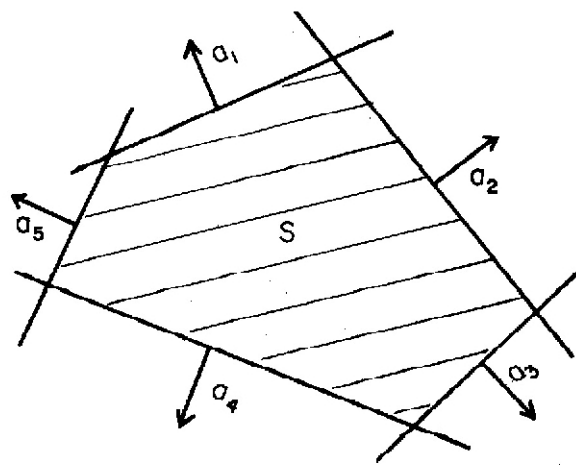


Figura 9.2

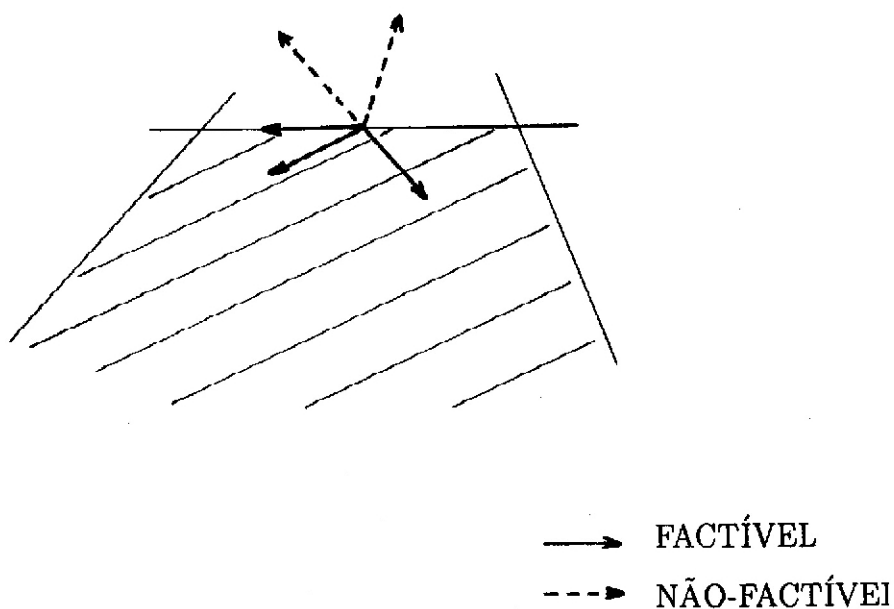


Interessa-nos caracterizar, dado um ponto  $x \in S$ , as direções factíveis a partir de  $x$ . Essas direções são aquelas nas quais há espaço para se movimentar dentro da região  $S$ .

Mais precisamente  $d \in \mathbb{R}^n$  é uma direção factível a partir de  $x \in S$  se e somente se

$$\text{Existe } \bar{\gamma} > 0 \text{ tal que } x + \gamma d \in S \text{ para todo } \gamma \in [0, \bar{\gamma}]. \quad (9.2)$$

Ver Figura 9.3.



**Figura 9.3**

A cada  $x \in S$  pode ser associado um número  $r(x)$  com  $0 \leq r(x) \leq m$ , que representa a quantidade de restrições para as quais

$$a_i^t x = b_i.$$

Diremos que essas restrições estão *ativas* em  $x$ . Ver Figura 9.4.

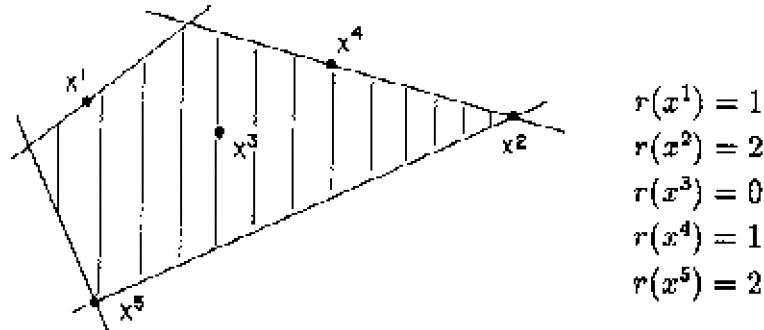


Figura 9.4

O conjunto de direções factíveis a partir de  $x$  depende das restrições ativas nesse ponto. Por exemplo, se  $r(x) = 0$ , qualquer  $d \in \mathbb{R}^n$  é factível.

Suponhamos que  $x \in S$  é tal que  $r(x) = p$  com  $0 < p \leq m$ .

Definimos  $\mathcal{I}(x) \subset \{1, 2, \dots, m\} \equiv \mathcal{M}$  por:

$$\mathcal{I}(x) = \{j \in \mathcal{M} \mid a_j^t x = b_j\}.$$

Dado  $d \in \mathbb{R}^n$  e  $\alpha > 0$ , temos que  $x + \alpha d \in S$  se e somente se  $A(x + \alpha d) \leq b$ , ou seja,  $a_j^t(x + \alpha d) \leq b_j$  para todo  $j \in \mathcal{M}$ .

Em particular se  $j \in \mathcal{I}(x)$  temos que  $a_j^t(x + \alpha d) = b_j + \alpha a_j^t d$ , portanto, para que  $b_j + \alpha a_j^t d \leq b_j$  necessariamente devemos ter  $a_j^t d \leq 0$ .

Vejamos que se  $a_j^t d \leq 0$  para todo  $j \in \mathcal{I}(x)$  então  $d$  é uma direção factível.

Se  $j \in \mathcal{I}(x)$  (portanto  $a_j^t d \leq 0$ ) temos que  $a_j^t(x + \alpha d) \leq b_j$  para todo  $\alpha \geq 0$ .

Se  $j \notin \mathcal{I}(x)$  (portanto  $a_j^t x < b_j$ ) temos que analisar as situações seguintes:

- (a) Se  $a_j^t d \leq 0$  resulta  $a_j^t(x + \alpha d) \leq b_j$ .
- (b) Se  $a_j^t d > 0$  podemos calcular o valor de  $\alpha$  tal que  $a_j^t(x + \alpha d) = b_j$ . Vemos que  $\alpha$  é dado por

$$\alpha = \frac{b_j - a_j^t x}{a_j^t d}.$$

Então, se definimos

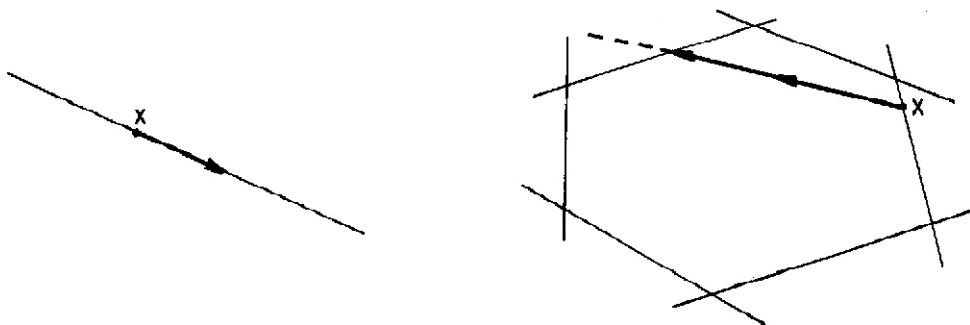
$$\tilde{\alpha} = \min_{\substack{j \in \mathcal{M} - \mathcal{I}(x) \\ a_j^t d > 0}} \left\{ \frac{b_j - a_j^t x}{a_j^t d} \right\},$$

teremos que  $a_j^t(x + \alpha d) \leq b_j$  para todo  $j \in \mathcal{M}$  e  $\alpha \in (0, \tilde{\alpha}]$  e, portanto,  $d$  será uma direção factível em  $x$ .

Acabamos de provar a seguinte afirmação:

$$d \in \mathbb{R}^n \text{ é factível em } x \text{ se e somente se } a_j^t d \leq 0 \text{ para todo } j \in \mathcal{I}(x). \quad (9.3)$$

Lembremos que no caso de restrições de igualdade, dada uma direção factível há total liberdade para se movimentar nessa direção. Isto pode não acontecer com restrições de desigualdade como mostra a Figura 9.5. Portanto, nos interessa saber quanto podemos andar sobre uma direção factível a partir de  $x$ .



**Figura 9.5**

Assim, dado  $x \in S$  e  $d$  uma direção factível em  $x$  precisamos determinar o maior valor de  $\alpha$  tal que  $a_j^t(x + \alpha d) \leq b_j$  para todo  $j \in \mathcal{M}$ , ou seja, o menor valor de  $\alpha$  para o qual  $a_j^t(x + \alpha d) = b_j$  para algum  $j \in \mathcal{M}$ .

Se  $j$  é tal que  $a_j^t d \leq 0$ ,  $\alpha$  pode ser arbitrariamente grande.

Se  $j$  é tal que  $a_j^t d > 0$  o valor procurado é

$$\tilde{\alpha} = \min_{\substack{j \in \mathcal{M} \\ a_j^t d > 0}} \left\{ \frac{b_j - a_j^t x}{a_j^t d} \right\}. \quad (9.4)$$

Observemos que  $a_j^t(x + \alpha d) \leq b_j$  para todo  $j \in \mathcal{M}$  e  $\alpha \in (0, \tilde{\alpha}]$ . Se  $\alpha > \tilde{\alpha}$ , existe  $j \in \mathcal{M}$  tal que  $a_j^t(x + \alpha d) > b_j$ .

## 9.2 CONDIÇÕES NECESSÁRIAS DE PRIMEIRA ORDEM

Agora que já temos uma caracterização das direções factíveis para qualquer ponto  $x \in S$ , estamos prontos para discutir as condições necessárias de otimalidade do problema (9.1).

Dado um ponto  $x \in S$ , queremos saber se existem direções de descida factíveis, ou seja, direções factíveis tais que

$$\nabla^t f(x)d < 0.$$

Se existe uma direção assim, o ponto  $x$  dado certamente não é um minimizador local de nosso problema. Mais uma vez, a análise dependerá das restrições ativas em  $x$ .

Se  $r(x) = 0$ , o ponto está no interior de  $S$  e as condições necessárias e suficientes são as que obtivemos para problemas sem restrições.

Suponhamos que  $r(x) \geq 1$ .

Para fixar idéias observemos algumas situações possíveis na Figura 9.6. Nessa figura tratamos de minimizar  $f$  sujeita a  $a_1^t x \leq b_1$ ,  $a_2^t x \leq b_2$ ,  $a_3^t x \leq b_3$ ,  $a_4^t x \leq b_4$ .

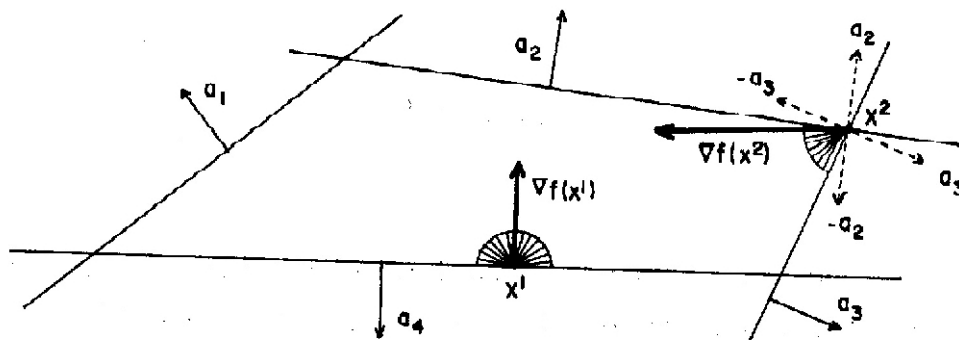


Figura 9.6

As direções factíveis em  $x^1$  e  $x^2$  estão na região hachurada.

Em  $x^1$  há uma única restrição ativa:  $a_4^t x^1 = b_4$ . Como  $\nabla f(x^1) = \lambda a_4$  com  $\lambda \leq 0$ , temos que  $\nabla^t f(x^1)d \geq 0$  para toda direção factível. Se tivéssemos  $\lambda > 0$ , existiria uma direção factível tal que  $\nabla^t f(x^1)d < 0$ . Portanto, encontramos uma condição necessária para que  $x^1$  seja um minimizador local.

Em  $x^2$  há duas restrições ativas, dadas por  $a_2^t x^2 = b_2$  e  $a_3^t x^2 = b_3$ . Como

$$\nabla f(x^2) = \lambda_1 a_2 + \lambda_2 a_3 \text{ com } \lambda_1 \leq 0 \text{ e } \lambda_2 \leq 0, \quad (9.5)$$

então  $\nabla^t f(x^2)d \geq 0$  para toda  $d$ , direção factível em  $x^2$ .

Em qualquer outro caso existe uma direção factível tal que

$$\nabla^t f(x^2)d < 0.$$

Portanto, (9.5) é condição necessária para que  $x^2$  seja um minimizador local.

Vamos generalizar essas idéias para  $\mathbb{R}^n$ .

