SOCIEDADE PORTUGUESA
DE MATEMÁTICA
Contribuinte Nº 501 065 792

ANO XXXI — N.os 117-120 **GAZETA DE MATEMÁTICA** JANEIRO/DEZ. — 1970

EDITOR: *Gazeta de Matemática, Lda.* ADMINISTRADOR: *A. Sá da Costa*

REDACTORES: *J. Gaspar Teixeira, J. Morgado e J. da Silva Paulo*

Composto na Tipografia Matemática, Lda. — Rua Diário de Notícias, 134 - 1.º - Esq. — Telef. 36 94 49 — LISBOA

Aspectos fundamentais da teoria geral da programação matemática

por **Fernando de Jesus**

1. Introdução

Designando por $x = (x_1, \cdots, x_n)$ um ponto ou vector (1) do espaço euclideano R^n, seja $f : D \subseteq R^n \to R$. O problema geral da programação matemática consiste em optimizar a *função-objectivo* (ou *função-critério*) $f(x)$ sujeita às restrições $g_i(x) \leqq 0$ $(i = 1, \cdots, m)$ e $x \geqq 0$. Em alguns problemas de programação matemática podem estar ausentes todas ou parte das restrições $x_j \geqq 0$ $(j = 1, \cdots, n)$ mas, como toda a variável livre pode considerar-se diferença de duas variáveis não-negativas, tais problemas podem reduzir-se à à forma geral apresentada; reciprocamente, se as variáveis são não-negativas, o programa pode apresentar-se na forma seguinte: optimizar $f(x)$ com as restrições $g_i(x) \leqq 0$, supondo que estas incluem as inequações $-x_j \leqq 0$ $(j = 1, \cdots, n)$.

Caso especial importante e já estudado desenvolvidamente é o da *programação linear* em que $f(x)$ é função linear e $g_i(x)$ é linear afim, isto é,

$$f(x) = \sum_{j=1}^{n} c_j x_j \quad \text{e} \quad g_i(x) = \sum_{j=1}^{n} a_{ij} x_j - b_i .$$

Se a função $f(x)$ ou alguma das funções $g_i(x)$ não assumir a forma indicada, o problema de programação diz-se *não-linear*.

Ao contrário do que acontece na programação linear cujo estudo está muito avançado, dispondo-se de algoritmos eficientes para a pesquisa do óptimo da função-objectivo, os progressos na teoria dos programas não-lineares têm sido mais modestos, podendo dizer-se que não há ainda métodos gerais eficientes para resolver todos os tipos de programas não-lineares.

A classe de programas não-lineares que tem sido estudada com maior desenvolvimento é constituída pelos problemas em que a função-objectivo é não-linear e $g_i(x)$ são lineares afins. Entre estes, há dois casos especiais que interessa referir. No primeiro, a função-objectivo é separável, isto é, $f(x) = \sum_{j=1}^{n} f_j(x_j)$; no segundo, a função-objectivo pode escrever-se como soma de uma forma linear com uma forma quadrática:

(1) Quando for considerado vector, x designará sempre um vector-coluna.

$f(x) = \sum_{j=1}^{n} c_j x_j + \sum_{j=1}^{n} \sum_{k=1}^{n} d_{jk} x_j x_k$. Este último é conhecido por problema da *programação quadrática.*

Outra classe de programas não-lineares que apresenta grande interesse na prática é constituída pelos programas lineares onde se impõe a restrição adicional de que algumas ou todas as variáveis só podem tomar valores inteiros. A maior parte dos autores continua a chamar a estes problemas *programas lineares em números inteiros,* quando todas as variáveis só tomam valores inteiros, e *programas lineares mistos* quando só algumas variáveis tomam valores inteiros.

Os programas matemáticos aparecem frequentemente em estudos no âmbito de várias disciplinas. Em particular, podem surgir em economia matemática, econometria, estatística e investigação operacional. A título exemplificativo, consideramos seguidamente alguns problemas que se podem formalizar por meio