### Teorema 9.1

Consideremos o problema (9.1) com  $f \in C^1$  e  $x^* \in S$  tal que  $1 \leq r(x^*) \leq n$ . Seja  $\mathcal{I} \subset \mathcal{M}$ ,  $\mathcal{I} = \{i_1, i_2, \dots, i_{r(x^*)}\}$  tal que  $a_j^t x = b_j$  se e somente se  $j \in \mathcal{I}$ . ( $\mathcal{I}$  é o conjunto dos índices que correspondem às restrições ativas em  $x^*$ ). Seja  $A_{\mathcal{I}} \in \mathbb{R}^{r(x^*) \times n}$  a submatriz de  $A$  cujas linhas são as que têm os índices em  $\mathcal{I}$

$$e \ b_{\mathcal{I}} = \begin{pmatrix} b_{i_1} \\ b_{i_2} \\ \vdots \\ b_{i_{r(x^*)}} \end{pmatrix}.$$

Supomos que  $\text{posto } A_{\mathcal{I}} = r(x^*)$ .

Se  $x^*$  é minimizador local de (9.1), então existe  $\lambda \in \mathbb{R}^{r(x^*)}$  tal que

$$\nabla f(x^*) = \sum_{k=1}^{r(x^*)} \lambda_k a_{i_k} \text{ e } \lambda_k \leq 0, \ 1 \leq k \leq r(x^*),$$

ou, equivalentemente

$$\nabla f(x^*) = A_{\mathcal{I}}^t \lambda, \ \lambda \in \mathbb{R}^{r(x^*)} \quad (9.6)$$

e

$$\lambda_k \leq 0, \quad 1 \leq k \leq r(x^*).$$

**Prova:** Suponhamos que (9.6) é falso. Isto pode acontecer por dois motivos:

(i)  $\nabla f(x^*) \neq A_{\mathcal{I}}^t \lambda$  para todo  $\lambda \in \mathbb{R}^{r(x^*)}$ .

Neste caso,  $x^*$  não é minimizador local do problema com restrições de igualdade definido por

Minimizar  $f(x)$

$$\text{sujeita a } A_{\mathcal{I}}x = b_{\mathcal{I}} \tag{9.7}$$

e, portanto,  $x^*$  tampouco pode ser minimizador local do problema (9.1).

(ii)  $\nabla f(x^*) = A_{\mathcal{I}}^t \lambda$  ( $\lambda \in \mathbb{R}^{r(x^*)}$ ) mas existe  $j$  tal que  $\lambda_j > 0$ .

Se  $r(x^*) = 1$  e  $\mathcal{I} = \{i_1\}$ , então  $\nabla f(x^*) = \lambda_1 a_{i_1}$  e  $\lambda_1 > 0$ . Se  $d = -\nabla f(x^*)$  temos  $a_{i_1}^t d = -\lambda_1 a_{i_1}^t a_{i_1} = -\lambda_1 \|a_{i_1}\|^2 < 0$ . Portanto,  $d$  é uma direção de descida factível.

Se  $2 \leq r(x^*) \leq n$ , denotamos por  $\tilde{A}_{\mathcal{I}}$  a matriz obtida retirando a linha  $a_{i_j}$  correspondente ao multiplicador  $\lambda_j > 0$ .

Consideramos  $d = P_{Nu(\tilde{A}_{\mathcal{I}})}(-\nabla f(x^*))$  onde  $P_{Nu(\tilde{A}_{\mathcal{I}})}$  é o operador projeção ortogonal sobre  $Nu(\tilde{A}_{\mathcal{I}})$ .

Então resulta

$$(-\nabla f(x^*) - d)^t d = 0$$

ou

$$\nabla^t f(x^*) d = -d^t d = -\|d\|^2 < 0, \tag{9.8}$$

o que mostra que  $d$  é uma direção de descida. Ver Figura 9.7.

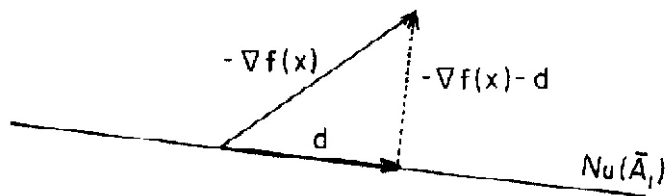


Figura 9.7

Agora,

$$\nabla f(x^*) = \lambda_1 a_{i_1} + \lambda_2 a_{i_2} + \cdots + \lambda_j a_{i_j} + \cdots + \lambda_{r(x^*)} a_{i_{r(x^*)}}$$

e, por construção,  $a_{i_k}^t d = 0$  para todo  $k \neq j$  ( $d \in Nu(\tilde{A}_I)$  e posto  $A_I = r(x^*)$ ).  
Portanto,

$$\nabla^t f(x^*) d = \lambda_j a_{i_j}^t d$$

e por (9.8) temos que  $\lambda_j a_{i_j}^t d < 0$ , que, junto com  $\lambda_j > 0$ , implica que  $a_{i_j}^t d < 0$ . Portanto,  $a_{i_k}^t d \leq 0$  para todo  $k$  tal que  $1 \leq k \leq r(x^*)$ , ou seja,  $d$  é uma direção factível e de descida.

Assim, o teorema fica demonstrado, já que sempre que a condição (9.6) não se verifica é possível construir uma direção de descida factível para  $x^*$ , contradizendo a hipótese de  $x^*$  ser minimizador local de (9.1). ■

Na Figura 9.8 ilustramos o teorema.

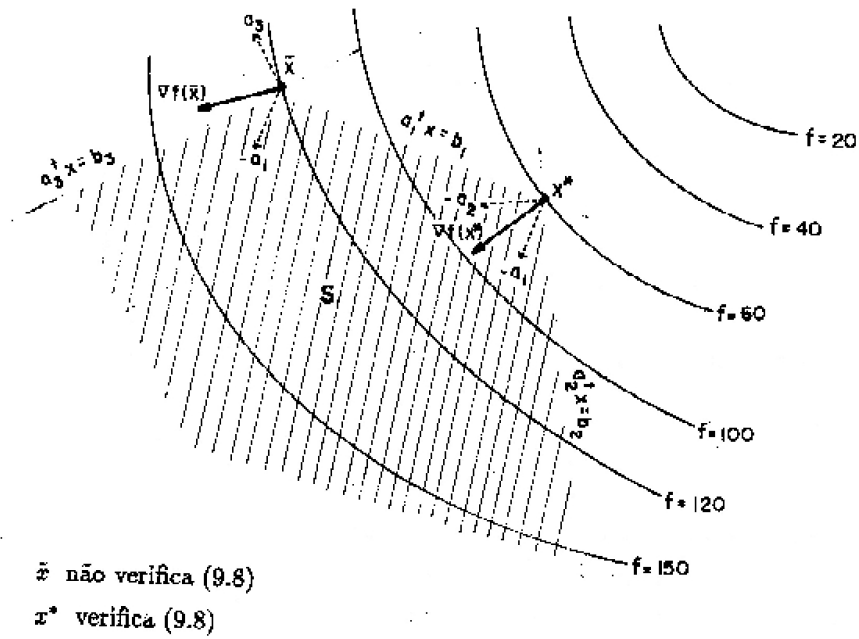


Figura 9.8

A condição necessária de que fala o Teorema 9.1 não é suficiente. Isso é evidenciado pela Figura 9.9.

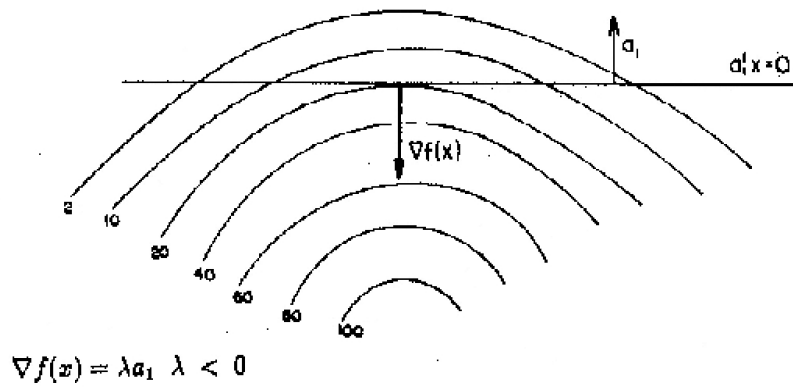


Figura 9.9



### 9.3 CONDIÇÕES DE SEGUNDA ORDEM

#### Teorema 9.2

Sejam  $f \in C^2$ ,  $x^*$  um minimizador local do problema (9.1), e  $r(x^*)$  e  $\mathcal{I}$  definidos como anteriormente, então

(i) Existe  $\lambda \in \mathbb{R}^{r(x^*)}$  tal que  $\nabla f(x^*) = A_{\mathcal{I}}^t \lambda$  e  $\lambda_i \leq 0$  para todo  $i \in \{1, 2, \dots, r(x^*)\}$ ;

(ii) Para todo  $y \in \text{Nu}(A_{\mathcal{I}})$  temos que  $y^t \nabla^2 f(x^*) y \geq 0$ .

#### Teorema 9.3

Sejam  $f \in C^2$ ,  $x^* \in S$ , e  $r(x^*)$  e  $\mathcal{I}$  definidos como acima. Se  $\nabla f(x^*) = A_{\mathcal{I}}^t \lambda$  com  $\lambda_i \leq 0$  para todo  $i \in \{1, 2, \dots, r(x^*)\}$  e  $y^t \nabla^2 f(x^*) y > 0$  para todo  $y \in \text{Nu}(A_{\mathcal{J}})$ ,  $y \neq 0$ , onde  $\mathcal{J} = \{i \in \{1, \dots, r(x^*)\} \mid \lambda_i < 0\}$ , então  $x^*$  é um minimizador local de (9.1).

As provas dos Teoremas 9.2 e 9.3 podem ser obtidas como casos particulares dos resultados provados em Luenberger [11].

#### Exercícios

9.1 Considere o seguinte problema

$$\text{Maximizar } 2x_1 + 3x_2$$

$$\text{s.a. } x_1 + x_2 \leq 8, \quad -x_1 + 2x_2 \leq 4, \quad x_1, x_2 \geq 0$$

(a) Escreva as condições de otimalidade;

(b) Para cada ponto extremo verifique se as condições de otimalidade são satisfeitas. Encontre a solução ótima.

9.2 Considere o problema (P):

$$\text{Minimizar } f(x)$$

$$\text{s.a. } Ax \leq b,$$

onde  $A \in \mathbb{R}^{m \times n}$ ,  $m < n$ ,  $b \in \mathbb{R}^m$  e considere também o sistema não-linear (S):

$$\nabla f(x) + A^t \mu = 0$$

$$(a_i^t x - b_i) \mu_i = 0, \quad i = 1, \dots, m,$$

onde  $A^t = [a_1 \dots a_m]$ . Qual é a relação entre as soluções de (P) e (S)?

9.3 Resolva o problema de otimização

Minimizar  $f(x, y)$

s.a.  $0 \leq x \leq 1, 0 \leq y \leq 1$

com  $f(x, y) = g(x) - x^2 + y^2$ , onde  $g(x)$  é o valor ótimo da função objetivo do seguinte problema

Minimizar  $u^2 + v^2$

s.a.  $u + 2v \geq x,$

$u, v \geq 0.$

9.4 Considere o seguinte problema canalizado:

Minimizar  $f(x)$

s.a.  $a_i \leq x_i \leq b_i, i = 1, \dots, m.$

Seja  $x$  um ponto factível e  $g = \nabla f(x)$ . Seja a direção  $d$  definida por

$$d_i = \begin{cases} 0 & \text{se } (x_i = a_i \text{ e } g_i \geq 0) \text{ ou } (x_i = b_i \text{ e } g_i \leq 0) \\ -g_i, & \text{caso contrário} \end{cases}$$

- (a) Prove que  $d$  é uma direção factível e de descida em  $x$ ;  
 (b) Prove que  $d = 0$  se e somente se  $x$  satisfaz as condições de otimalidade de primeira ordem;  
 (c) Usando essa direção e  $x^0 = (0, 3)^t$  ache a solução do seguinte problema:

Minimizar  $x^2 + y^2$

s.a.  $0 \leq x \leq 4, 1 \leq y \leq 3.$

9.5 Considere o seguinte problema:

Minimizar  $f(x)$

s.a.  $a_1^t x \leq b_1, a_2^t x \leq b_2.$

Suponha que as duas restrições são ativas em  $\tilde{x}$  e que  $\nabla f(\tilde{x})$  é combinação linear positiva de  $a_1$  e  $a_2$ . Construa duas direções factíveis e de descida diferentes em  $\tilde{x}$ . Justifique!

9.6 Considere os problemas primal e dual de programação linear:

Minimizar  $c^t x$   
 s.a.  $Ax = b$   
 $x \geq 0$

Maximizar  $b^t y$   
 s.a.  $A^t y \leq c$

Seja  $\tilde{x}$  solução do primal.

- (a) Prove que  $b^t y \leq c^t x$  para quaisquer  $x$  e  $y$  factíveis;  
 (b) Prove que o vetor dos multiplicadores de Lagrange  $\tilde{\lambda}$  associado às restrições de igualdade em  $\tilde{x}$  é solução ótima do dual;  
 (c) Prove que  $c^t \tilde{x} = b^t \tilde{\lambda}$ .

9.7 Considere o problema de programação quadrática

$$\begin{aligned} \text{Minimizar } f(x) &= \frac{1}{2} x^t B x + c^t x \\ \text{s.a. } Ax &= b \\ x &\geq 0. \end{aligned}$$

Seja  $\tilde{x}$  uma solução regular do problema, e  $\tilde{\lambda}$  o vetor de multiplicadores de Lagrange associado às restrições de igualdade. Prove que

$$f(\tilde{x}) = \frac{1}{2} (c^t \tilde{x} + b^t \tilde{\lambda}).$$

9.8 Resolva o seguinte problema de otimização

$$\begin{aligned} \text{Maximizar } P(x) &= x_1 x_2 \dots x_n \\ \text{s.a. } x_1 + x_2 + \dots + x_n &= c, \\ x &\geq 0. \end{aligned}$$

Deduz a seguinte desigualdade entre as médias aritmética e geométrica:

$$\frac{1}{n} \sum_{i=1}^n x_i \geq \left( \prod_{i=1}^n x_i \right)^{1/n}.$$

9.9 Suponha que  $S \equiv \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$  é não-vazio, onde  $A \in \mathbb{R}^{m \times n}$  e  $b \in \mathbb{R}^m$ . Seja  $0 \leq z \in \mathbb{R}^n$  tal que  $A^t(Az - b) = \gamma \geq 0$  e  $z^t \gamma = 0$ . Prove que  $Az = b$ .



## Capítulo 10

# MÉTODO DE RESTRIÇÕES ATIVAS

Neste capítulo descrevemos um modelo de algoritmo para resolver problemas de minimização com restrições lineares de desigualdade.

A idéia básica é a seguinte: dado um ponto  $x^k \in S$ , definimos um subproblema de minimização com restrições de igualdade determinadas pelas restrições ativas em  $x^k$ . Se  $x^k$  não for ótimo para este subproblema, continuamos tentando resolver o subproblema escolhendo uma direção factível de descida e fazendo uma busca linear. Ao dar este passo existe a possibilidade de acrescentar uma ou mais restrições. Se isto acontecer o subproblema muda e continuamos trabalhando com um subproblema novo. Se  $x^k$  for o ótimo do subproblema (geometricamente,  $x^k$  é o minimizador na face do poliedro determinada pelas restrições ativas em  $x^k$ ), testamos se  $x^k$  é solução ótima do problema. Se não for, escolhemos uma nova direção de descida factível e fazemos uma busca linear para determinar  $x^{k+1}$ . Este movimento nos faz abandonar a face que contém  $x^k$ , e podemos ter certeza que não voltaremos mais a esta face. Também, neste deslocamento mudamos de subproblema e o processo descrito se repete. Como o poliedro tem um número finito de faces que vão sendo descartadas, pode-se provar que este processo é finito.

O seguinte algoritmo formaliza a descrição do método.

### Algoritmo 10.1 (Método de restrições ativas)

*Dado  $x^k \in S$ , executar os seguintes passos.*

**Passo 1:** Determinar  $\mathcal{I}_k \equiv \mathcal{I}(x^k)$  e  $r(x^k)$ .

*Se  $\mathcal{I}_k = \emptyset$  e  $\nabla f(x^k) = 0$ , parar. ( $x^k$  é um ponto estacionário).*

*Se  $\mathcal{I}_k = \emptyset$  e  $\nabla f(x^k) \neq 0$ , ir ao Passo 7.*

*Se  $\mathcal{I}_k \neq \emptyset$ , ir ao Passo 2.*

**Passo 2:** Resolver o seguinte sistema linear

$$\nabla f(x^k) = A_{\mathcal{I}_k}^t \lambda.$$

Se o sistema não admite solução, (ou seja,  $x^k$  não é ponto estacionário do subproblema  $\langle \min f(x)$  sujeita a  $A_{\mathcal{I}_k} x = b_{\mathcal{I}_k} \rangle$ ), ir ao Passo 4.

Se o sistema tem solução ir ao Passo 3.

**Passo 3:** Se  $\lambda_i \leq 0$  para  $1 \leq i \leq r(x^k)$ , parar. ( $x^k$  é um ponto estacionário). Se  $\lambda_j > 0$  para algum  $j$  ir ao Passo 7.

**Passo 4:** Achar  $d_k \in \text{Nu}(A_{\mathcal{I}_k})$  tal que  $\nabla^t f(x^k) d_k < 0$ .

**Passo 5:** Determinar

$$\bar{\alpha} = \min_{a_j^t d_k > 0} \left\{ \frac{b_j - a_j^t x^k}{a_j^t d_k} \right\}.$$

**Passo 6:** Realizar uma busca linear na direção  $d_k$  para obter um tamanho do passo  $\alpha_k \in (0, \bar{\alpha}]$  que garanta descenso suficiente.

Se  $\alpha_k < \bar{\alpha}$ , fazer  $x^{k+1} = x^k + \alpha_k d_k$ ,  $k = k + 1$  e ir ao Passo 2.

Se  $\alpha_k = \bar{\alpha}$ , fazer  $x^{k+1} = x^k + \alpha_k d_k$ ,  $k = k + 1$  e ir ao Passo 1.

**Passo 7:** Escolher uma direção factível e de descida  $d_k$  em  $x^k$ .

**Passo 8:** Igual ao Passo 5.

**Passo 9:** Realizar busca linear em  $(0, \bar{\alpha}]$  garantindo descenso suficiente. Fazer  $x^{k+1} = x^k + \alpha_k d_k$ ,  $k = k + 1$  e ir ao Passo 1.

Lembramos que pontos estacionários são aqueles que satisfazem as condições necessárias de otimalidade de primeira ordem.

A eficiência de um método particular de restrições ativas depende em grande parte dos métodos utilizados para resolver os subproblemas, que são métodos para resolver problemas com restrições de igualdade. Obviamente, se não dispomos de um método finito para os subproblemas há o risco de permanecer indefinidamente numa face não ótima do poliedro.

Para certas funções (as quadráticas) conhecemos métodos finitos e este esquema é viável. Contudo, também pode acontecer que sejam necessárias

“demasiadas” iterações para chegar na face ótima. Naturalmente, o desejável é que este processo de identificação das restrições corretas seja rápido.

Estas observações sugerem que a construção de algoritmos eficientes para este tipo de problema não é uma tarefa simples. Em Fletcher [5] e Gill et al. [7] podem ser encontradas descrições e discussões de alguns métodos deste tipo.

## Exercícios

10.1 Resolva graficamente o problema

$$\text{Minimizar } x^2 - xy + y^2 - 3x$$

$$\text{s.a. } x + y \leq 4, x, y \geq 0$$

usando um método de restrições ativas a partir do ponto  $x^0 = (0, 0)^t$ .

10.2 Considere o problema de maximizar  $f(x, y) = xy$  sujeita a  $x + y \geq 1$  e  $x + 2y \leq 2$ . Aplique um método de restrições ativas, algébrica e geometricamente, a partir de  $(a)(1, 0)^t$  e  $(b)(2, 0)^t$ , até encontrar a solução.

10.3 Resolva algébrica ou graficamente o problema abaixo por um método de restrições ativas, tomando como ponto inicial  $(2, 1)^t$  e *justificando* todos os passos.

$$\text{Minimizar } (x + 1)^2 + (y - 1)^2$$

$$\text{s.a. } x + y \geq 1, x + y \leq 3, x, y \geq 0.$$

10.4 Aplique um método de restrições ativas para resolver

$$\text{Minimizar } x^2 + xy + 2y^2 - 6x - 2y - 12z$$

$$\text{s.a. } x + y + z = 2, -x + 2y \leq 3, x, y, z \geq 0.$$





## Capítulo 11

# MINIMIZAÇÃO COM RESTRIÇÕES LINEARES DE IGUALDADE E DESIGUALDADE

### 11.1 CONDIÇÕES NECESSÁRIAS DE PRIMEIRA ORDEM

O caso mais geral do problema de minimização de funções sujeitas a restrições lineares pode ser expressado como

Minimizar  $f(x)$

$$\text{sujeita a } Ax = b, Wx \leq c, \quad (11.1)$$

onde  $A \in \mathbb{R}^{m \times n}$  com  $m < n$  e posto de  $A = m$ ,  $W \in \mathbb{R}^{p \times n}$ ,  $b \in \mathbb{R}^m$  e  $c \in \mathbb{R}^p$ .

O conjunto de factibilidade  $S$  é um poliedro em  $\mathbb{R}^n$ .  $S \equiv \{x \in \mathbb{R}^n \mid Ax = b \text{ e } Wx \leq c\}$ .

As restrições correspondentes às linhas de  $A$  estão sempre ativas. Então, dado um ponto factível  $x$ , o conjunto dos índices das restrições ativas em  $x$  é

$$\mathcal{I}(x) = \{1, 2, \dots, m, i_1, i_2, \dots, i_{s(x)}\},$$

onde  $\mathcal{J}(x) \equiv \{i_1, i_2, i_3, \dots, i_{s(x)}\}$  é o conjunto de índices que correspondem às restrições (linhas de  $W$ ) que estão ativas em  $x$ . Temos que  $0 \leq s(x) \leq p$ . Se  $r(x)$  é o número total de restrições ativas em  $x$ , temos que

$$m \leq r(x) \leq m + p.$$

Repetindo os argumentos usados nos Capítulos 7 e 9 para caracterizar o conjunto de direções factíveis a partir de um ponto factível  $x$ , é fácil provar que neste caso  $d \in \mathbb{R}^n$  é factível em  $x$  se e somente se  $Ad = 0$  e  $w_j^t d \leq 0$  para todo  $j \in \mathcal{J}(x)$ .

As condições necessárias de otimalidade de primeira ordem que obtemos neste caso são uma generalização das condições (7.9) e (9.6).

**Teorema 11.1**

Consideremos o problema (11.1) com  $f \in C^1$  e  $x^* \in S$  tal que  $m \leq r(x^*) \leq n$  e  $s(x^*) \geq 1$ . Sejam  $\mathcal{I} = \{1, 2, \dots, m, i_1, i_2, \dots, i_{s(x^*)}\}$ ,  $\mathcal{J} = \{i_1, i_2, \dots, i_{s(x^*)}\}$  tal que  $w_j^t d = c_j$  se e somente se  $j \in \mathcal{J}$ ,  $W_{\mathcal{J}}$  a submatriz de  $W$ , cujas linhas são as que têm os índices em  $\mathcal{J}$ , e  $c_{\mathcal{J}} \in \mathbb{R}^{s(x^*)}$  formado pelas componentes de  $c$  correspondentes a  $\mathcal{J}$ .

Seja  $B \in \mathbb{R}^{[m+s(x^*)] \times n}$  dada por

$$B = \begin{pmatrix} A \\ W_{\mathcal{J}} \end{pmatrix} \text{ e posto } B = r(x^*).$$

Se  $x^*$  é minimizador local de (11.1), então existem  $\lambda \in \mathbb{R}^m$  e  $\mu \in \mathbb{R}^{s(x^*)}$  tais que

$$\nabla f(x^*) = A^t \lambda + W_{\mathcal{J}}^t \mu$$

e

$$(11.2)$$

$$\mu_k \leq 0 \text{ para todo } k \text{ tal que } 1 \leq k \leq s(x^*).$$

**Prova:** Os argumentos são os mesmos que usamos para provar (9.6). Deixamos esta prova para o leitor. ■

As condições (11.2) também são chamadas condições Kuhn-Tucker.

**11.2 CONDIÇÕES DE SEGUNDA ORDEM**

**Teorema 11.2**

Sejam  $f \in C^2$ ,  $x^*$  um minimizador local do problema (11.1),  $r(x^*)$ ,  $s(x^*)$ ,  $\mathcal{J}$  e  $B$  definidos como acima, então

(i) Existem  $\lambda \in \mathbb{R}^m$  e  $\mu \in \mathbb{R}^{s(x^*)}$  tais que

$$\nabla f(x^*) = A^t \lambda + W_{\mathcal{J}}^t \mu, \quad \mu_k \leq 0 \text{ para todo } k \in \{1, 2, \dots, s(x^*)\};$$

(ii)  $y^t \nabla^2 f(x^*) y \geq 0$  para todo  $y \in \text{Nu}(B)$ .

**Teorema 11.3**

Sejam  $f \in C^2$ ,  $x^* \in S$ ,  $r(x^*)$ ,  $s(x^*)$  e  $\mathcal{J}$  como acima, então se  $x^*$  verifica

(i) Existem  $\lambda \in \mathbb{R}^m$  e  $\mu \in \mathbb{R}^{s(x^*)}$  tais que

$$\nabla f(x^*) = A^t \lambda + W_{\mathcal{J}}^t \mu$$

e

$$\mu_k \leq 0 \text{ para todo } k \in \{1, 2, \dots, s(x^*)\};$$

(ii) Se  $y^t \nabla^2 f(x^*) y > 0$  para todo  $y \in \text{Nu}(\tilde{B})$ , onde

$$\tilde{B} = \begin{pmatrix} A \\ W_{\mathcal{K}} \end{pmatrix}$$

e

$$\mathcal{K} = \{j \in \mathcal{J} \mid \mu_j < 0\},$$

então  $x^*$  é um minimizador local de (11.1).

Os Teoremas 11.2 e 11.3 são casos particulares das condições de otimalidade provadas em Luenberger [11].

### Exercícios

11.1 Considere o problema

$$\text{Minimizar } \sum_{j=1}^n f_j(x_j)$$

$$\text{s.a. } e^t x = 1, \quad x \geq 0,$$

com  $f_j : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f_j \in C^1$ ,  $j = 1, \dots, n$  e  $e = (1, \dots, 1)^t$ . Prove que se  $\tilde{x}$  é a solução do problema acima, então existe  $\alpha \in \mathbb{R}$  tal que  $f'_j(\tilde{x}_j) = \alpha$  se  $\tilde{x}_j > 0$  e  $f'_j(\tilde{x}_j) \geq \alpha$  se  $\tilde{x}_j = 0$ .

11.2 Considere o problema de programação quadrática

$$\text{Minimizar } \frac{1}{2} x^t H x + c^t x$$

$$\text{s.a. } Ax \leq b,$$

onde  $H \in \mathbb{R}^{n \times n}$  é simétrica,  $c \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{m \times n}$  e  $b \in \mathbb{R}^m$ .

- (a) Escreva as condições de otimalidade de segunda ordem;
- (b) Para  $H = I$  e  $c = 0$ , interprete esse problema geometricamente.



## Capítulo 12

# MINIMIZAÇÃO COM RESTRICÇÕES NÃO-LINEARES DE IGUALDADE

Consideraremos problemas da forma

Minimizar  $f(x)$

sujeita a  $h(x) = 0$ , (12.1)

onde  $f, h \in C^1$ ,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$  e  $m < n$ . Permitiremos a partir de agora que a função  $h$  seja não-linear. Os resultados que apresentaremos são extensões dos que existem para problemas com restrições lineares.

### 12.1 A REGIÃO DE FACTIBILIDADE

No caso de restrições lineares o estudo do conjunto de factibilidade e das direcções factíveis é feito utilizando exclusivamente os conceitos da álgebra linear. A presença de restrições não-lineares exige o uso de conceitos algo mais complicados.

Na Figura 12.1 ilustramos o caso em que há uma única restrição de igualdade em  $\mathbb{R}^2$ . Neste caso, a região factível  $S \equiv \{x \in \mathbb{R}^2 \mid h(x) = 0\}$  é uma curva.



Figura 12.1

Notemos que não há direções factíveis a partir de um ponto  $x \in S$ , ao contrário do que acontecia quando as restrições eram lineares. Para permanecer em  $S$ , os movimentos a partir de  $x \in S$  devem ser curvilíneos. É claro que, dado um ponto  $x \in S$ , a dificuldade em determinar outro ponto  $\tilde{x} \in S$  depende da função  $h$ .

Na Figura 12.2 ilustramos o caso em que há uma restrição de igualdade em  $\mathbb{R}^3$ . A região  $S \equiv \{x \in \mathbb{R}^2 \mid h(x) = 0\}$  é uma superfície em  $\mathbb{R}^3$ .

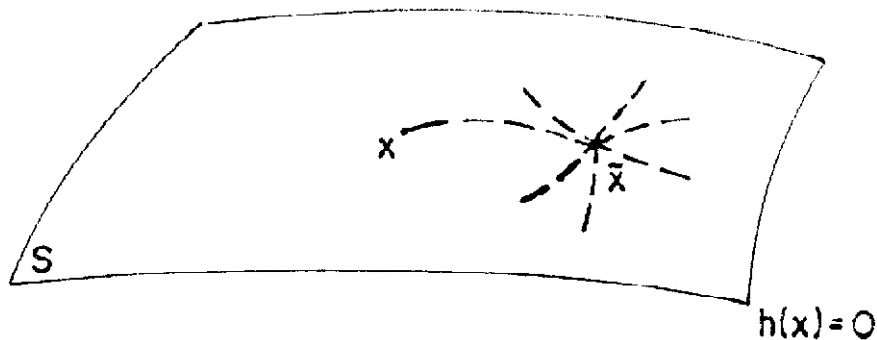


Figura 12.2

Dado o ponto  $\tilde{x} \in S$ , observamos que há infinitos caminhos que passam por  $\tilde{x}$ , contidos em  $S$ , todos eles curvilíneos. Podem-se unir dois pontos  $x$  e  $\tilde{x}$  por arcos de curva contidos em  $S$ , que chamamos arcos factíveis. Em geral,  $m$  equações (não-lineares) em  $\mathbb{R}^n$ ,  $m < n$ , determinam uma “superfície” de dimensão  $n - m$ .

Na Figura 12.2 vemos que por um ponto  $\tilde{x} \in S$  passa uma família de curvas contidas em  $S$ . Cada curva é uma “superfície” de dimensão 1 e sob certas condições é possível expressar esta curva, numa vizinhança do ponto  $\tilde{x}$ , mediante uma parametrização contínua  $x : \mathbb{R} \rightarrow S$ , tal que  $x(t) \in S$  para todo  $t \in (a, b)$  e, ademais, existe  $\tilde{t} \in (a, b)$  tal que  $x(\tilde{t}) = \tilde{x}$ .

O arco  $x(t)$ , assim definido, é diferenciável se  $x'(t)$  existe para todo  $t \in (a, b)$ .

Por exemplo, para  $x : \mathbb{R} \rightarrow \mathbb{R}^2$ ,

$$x(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}$$

e

$$x'(t) = \begin{pmatrix} x'_1(t) \\ x'_2(t) \end{pmatrix}.$$

O vetor  $x'(t)$  é tangente ao arco de curva no ponto  $\tilde{x}$ , portanto, tangente à superfície.

Se considerarmos todas as curvas diferenciáveis que passam por  $\tilde{x}$ , intuitivamente vemos que seus vetores tangentes em  $\tilde{x}$  definem um “plano tangente”  $T$ , como ilustra a Figura 12.3.

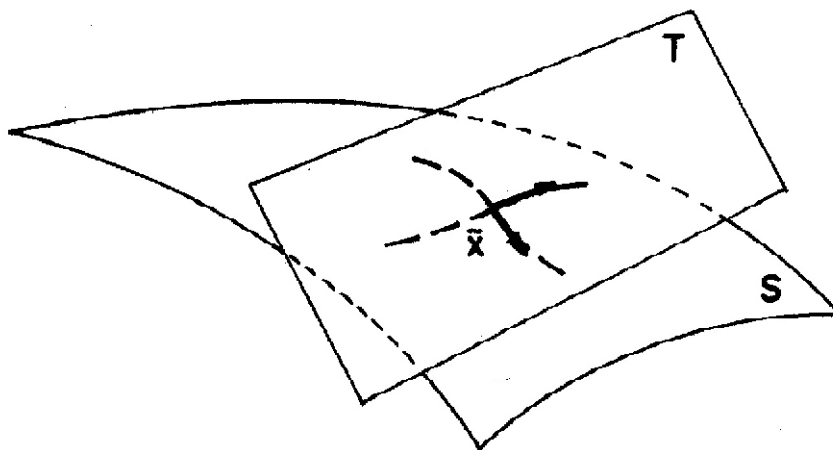


Figura 12.3

Se a superfície  $S$  for  $(n - m)$ -dimensional em  $\mathbb{R}^n$ , generalizando essas idéias, observamos que o plano tangente será gerado por vetores da forma

$$x'(t) = (x'_1(t), x'_2(t), \dots, x'_n(t))^t.$$

A dimensão do “plano tangente” também é  $(n - m)$ .

### Definição 12.1

O plano tangente a uma superfície  $S \subseteq \mathbb{R}^n$ , em um ponto  $\tilde{x} \in \mathbb{R}^n$ , é o conjunto de vetores de  $\mathbb{R}^n$ , que são tangentes em  $\tilde{x}$  a alguma curva diferenciável contida em  $S$  e que passa por  $\tilde{x}$ .

Temos agora alguns elementos para tentar caracterizar os arcos factíveis

que passam por um ponto factível  $\tilde{x}$ . Se

$$S = \{x \in \mathbb{R}^n \mid h(x) = 0\}$$

e  $x(t) : (a, b) \rightarrow S$  é a parametrização de um arco factível, temos que

$$h(x(t)) = 0 \text{ para todo } t \in (a, b),$$

Derivando a equação acima em relação a  $t$ , temos

$$J_h(x(t))x'(t) = 0 \text{ para todo } t \in (a, b), \quad (12.2)$$

ou seja,  $\nabla^t h_i(x(t))x'(t) = 0$  para todo  $t \in (a, b)$  e  $1 \leq i \leq m$ . (12.3)

Em particular, para  $x(\tilde{t}) = \tilde{x}$ , obtemos

$$\nabla^t h_i(\tilde{x})x'(\tilde{t}) = 0, \quad 1 \leq i \leq m,$$

o que significa que dado um arco factível diferenciável, é necessário que o vetor tangente ao arco em  $\tilde{x}$  seja ortogonal aos gradientes das restrições avaliados em  $\tilde{x}$ .

Notemos que (12.2) é uma extensão da caracterização das direcções factíveis obtida para restrições lineares no Capítulo 7.

Para que (12.2) seja uma caracterização dos arcos factíveis diferenciáveis, precisamos que para todo  $p \in \mathbb{R}^n$  tal que  $J_h(\tilde{x})p = 0$  exista um arco factível diferenciável  $z(t)$  tal que  $z(\tilde{t}) = \tilde{x}$  e  $z'(\tilde{t}) = p$ .

Infelizmente, isso nem sempre acontece, como mostra o seguinte exemplo devido a Kuhn e Tucker.

$$h_1(x) = (1 - x_1)^3 - x_2, \quad h_2(x) = x_2$$

$$S = \{x \in \mathbb{R}^2 \mid h_1(x) = 0 \text{ e } h_2(x) = 0\} = \{(1, 0)^t\}.$$

Como  $S$  consiste num único ponto, não existem arcos factíveis, mas

$$\nabla^t h_1(\tilde{x}) = (0, -1), \quad \nabla^t h_2(\tilde{x}) = (0, 1)$$

e, portanto, todo vetor  $p \in \mathbb{R}^2$  da forma  $(\gamma, 0)^t$ ,  $\gamma \in \mathbb{R}$  verifica

$$J_h(\tilde{x})p = 0.$$

Assim, caracterizar o conjunto de arcos factíveis diferenciáveis através do plano tangente mediante a equação (12.2) não é possível sem alguma hipótese adicional. A mais simples é a da regularidade do ponto  $\tilde{x}$  em relação às restrições.

### Definição 12.2



Um ponto  $\tilde{x}$  que satisfaz as equações  $h(x) = 0$  é regular em relação às restrições se e somente se o conjunto de vetores  $\{\nabla h_1(\tilde{x}), \dots, \nabla h_m(\tilde{x})\}$  é linearmente independente.

Com esta hipótese sobre  $\tilde{x}$  é possível caracterizar os arcos factíveis diferenciáveis.

### Teorema 12.1

Se  $\tilde{x}$  é um ponto regular da superfície  $S \equiv \{x \in \mathbb{R}^n \mid h(x) = 0\}$ , então o plano tangente  $T$  verifica

$$T = \{y \in \mathbb{R}^n \mid J_h(\tilde{x})y = 0\}. \quad (12.4)$$

**Prova:** Ver Luenberger [11].■

## 12.2 CONDIÇÕES NECESSÁRIAS DE PRIMEIRA ORDEM

### Teorema 12.2

Seja  $x^*$  um minimizador local de (12.1). Suponhamos que  $x^*$  é um ponto regular das restrições. Então, existe  $\lambda^* \in \mathbb{R}^m$  tal que

$$\nabla f(x^*) = \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*),$$

ou, equivalentemente,

$$Z^t(x^*) \nabla f(x^*) = 0,$$

onde  $Z(x^*) \in \mathbb{R}^{n \times (n-m)}$  e suas colunas formam uma base de  $Nu(J_h(x^*))$ .

**Prova:** Seja  $x^*$  um ponto regular de  $S \equiv \{x \in \mathbb{R}^n \mid h(x) = 0\}$ , minimizador local de (12.1). Então, para qualquer parametrização

$$x : (a, b) \rightarrow S, \quad x(t^*) = x^*, \quad t^* \in (a, b)$$

temos que  $t^*$  é solução do problema

$$\underset{t \in (a, b)}{\text{Minimizar}} \varphi(t) = f(x(t)). \quad (12.5)$$

A condição necessária de primeira ordem para (12.5) é  $\varphi'(t^*) = 0$ , portanto,

$$\varphi'(t^*) = J_f(x(t^*))x'(t^*) = \nabla^t f(x^*)x'(t^*) = 0, \quad (12.6)$$

ou seja,  $\nabla f(x^*)$  deve ser ortogonal a qualquer vetor do plano tangente  $T$  à superfície  $S$  em  $x^*$ .

Como  $x^*$  é regular, vale a caracterização de  $T$  dada em (12.4) e deduzimos que existe  $\lambda^* \in \mathbb{R}^m$  tal que

$$\nabla f(x^*) = J_h^t(x^*)\lambda^*, \quad (12.7)$$

ou

$$\nabla f(x^*) = \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*),$$

e

$$Z^t(x^*)\nabla f(x^*) = 0. \blacksquare \quad (12.8)$$

Os argumentos para obter (12.7) e (12.8) são idênticos aos usados em 7.2.

Observemos que estas condições são extensões imediatas das obtidas em 7.2 para restrições de igualdade lineares. O vetor  $\lambda^* \in \mathbb{R}^m$  é o vetor de multiplicadores de Lagrange associado às restrições.

### 12.3 CONDIÇÕES DE SEGUNDA ORDEM

#### Teorema 12.3

Sejam  $x^*$  um ponto regular, minimizador local de (12.1) e  $T$  como em (12.4). Supomos  $f, h \in C^2$ . Então existe  $\lambda^* \in \mathbb{R}^m$  tal que

$$\nabla f(x^*) + \sum_{j=1}^m \lambda_j^* \nabla h_j(x^*) = 0 \quad (12.9)$$

e

$$y^t \nabla_x^2 L(x^*, \lambda^*) y \geq 0 \text{ para todo } y \in T, \quad (12.10)$$

onde

$$L(x, \lambda) = f(x) + \lambda^t h(x), \quad x \in \mathbb{R}^n, \quad \lambda \in \mathbb{R}^m$$

é a chamada função lagrangeana.

**Prova:** (12.9) é o Teorema 12.2.

Supomos agora que  $x(t) \in C^2$ .

A condição necessária de segunda ordem para (12.5) é  $\varphi''(t^*) \geq 0$ .

Agora,

$$\varphi'(t) = \nabla^t f(x(t))x'(t) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x(t))x'_i(t),$$

portanto,

$$\varphi''(t) = \sum_{i=1}^n \left( \frac{\partial f}{\partial x_i}(x(t))x'_i(t) \right)'. \quad (12.11)$$

Mas

$$\left(\frac{\partial f}{\partial x_i}(x(t))x'_i(t)\right)' = \nabla^t \left[ \frac{\partial f}{\partial x_i}(x(t)) \right] x'(t)x'_i(t) + \frac{\partial f}{\partial x_i}(x(t))x''_i(t). \quad (12.12)$$

De (12.11) e (12.12) obtemos

$$\varphi''(t) = x'(t)^t \nabla^2 f(x(t)) x'(t) + \nabla^t f(x(t)) x''(t). \quad (12.13)$$

Por outro lado, para qualquer  $\lambda \in \mathbb{R}^m$ ,  $j \in \{1, \dots, m\}$  e  $t \in (a, b)$ ,

$$\rho_j(t) \equiv \lambda_j h_j(x(t)) = 0.$$

Portanto, para todo  $j \in \{1, \dots, m\}$  e  $t \in (a, b)$ ,

$$\rho'_j(t) = \lambda_j \nabla^t h_j(x(t)) x'(t) = 0$$

e

$$\rho''_j(t) = \lambda_j [x'(t)^t \nabla^2 h_j(x(t)) x'(t) + \nabla^t h_j(x(t)) x''(t)] = 0.$$

Então,

$$x'(t)^t \left[ \sum_{j=1}^m \lambda_j \nabla^2 h_j(x(t)) \right] x'(t) + \left[ \sum_{j=1}^m \lambda_j \nabla h_j(x(t)) \right]^t x''(t) = 0. \quad (12.14)$$

De (12.13) e (12.14) resulta, para  $t = t^*$ ,

$$\varphi''(t^*) = x'(t^*) \left[ \nabla^2 f(x^*) + \sum_{j=1}^m \lambda_j \nabla^2 h_j(x^*) \right] x'(t^*) + \left[ \nabla f(x^*) + \sum_{j=1}^m \lambda_j \nabla h_j(x^*) \right]^t x''(t^*). \quad (12.15)$$

Se  $x^*$  é minimizador local de (12.1), sabemos que existe  $\lambda^* \in \mathbb{R}^m$  tal que  $\nabla f(x^*) + \sum_{j=1}^m \lambda_j^* \nabla h_j(x^*) = 0$ , portanto de (12.15) obtemos

$$\varphi''(t^*) = x'(t^*) \left[ \nabla^2 f(x^*) + \sum_{j=1}^m \lambda_j^* \nabla^2 h_j(x^*) \right] x'(t^*) \geq 0, \quad (12.16)$$

onde  $x'(t^*)$  é qualquer vetor de  $T$ .

Dado que

$$\nabla_x^2 L(x, \lambda) = \nabla^2 f(x) + \sum_{j=1}^m \lambda_j \nabla^2 h_j(x),$$

(12.10) se segue de (12.16).■

É importante entender o significado de (12.10). As condições de segunda ordem expressam sempre informações sobre a curvatura das funções. No caso de restrições lineares, nas condições de segunda ordem aparece somente a função objetivo. Se consideramos restrições não-lineares, (12.10) significa que as curvaturas, tanto da função objetivo como das restrições, devem ser levadas em conta para caracterizar um minimizador local. De novo, isto estende a análise feita em 7.3, já que para restrições lineares  $\nabla^2 h_i(x) \equiv 0$ .

Na Figura 12.4 ilustramos a importância da curvatura das restrições na caracterização de um minimizador.

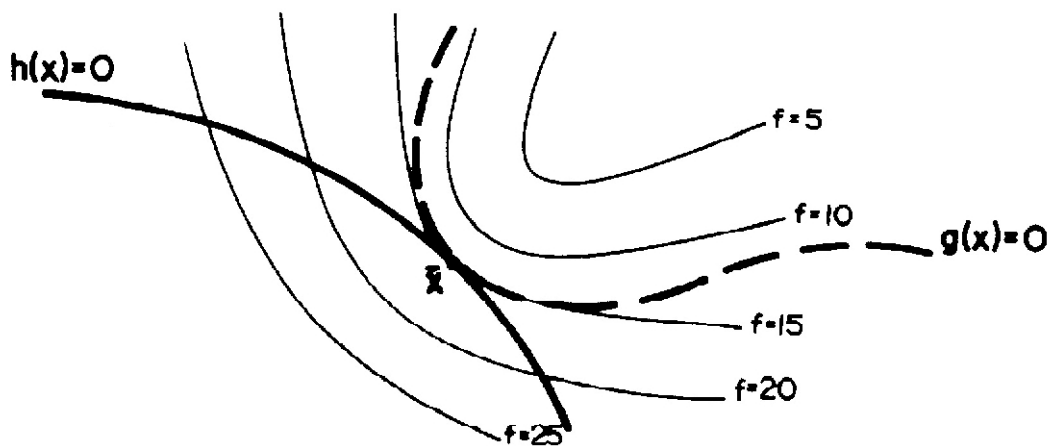


Figura 12.4

Nesta figura observamos que  $\tilde{x}$  é minimizador local de  $f$  sujeita a  $h(x) = 0$ , mas é maximizador de  $f$  sujeita a  $g(x) = 0$ .

No seguinte teorema, damos condições suficientes de segunda ordem para que um ponto regular seja minimizador estrito de  $f$  com restrições não-lineares de igualdade.

#### Teorema 12.4

*Sejam  $x^*$  um ponto regular tal que  $h(x^*) = 0$  e  $T$  como em (12.4). Se*

$\lambda^* \in \mathbb{R}^m$  é tal que

$$\nabla f(x^*) + \sum_{j=1}^m \lambda_j^* \nabla h_j(x^*) = 0$$

e

$$y^t \nabla_x^2 L(x^*, \lambda^*) y > 0 \text{ para todo } y \in T - \{0\},$$

então  $x^*$  é um minimizador local estrito de (12.1).

**Prova:** Ver Luenberger [11].■

### Exercícios

12.1 Considere o problema de encontrar o ponto da superfície  $f(x, y, z) = 0$  mais próximo da superfície  $g(x, y, z) = 0$ . Formule esse problema como um sistema não-linear. Invente exemplos!

12.2 Sejam  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $f, g \in C^2(\mathbb{R}^n)$ . Seja  $\tilde{x} \in \mathbb{R}^n$  tal que  $g(\tilde{x}) = 0$ ,  $\nabla f(\tilde{x}) = J_g^t(\tilde{x})\lambda$  e  $\nabla^2 f(\tilde{x}) > 0$ . Isso implica que  $\tilde{x}$  é minimizador local de  $f$  sujeita a  $g(x) = 0$ ? Prove ou dê um contra-exemplo.

12.3 Desejamos minimizar  $f$  sujeita a  $h_i(x) = 0$ ,  $i = 1, \dots, m$ . Suponha que  $\tilde{x}$  é uma solução desse problema e que  $\tilde{x}$  é regular. Suponha também que  $\nabla f(\tilde{x}) = 0$ . Calcule os multiplicadores de Lagrange. Interprete geometricamente.

12.4 Encontre todos os pontos estacionários da função

$$f(x) = -x_1^2 - 4x_2^2 - 16x_3^2$$

sujeita à restrição  $c(x) = 0$ , onde  $c(x)$  é dada por:

- (a)  $c(x) = x_1 - 1$ ;
- (b)  $c(x) = x_1 x_2 - 1$ ;
- (c)  $c(x) = x_1 x_2 x_3 - 1$ .

12.5 Seja  $\tilde{x}$  um ponto regular, minimizador de  $f$  sujeita a  $h(x) = 0$ , onde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$  e  $f, h \in C^2$ , com multiplicadores de Lagrange associados  $\tilde{\lambda} \in \mathbb{R}^m$ . Denotemos por  $H$  a matriz hessiana da lagrangeana em  $(\tilde{x}, \tilde{\lambda})$ ,  $H = \nabla^2 f(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i \nabla^2 h_i(\tilde{x})$ , e por  $A$  o jacobiano de  $h$  em  $\tilde{x}$ ,  $A = J_h(\tilde{x})$ . Seja  $P$  a matriz de projeção sobre o núcleo de  $A$ . Prove que a matriz definida por

$$B = P^t H P + A^t A$$

é semidefinida positiva.



## Capítulo 13

# MINIMIZAÇÃO COM RESTRICÇÕES NÃO-LINEARES DE IGUALDADE E DESIGUALDADE

Neste capítulo, consideramos problemas da forma

Minimizar  $f(x)$

$$\text{sujeita a } h(x) = 0, \quad g(x) \leq 0, \quad (13.1)$$

onde  $f, h, g \in C^1$ ,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$  com  $m < n$  e  $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ .

### 13.1 A REGIÃO DE FACTIBILIDADE

Começemos com um exemplo. Se  $n = 2$ , podemos ter, essencialmente, os casos ilustrados na Figura 13.1.

Em geral, em  $\mathbb{R}^n$ , podemos pensar que uma equação divide o espaço em duas partes e que a região factível é a intersecção dos semi-espacos determinados por cada  $g_i$  com a superfície  $h(x) = 0$ .

A complicação na caracterização dos deslocamentos factíveis a partir de um ponto  $x \in S$  é devida às restrições de igualdade e à existência de restrições de desigualdade ativas no ponto  $x$  ( $g_i(x) = 0$ ). Como no caso de restrições de igualdade, é preciso trabalhar com o conceito de arco factível. Claramente, dado um ponto  $x \in S$ , a caracterização dos arcos factíveis depende somente das restrições de igualdade e das de desigualdade que estão ativas em  $x$ . Neste caso, também precisamos do conceito de regularidade.

Seja  $S = \{x \in \mathbb{R}^n \mid h(x) = 0 \text{ e } g(x) \leq 0\}$ . Sejam  $x \in S$  e  $\mathcal{K}(x) = \{i_1, i_2, \dots, i_{s(x)}\}$  tais que  $j \in \mathcal{K}(x)$  se e somente se  $g_j(x) = 0$ . Temos que  $0 \leq s(x) \leq p$ . As  $m$  restrições correspondentes a  $h$  estão sempre ativas. Se  $r(x)$  é o número total de restrições ativas em  $x$ , temos que  $m \leq r(x) = m + s(x) \leq m + p$ .

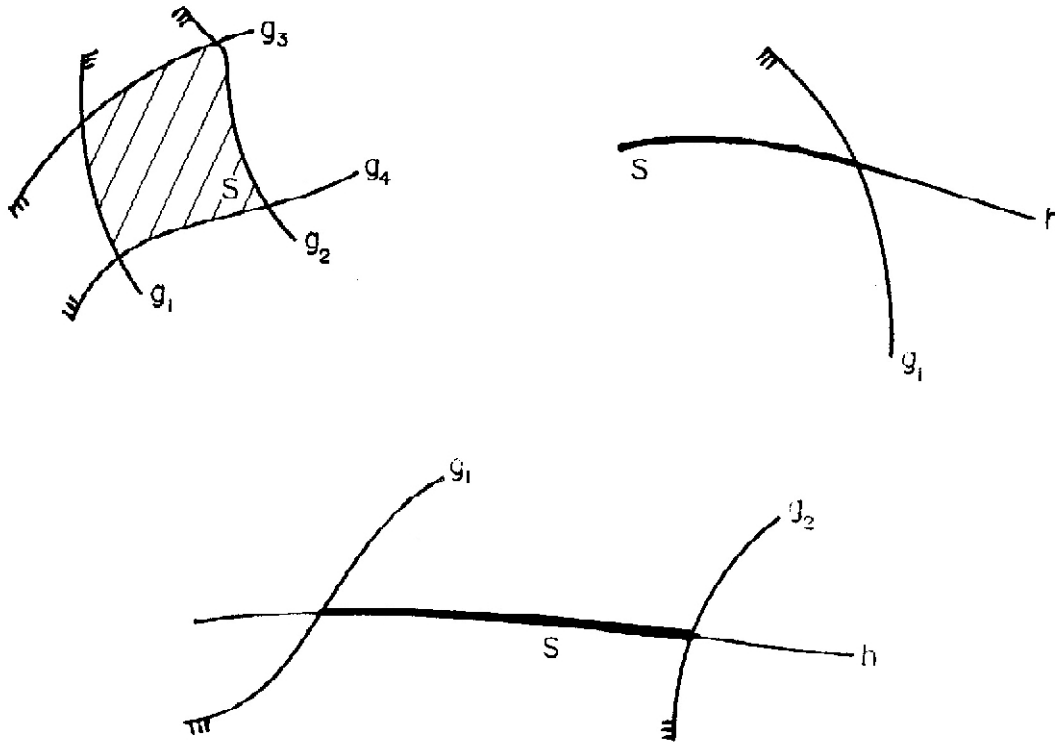


Figura 13.1

**Definição 13.1**

Dizemos que  $x \in S$  é um ponto regular se e somente se o conjunto de vetores  $\{\nabla h_1(x), \dots, \nabla h_m(x), \nabla g_{i_1}(x), \dots, \nabla g_{i_s(x)}(x)\}$  é linearmente independente.

É possível mostrar que um arco factível diferenciável tal que  $x(\tilde{t}) = \tilde{x}$  está caracterizado por

$$J_h(\tilde{x})x'(\tilde{t}) = 0$$

e

$$\nabla^t g_j(\tilde{x})x'(\tilde{t}) \leq 0, \text{ para todo } j \in \mathcal{K}(\tilde{x}).$$

Ver a Figura 13.2.



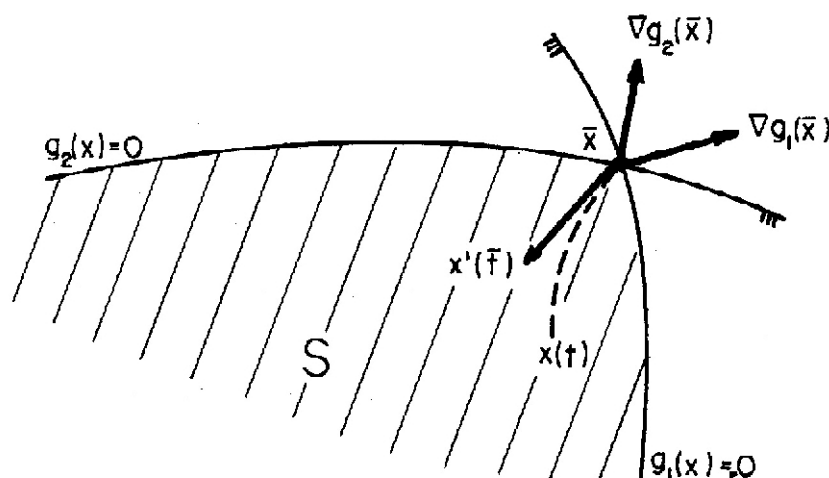


Figura 13.2

### 13.2 CONDIÇÕES NECESSÁRIAS DE PRIMEIRA ORDEM (KUHN-TUCKER)

#### Teorema 13.1

Consideremos o problema (13.1). Seja  $x^*$  um ponto factível e regular. Seja  $\mathcal{K}(x^*) = \{i_1(x^*), \dots, i_{s(x^*)}(x^*)\}$  o conjunto de índices correspondentes às restrições de desigualdade que estão ativas em  $x^*$ . Seja  $W_{\mathcal{K}} \in \mathbb{R}^{s(x^*) \times n}$ ,

$$W_{\mathcal{K}} = \begin{bmatrix} \nabla^t g_{i_1}(x^*) \\ \vdots \\ \nabla^t g_{i_{s(x^*)}}(x^*) \end{bmatrix}. \quad (13.2)$$

Se  $x^*$  é um minimizador local de (13.1), então existem  $\lambda^* \in \mathbb{R}^m$  e  $\mu^* \in \mathbb{R}^{s(x^*)}$  tais que

$$\nabla f(x^*) + J_h^t(x^*)\lambda^* + W_{\mathcal{K}}^t \mu^* = 0 \quad (13.3)$$

e

$$\mu_k^* \geq 0 \text{ para todo } k \text{ tal que } 1 \leq k \leq s(x^*). \quad (13.4)$$

**Prova:** Se  $x^*$  é um minimizador local de (13.1), então também é minimizador local do problema:

Minimizar  $f(x)$

$$\text{sujeita a } h(x) = 0, g_{i_1}(x) = 0, \dots, g_{i_{s(x^*)}}(x) = 0. \quad (13.5)$$

O Teorema 12.2 garante que existem  $\lambda^* \in \mathbb{R}^m$  e  $\mu^* \in \mathbb{R}^{s(x^*)}$  que verificam (13.3).

Para provar (13.4), suponhamos que existe  $k$  tal que  $1 \leq k \leq s(x^*)$  e  $\mu_k^* < 0$ . Sejam  $\bar{S}$  e  $\bar{T}$  respectivamente a superfície e o plano tangente definidos pelas igualdades que restam em (13.5) se retirarmos a correspondente a  $\mu_k^*$  ( $g_{i_k}(x^*) = 0$ ). Como  $x^*$  é um ponto regular temos que as linhas da matriz  $B \in \mathbb{R}^{(m+s(x^*)) \times n}$  dada por

$$B = \begin{bmatrix} J_h(x^*) \\ W_{\mathcal{K}} \end{bmatrix}$$

são linearmente independentes e, portanto, existe  $y \in \bar{T}$  tal que  $\nabla^t g_{i_k}(x^*) y < 0$ . Seja  $x(t)$  um arco factível em  $\bar{S}$  tal que  $x(t^*) = x^*$  e  $x'(t^*) = y$ . Para  $t \geq t^*$ , suficientemente pequeno,  $x(t) \in S$ . Então,  $t^*$  deve ser solução de

Minimizar  $\varphi(t) \equiv f(x(t))$

$$\text{sujeita a } t \geq t^*. \quad (13.6)$$

Usando a regra da cadeia obtemos

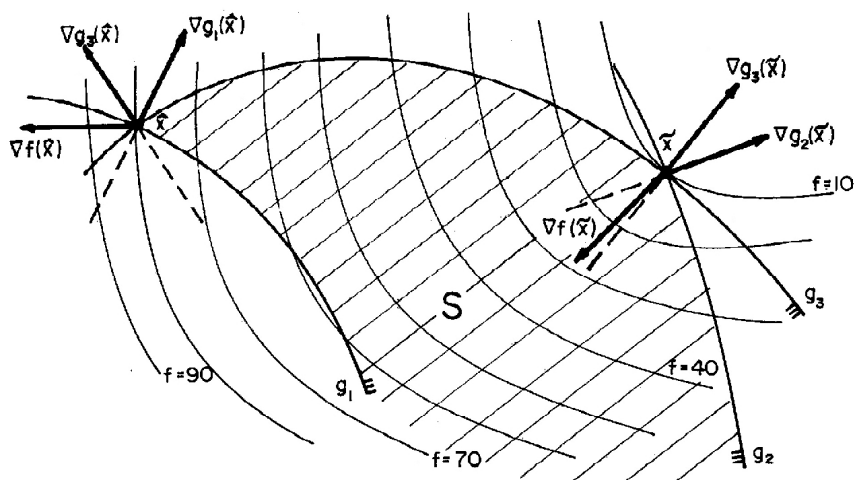
$$\varphi'(x(t^*)) = \nabla^t f(x^*) x'(t^*) = \nabla^t f(x^*) y.$$

Logo, por (13.3) e a definição de  $\bar{T}$ , resulta

$$\varphi'(x(t^*)) = -\mu_k^* \nabla^t g_{i_k}(x^*) y.$$

Porém, de  $\mu_k^* < 0$  resulta  $\varphi'(x(t^*)) < 0$ , o que contradiz o fato de  $t^*$  ser solução de (13.6). Portanto, necessariamente  $\mu_k^* \geq 0$ . ■

**Observação:** Se compararmos esta prova com aquela feita para restrições lineares em 12.2, notaremos que as idéias são essencialmente as mesmas, e que a única dificuldade é a necessidade de introduzir os “arcos” factíveis. Ver Figura 13.3.



$\tilde{x}$  verifica (13.3)–(13.4)

$\hat{x}$  não verifica (13.3)–(13.4)

Figura 13.3

### 13.3 CONDIÇÕES DE SEGUNDA ORDEM

#### Teorema 13.2

Suponhamos  $f, h, g \in C^2$ . Seja  $x^*$  um minimizador local de (13.1). Supomos que  $x^*$  é regular. Então, existem  $\lambda^* \in \mathbb{R}^m$ ,  $\mu^* \in \mathbb{R}^{s(x^*)}$  tais que (13.3) e (13.4) se verificam e, além disso, a matriz  $\nabla_x^2 L(x^*, \lambda^*, \mu^*)$ , definida por

$$\nabla_x^2 L(x^*, \lambda^*, \mu^*) \equiv \nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 h_i(x^*) + \sum_{j=1}^{s(x^*)} \mu_j^* \nabla^2 g_{i_j}(x^*), \quad (13.7)$$

verifica

$$y^t \nabla_x^2 L(x^*, \lambda^*, \mu^*) y \geq 0 \text{ para todo } y \in \tilde{T} \equiv \{y \in \mathbb{R}^n \mid By = 0\},$$

onde

$$B = \begin{bmatrix} J_h(x^*) \\ W_{\mathcal{K}} \end{bmatrix}$$

e  $W_{\mathcal{K}}$  é como no Teorema 13.1.

**Prova:**  $x^*$  deve ser solução do problema:

Minimizar  $f(x)$ , sujeita a  $h(x) = 0$ ,  $g_{i_1}(x) = 0, \dots, g_{i_{s(x^*)}}(x) = 0$ , e o resultado desejado é deduzido do Teorema 12.3. ■

**Teorema 13.3**

Sejam  $f, h, g \in C^2$ . Seja  $x^*$  factível em (13.1) tal que existem  $\lambda^* \in \mathbb{R}^m$  e  $\mu^* \in \mathbb{R}^{s(x^*)}$  com  $\mu_k^* \geq 0$  para todo  $k \in \{1, \dots, s(x^*)\}$  e

$$\nabla f(x^*) + \sum_{j=1}^m \lambda_j^* \nabla h_j(x^*) + \sum_{j=1}^{s(x^*)} \mu_j^* \nabla g_{i_j}(x^*) = 0.$$

Suponhamos que  $\nabla_x^2 L(x^*, \lambda^*, \mu^*)$ , definida como em (13.7), verifica

$$y^t \nabla_x^2 L(x^*, \lambda^*, \mu^*) y > 0$$

para todo  $y \in T' \equiv \{y \mid J_h(x^*)y = 0 \text{ e } \nabla g_j(x^*)y = 0 \forall j \in \mathcal{K}\}$ , onde

$$\mathcal{K} = \{j \in \mathcal{K}(x^*) \mid \mu_j^* > 0\}.$$

Então,  $x^*$  é um minimizador local estrito de (13.1).

**Prova:** Ver Luenberger [11]. ■

**Exercícios**

13.1 Em  $\mathbb{R}^2$  considere as seguintes restrições:

$$x_1 \geq 0 \qquad x_2 \geq 0 \qquad x_2 - (x_1 - 1)^2 \leq 0$$

Prove que  $(1, 0)^t$  é factível mas não é regular.

13.2 Considere o problema

$$\begin{aligned} &\text{Minimizar } (x + 1)^2 + (y - 1)^2 \\ &\text{s.a. } 2y - 1 = 0 \\ &\quad (1 - x)(4 - x^2 - y^2) \leq 0 \\ &\quad 100 - 2x^2 - y^2 \geq 0. \end{aligned}$$

Resolva o problema graficamente e encontre os valores exatos dos multiplicadores de Lagrange usando as condições Kuhn-Tucker.

13.3 Considere o problema

$$\begin{aligned} & \text{Maximizar } x_2^3 \\ & \text{s.a. } (x_1 - x_2)^3 \geq 0 \\ & \quad (x_1 + x_2 - 2)^3 \leq 0. \end{aligned}$$

Resolva e analise as condições de otimalidade.

13.4 Considere o problema

$$\text{Minimizar } f(x)$$

$$\text{s.a. } u(x) \leq 0, v(x) \leq 0.$$

Suponha que  $\tilde{x}$  é uma solução regular do problema acima. Defina problemas onde isso acontece e:

- (a)  $u(\tilde{x}) = v(\tilde{x}) = 0$ ;
- (b)  $u(\tilde{x}) < 0, v(\tilde{x}) = 0$ ;
- (c)  $u(\tilde{x}) < 0, v(\tilde{x}) < 0$ ;
- (d)  $u(\tilde{x}) = v(\tilde{x}) = 0$  e um dos multiplicadores é zero.

13.5 Encontre todas as soluções globais do problema de maximizar  $x_1$  sujeita às restrições:

$$\begin{aligned} x_2 - \text{sen } x_1 &= 0 \\ x_2^2 - 1 &= 0 \\ -10 &\leq x_1 \leq 10. \end{aligned}$$

13.6 Considere o problema

$$\begin{aligned} & \text{Minimizar } x_1 \\ & \text{s.a. } x_2 \geq 0 \\ & \quad x_2 \leq x_1^3. \end{aligned}$$

Qual é a solução? Por que não se verificam as condições Kuhn-Tucker?

13.7 Resolva os problemas abaixo usando as condições Kuhn-Tucker:

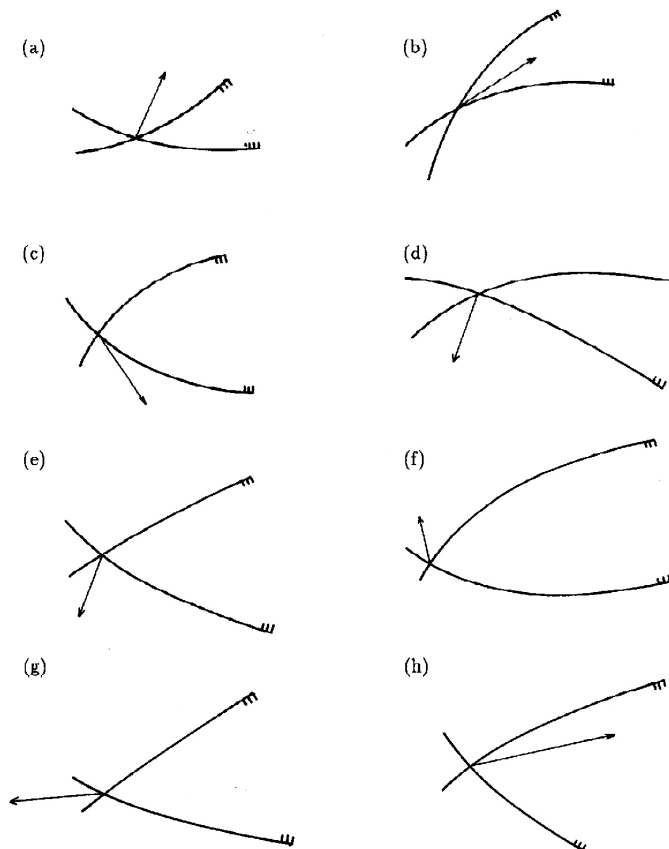
- (a) Minimizar  $\sum_{i=1}^n (1/x_i)$  s.a.  $\sum_{i=1}^n x_i^2 = n$ ,  $x_i \geq 0$ ,  $i = 1, \dots, n$ ;
- (b) Maximizar  $\prod_{i=1}^n x_i$  s.a.  $\sum_{i=1}^n x_i^2 = n$ .

13.8 Considere o seguinte problema

$$\begin{aligned} &\text{Minimizar } -x_1 + x_2 \\ &\text{s.a. } x_1^2 + x_2^2 - 2x_1 = 0 \\ &\quad (x_1, x_2) \in X, \end{aligned}$$

onde  $X$  é o conjunto formado pelas combinações convexas dos pontos  $(-1, 0)$ ,  $(0, 1)$ ,  $(1, 0)$  e  $(0, -1)$ . Encontre a solução ótima graficamente e verifique se as condições Kuhn-Tucker são cumpridas na solução obtida.

13.9 Os seguintes desenhos mostram duas restrições  $g(x) \leq 0$ ,  $h(x) \leq 0$  e o gradiente de uma função  $f$  num ponto factível  $\tilde{x}$ . Em cada caso, diga se  $\tilde{x}$  é um maximizador, minimizador ou “nada”.



13.10 Sejam  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $r : \mathbb{R}^p \rightarrow \mathbb{R}$  e  $h : \mathbb{R}^p \rightarrow \mathbb{R}^q$ . Considere os problemas

(P) Minimizar  $f(x)$   
s.a.  $g(x) \leq 0$ .

(Q) Minimizar  $r(x)$   
s.a.  $h(x) = 0$ .

Mostre como transformar (P) em (Q) e vice-versa.

13.11 Encontre a solução  $(\tilde{x}, \tilde{y})$  do problema abaixo em função do parâmetro não-negativo  $a$ :

$$\begin{aligned} &\text{Minimizar } -x + y \\ &\text{s.a. } y \geq x^2 \\ &\quad 0 \leq x \leq a \\ &\quad 0 \leq y \leq 1. \end{aligned}$$

13.12 Considere o conjunto  $S = \{(x, y) \in \mathbb{R}^2 \mid y \geq \sin x, y \leq x, x \leq \pi\}$ . Exiba uma função  $f$  tal que o minimizador dela no conjunto  $S$  não satisfaça as condições Kuhn-Tucker. Justifique.

13.13 Considere o problema

$$\begin{aligned} &\text{Maximizar } x^2 + (y - 1)^2 \\ &\text{s.a. } y \leq 2 \\ &\quad y \geq \cos \pi x \\ &\quad x + 1 \geq 0 \\ &\quad x - 1 \leq 0. \end{aligned}$$

Resolva o problema graficamente e encontre os multiplicadores de Lagrange utilizando as condições Kuhn-Tucker.

13.14 Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f \in C^1$ . Seja  $\tilde{d} \in \mathbb{R}^n$  a solução do seguinte problema:

$$\begin{aligned} &\text{Minimizar } \nabla^t f(x) \tilde{d} \\ &\text{s.a. } A \tilde{d} \leq 0, \|\tilde{d}\|^2 \leq c, \end{aligned}$$

onde  $A \in \mathbb{R}^{m \times n}$ ,  $m \leq n$  e posto  $A = m$  e  $c$  é uma constante positiva. Escreva as condições de otimalidade e interprete geometricamente. Prove que  $\nabla^t f(x) \tilde{d} \leq 0$ .





## Capítulo 14

# ALGORITMOS PARA RESTRICÇÕES NÃO-LINEARES

O desenvolvimento de algoritmos para resolver o problema geral da programação não-linear (função objetivo não-linear e restrições não-lineares) é uma tarefa difícil. Este continua sendo um campo de pesquisa aberto e trabalhos novos surgem continuamente.

Podemos considerar que há basicamente três categorias de métodos:

1. Métodos de penalização e barreira.
2. Programação quadrática seqüencial.
3. Gradiente reduzido generalizado.

Apresentamos a seguir as idéias básicas que caracterizam cada uma destas categorias.

### 14.1 MÉTODOS DE PENALIZAÇÃO E BARREIRA

Estes métodos são os primeiros que surgiram na tentativa de lidar com restrições não-lineares. Essencialmente a forma de lidar com elas é: não lidar com elas!

Para facilitar a exposição, nos métodos de penalização consideraremos apenas o problema

$$\text{Minimizar } f(x)$$

$$\text{sujeita a } h(x) = 0, \tag{14.1}$$

onde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$  ,  $m < n$ .

Dado o problema (14.1), associa-se uma seqüência de problemas irrestritos, de modo que as soluções desses problemas se aproximem da solução do problema original.

A função objetivo do problema irrestrito associado é:

$$\phi(x, \mu) = f(x) + \mu \sum_{i=1}^m (h_i(x))^2,$$

onde  $\mu > 0$  é um parâmetro. Quanto maior for o valor de  $\mu$ , estamos penalizando mais o fato de  $h_i(x)$  ser diferente de 0, para algum  $i$ .

A idéia fundamental é que se  $\mu$  cresce indefinidamente, a solução de  $\phi(x, \mu)$  será cada vez mais próxima da solução de (14.1).

Basicamente, a resolução de um problema de programação não-linear pelo método de penalização consta dos seguintes passos:

### Algoritmo 14.1

**Passo 1:** Dado  $\mu_k$ , obter  $x(\mu_k)$  solução do problema irrestrito

Minimizar  $\phi(x, \mu_k)$ .

**Passo 2:** Se  $\|h(x(\mu_k))\|$  é suficientemente pequeno (ou seja,  $x(\mu_k)$  é “quase factível”), parar. Caso contrário, escolher  $\mu^{k+1} > \mu^k$  e repetir o Passo 1.

No livro de Luenberger são apresentadas as propriedades teóricas destes métodos. Com hipóteses bastantes fracas é possível demonstrar que o processo descrito acima converge à solução de (14.1). Na prática, quando o parâmetro de penalização  $\mu_k$  é muito grande, os resultados computacionais obtidos na resolução dos problemas irrestritos associados podem não ser confiáveis. Na tentativa de evitar esta “falha” dos métodos de penalização foram introduzidas modificações que dão lugar a métodos mais eficientes.

Assim, surgem os métodos de lagrangeano aumentado, que resolvem uma seqüência de problemas irrestritos onde a função objetivo é

$$\phi(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i h_i(x) + \mu \sum_{i=1}^m (h_i(x))^2.$$

O leitor interessado em compreender as razões pelas quais isso é melhor que a simples penalização, achará matéria introdutória e referências sobre o assunto no livro de Fletcher [5].

Lancelot, um pacote computacional desenvolvido recentemente por Conn, Gould e Toint [3], que utiliza um método deste tipo, está disponível para os usuários interessados.

Os métodos de barreira são parecidos aos de penalização. Se aplicam a problemas do tipo

$$\begin{aligned} &\text{Minimizar } f(x) \\ &\text{sujeita a } h(x) \leq 0, \end{aligned}$$

onde a região factível deve ter interior não vazio. A diferença essencial é que nos métodos de penalização as aproximações sucessivas da solução não são factíveis, e nos métodos de tipo barreira, ao contrário, elas são sempre factíveis estritamente. Por isso, também são chamados *métodos de pontos interiores*.

Os problemas irrestritos típicos para os métodos de tipo barreira são

$$\text{Minimizar } \phi(x, \mu),$$

onde

$$\phi(x, \mu) = f(x) - \mu \sum_{i=1}^m \frac{1}{h_i(x)}$$

ou

$$\phi(x, \mu) = f(x) - \mu \sum_{i=1}^m \ln(-h_i(x)).$$

Estes métodos também são tratados nos livros clássicos. O interesse por estes métodos ressurgiu depois da revolução introduzida na programação linear pelo trabalho de Karmarkar [9]. Uma excelente referência para as relações entre a programação linear e os métodos de tipo barreira é Gonzaga [8].

## 14.2 PROGRAMAÇÃO QUADRÁTICA SEQUENCIAL

Programação quadrática é um caso particular do problema que analisamos no Capítulo 11. Trata-se de minimizar uma função quadrática sujeita a restrições lineares de igualdade e/ou desigualdade. Não é trivial desenvolver algoritmos eficientes para este problema, um dos mais simples de programação não-linear. Uma boa referência é o Capítulo 10 de Fletcher [5].

A programação quadrática sequencial é uma abordagem para resolver problemas gerais de programação não-linear, que consiste em resolver uma seqüência de problemas de programação quadrática.

Dada  $x^k$ , uma aproximação à solução de (14.1), associamos o seguinte problema de programação quadrática:

$$\text{Minimizar } q(d) \equiv \nabla^t f(x^k)d + \frac{1}{2}d^t Q^k d$$

$$\text{sujeita a } h(x^k) + J_h^t(x^k)d = 0. \quad (14.2)$$

Pelas condições de otimalidade de segunda ordem, vistas em (14.1), o ideal seria que  $Q^k$  fosse uma aproximação de  $\nabla_x^2 L(x^k, \lambda^k)$ , sendo  $\lambda^k$  uma estimativa dos multiplicadores de Lagrange associados a  $x^*$ .

Os métodos de programação quadrática sequencial têm a seguinte estrutura geral:

### Algoritmo 14.2

**Passo 1.** Dados  $x^k$  e  $\lambda^k$ , estimadores de  $x^*$  e  $\lambda^*$ , resolver (14.2), determinando  $d_k$  e  $\lambda^{k+1}$ , onde  $\lambda^{k+1}$  é o vetor de multiplicadores de Lagrange associado à solução  $d_k$  de (14.2).

**Passo 2.** Definir  $x^{k+1} = x^k + \alpha_k d_k$ , onde  $\alpha_k$  é escolhido de maneira a fazer decrescer uma “função de mérito” adequada.

**Passo 3.** Calcular  $Q^{k+1}$ , em geral, dependendo dos multiplicadores de Lagrange do subproblema quadrático resolvido no Passo 1.

Os problemas de como estimar  $\lambda^*$  e como atualizar  $Q^k$  são discutidos junto com as propriedades de convergência deste tipo de métodos no Capítulo 12 de Fletcher [5]. No Capítulo 6 de Gill, Murray e Wright [7], o leitor achará uma extensa lista de bibliografia relacionada.

Em relação ao software desenvolvido existe um trabalho recente de Mahidhara e Lasdon [12].

Nos problemas com desigualdades, os problemas quadráticos associados também têm restrições de desigualdade. Um método de restrições ativas pode ser utilizado neste caso para resolver os subproblemas.

### 14.3 GRADIENTE REDUZIDO GENERALIZADO

No Capítulo 8 fizemos uma breve referência ao método do gradiente reduzido para o caso de restrições lineares de igualdade. Lembramos aqui que a idéia era expressar algumas variáveis em função das outras. A generalização deste método para o caso de restrições não-lineares consiste em aproximar linearmente as restrições numa vizinhança de uma aproximação  $x^k$  da solução de (14.1). Com essa aproximação linear podemos proceder como no Capítulo 8. As matrizes usadas, que naquele caso eram constantes, agora dependerão de  $x^k$ . Os deslocamentos que produzimos com este processo são sobre o plano tangente à superfície de restrições, o que nos fornecerá um novo ponto, em geral, não factível.

Portanto, este método deve incorporar um processo para voltar à superfície definida pelas restrições. A implementação de um algoritmo para programação não-linear com estas características não é fácil.

Uma descrição e discussão das propriedades deste método, conhecido como *gradiente reduzido generalizado* (GRG), pode ser encontrada no Capítulo 11 de Luenberger [11].

Existem vários pacotes computacionais eficientes que utilizam o GRG. Ver Abadie [1] e Lasdon [10].

#### Exercícios

14.1 Proponha um método que combine penalização com barreira para minimizar  $c^t x$  sujeita a  $Ax = b$ ,  $x \geq 0$ , onde  $c$ ,  $x \in \mathbb{R}^n$ ,  $b \in \mathbb{R}^m$  e  $A \in \mathbb{R}^{m \times n}$ . Calcule o gradiente da função penalizada.

14.2 Considere a função de penalização

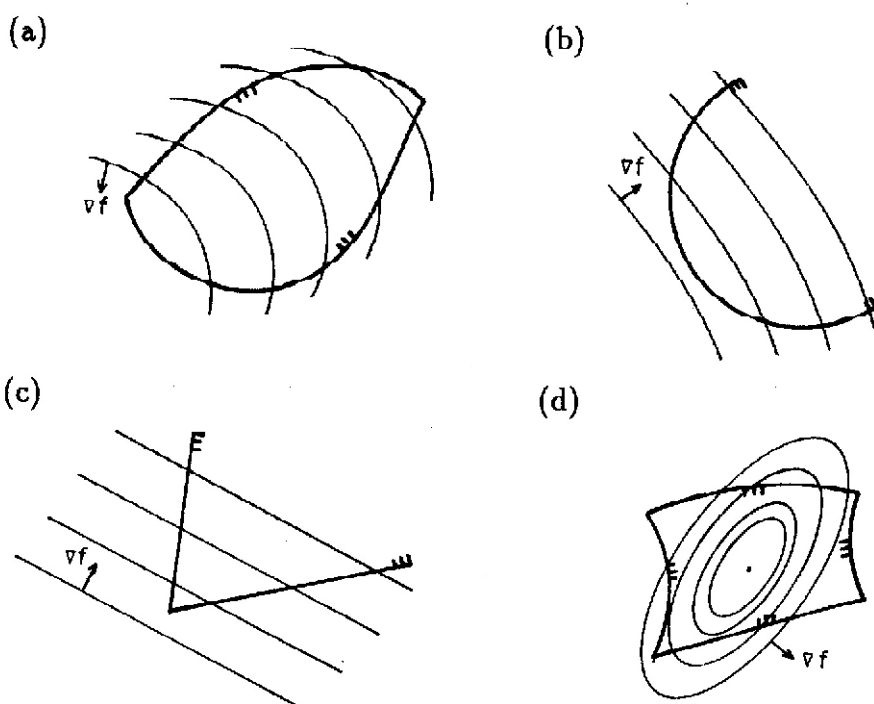
$$\phi_{\lambda, \mu}(x) = f(x) + \sum_{i=1}^m \mu_i \exp(\lambda_i h_i(x)/\mu_i),$$

onde  $\mu$ ,  $\lambda \in \mathbb{R}^m$ ,  $\mu_i > 0$ ,  $i = 1, \dots, m$ , para resolver o problema (P):

Minimizar  $f(x)$  s.a.  $h_i(x) = 0, i = 1, \dots, m$ .

Seja  $\tilde{x}$  uma solução regular de (P) com multiplicadores associados  $\tilde{\lambda} \in \mathbb{R}^m$ . Prove que  $\tilde{x}$  é um ponto estacionário de  $\phi_{\tilde{\lambda}, \mu}(x)$ .

14.3 Considere o problema de minimizar  $f$  sujeita a  $x \in R = \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$ . Para os seguintes casos, desenhe as curvas de nível da função penalizada.



14.4 Considere o problema de minimizar  $f$  sujeita a  $x \in S$ , onde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  e  $S \subseteq \mathbb{R}^n$ . Seja  $P$  uma função de penalização para  $S$  e suponha que a função penalizada  $q(x, \mu) = f(x) + \mu P(x)$  para  $\mu = \tilde{\mu}$  tem um minimizador global em  $\tilde{x}$  e que  $\tilde{x} \in S$ . Prove que  $\tilde{x}$  é um minimizador global do problema original. Interprete.

14.5 Seja  $\tilde{x}$  minimizador global de  $f$  sujeita a  $x \in S$ , onde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  e  $S \subseteq \mathbb{R}^n$ . Seja  $\bar{x}$  um ponto não factível para esse problema. Prove que existe  $\bar{\mu} > 0$  tal que  $q(\tilde{x}, \mu) \leq q(\bar{x}, \mu)$  para todo  $\mu \geq \bar{\mu}$ , onde  $q(x, \mu) = f(x) + \mu P(x)$  e  $P$  é uma função de penalização para  $S$ .

14.6 Considere o problema de minimizar  $c^t x$  sujeita a  $l \leq x \leq u$ , onde  $x, l, u \in \mathbb{R}^n$ .

- (a) Encontre as condições de otimalidade;
- (b) Faça um desenho em  $\mathbb{R}^2$ , considerando os diferentes casos possíveis relativos à localização da solução;
- (c) Nos diferentes casos, desenhe as curvas de nível da função penalizada.

14.7 Considere o problema de minimizar  $f$  sujeita a  $x \in S$ , onde  $S \subseteq \mathbb{R}^n$ . Seja  $\hat{x}(\mu)$  minimizador local da função penalizada  $q(x, \mu) = f(x) + \mu P(x)$ , onde  $P$  é uma função de penalização para  $S$ . Seja  $\tilde{x} = \lim_{\mu \rightarrow \infty} \hat{x}(\mu)$ . Suponha que  $\tilde{x} \in S$ . O que você pode afirmar sobre  $\tilde{x}$  em relação ao problema original?

14.8 Considere o problema de minimizar  $f(x)$  sujeita a  $h(x) = 0$ ,  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$  e  $f, h \in C^1$ . Seja  $\tilde{x}$  uma solução regular desse problema. Suponha que pelo menos um dos multiplicadores de Lagrange associados a  $\tilde{x}$  é diferente de zero. Prove que  $\tilde{x}$  não é minimizador local de  $q(x, \mu) \equiv f(x) + \mu \|h(x)\|^2$  para nenhum valor finito de  $\mu$ .

14.9 Considere o problema de minimizar  $f(x)$  sujeita a  $h(x) = 0$ , onde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  e  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Considere o problema penalizado, de minimizar  $q(x, \mu) \equiv f(x) + \mu \|h(x)\|_2^2$ . Mostre que os multiplicadores de Lagrange resultam ser os limites de certas quantidades que dependem do parâmetro de penalização.

14.10 Ao aplicarmos o método de penalização ao problema

$$\text{Minimizar } -x_1 - x_2 + x_3$$

$$\text{s.a. } x_1^3 + x_3 \leq 1, \quad x_1^2 + x_2^2 + x_3^2 \leq 1, \quad 0 \leq x_3 \leq 1$$

obtivemos os seguintes resultados:

$k$	$\mu_k$	$x^k$
0	1	$(0.8344, 0.8344, -0.4548)^t$
1	10	$(0.7283, 0.7283, -0.0879)^t$
2	100	$(0.7096, 0.7096, -0.0099)^t$
3	1000	$(0.7074, 0.7074, -0.0010)^t$

Utilize os dados acima para estimar a solução ótima e os multiplicadores de Lagrange, juntamente com as restrições ativas. Discuta a precisão atingida.





## Apêndice A

# NOTAÇÕES

1.  $\mathbb{R}^n$  é o conjunto dos vetores coluna

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

2.  $x^t = (x_1, x_2, \dots, x_n)$  (vetor transposto).

3.  $x^t y = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$  (produto escalar).

4.  $\|x\| = (x^t x)^{\frac{1}{2}}$  (norma euclideana).

5. Para  $x, y \in \mathbb{R}^n$ ,  $x \leq y$  significa que  $x_i \leq y_i$  para todo  $i \in \{1, 2, \dots, n\}$ .

6.  $B(x, \varepsilon) = \{y \in \mathbb{R}^n \mid \|y - x\| < \varepsilon\}$ .

7.  $\mathbb{R}^{m \times n}$  é o conjunto de matrizes de  $m \times n$ . Se  $A \in \mathbb{R}^{m \times n}$ , denotamos  $A^t$  a matriz transposta.

8.  $I$  é a matriz identidade ( $a_{ij} = 0$  se  $i \neq j$  e  $a_{ii} = 1$ ).

9. Dada  $A \in \mathbb{R}^{m \times n}$ , posto  $A$  é o cardinal do maior conjunto de vetores linearmente independentes que é possível formar com as colunas de  $A$ .

10. Se  $A \in \mathbb{R}^{m \times n}$ ,  $\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$ .

11. Se  $S \subset \mathbb{R}^n$ ,  $\overset{\circ}{S}$  é o interior de  $S$ , ou seja,

$$\overset{\circ}{S} = \{x \in S \mid \exists \varepsilon > 0 \mid B(x, \varepsilon) \subset S\}.$$

12. Dizemos que uma função  $g(\alpha)$  é um  $o(\alpha)$  se e somente se  $\lim_{\alpha \rightarrow 0} \frac{g(\alpha)}{\alpha} = 0$ .

13. Gradiente de  $f$ :

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}.$$

14. Matriz hessiana de  $f$ :  $\nabla^2 f(x) = \left[ \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right]$ .

15. Se  $g : \mathbb{R}^m \rightarrow \mathbb{R}^p$ ,  $J_g(x) \in \mathbb{R}^{p \times m}$  denota a matriz jacobiana de  $g$  em  $x$ . A  $j$ -ésima linha de  $J_g(x)$  é  $\nabla^t g_j(x)$ .

16.  $C^k$  denota o conjunto de funções  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  tais que todas as derivadas de ordem menor ou igual a  $k$  são contínuas.

17. Se a matriz  $A$  é semidefinida positiva ( $x^t A x \geq 0$  para todo  $x \in \mathbb{R}^n$ ), escrevemos  $A \geq 0$ . Analogamente, se  $A$  é definida positiva ( $x^t A x > 0$  para todo  $x \neq 0$ ), escrevemos  $A > 0$ .

## Referências Bibliográficas

- [1] ABADIE, J. The GRG method for nonlinear programming. In *Design and Implementation of Optimization Software*. Holanda, J. Greenberg, Sijthoff and Noordhoff (editores), 1978.
- [2] BAZARAA, M.; SHETTY, C. M. *Nonlinear programming – theory and algorithms*. Nova York, John Wiley and Sons, 1979.
- [3] CONN, A. R.; GOULD, N.; TOINT Ph. L. A comprehensive description of Lancelot. Technical Report, Department of Mathematics, FUNDP, Namur, Bélgica, 1990.
- [4] DENNIS, J. E.; SCHNABEL, R. B. *Numerical methods for unconstrained optimization and nonlinear equations*. Englewood Cliffs, Prentice Hall, 1983.
- [5] FLETCHER, R. *Practical methods of optimization*. 2<sup>a</sup> ed., Nova York, John Wiley and Sons, 1986.
- [6] FRIEDLANDER, A.; MARTÍNEZ, J. M. New algorithms for maximization of concave functions with box constraints. *Rairo – Operations Research* 26, 1992, pp. 209-236.
- [7] GILL, P. E ; MURRAY, W. ; WRIGHT, M. *Practical optimization*. Nova York, Academic Press, 1981.
- [8] GONZAGA, C. C. *Algoritmos de pontos interiores para programação linear*. 17<sup>o</sup> Colóquio Brasileiro de Matemática, Rio de Janeiro, IMPA, Sociedade Brasileira de Matemática, 1989.
- [9] KARMARKAR, N. A new polynomial-time algorithm for linear programming. *Combinatorics* 4, 1984, pp. 373-395.

- 
- [10] LASDON, L. S. Nonlinear programming: algorithms, applications, software and comparisons. In *Numerical optimization 1984*, Philadelphia, P. T. Boggs, R. H. Byrd and R. B. Schnabel (editores), SIAM Publications, 1985.
- [11] LUENBERGER, D. G. *Linear and nonlinear programming*. 2<sup>a</sup> ed., Nova York, Addison – Wesley Publishing Company, 1986.
- [12] MAHIDHARA, D; LASDON, L. An SQP algorithm for large sparse non linear programs. Austin, MSIS Department – School of Business Administration, University of Texas, 1991.
- [13] MC CORMICK, G. P. *Nonlinear programming*. Nova York, John Wiley and Sons, 1983.
- [14] REY PASTOR, J.; PI CALLEJA, P.; TREJO, C. A. *Análisis matemático*, volumes 1 e 2, 7<sup>a</sup> ed., Buenos Aires, Editorial Kapelusz, 1963.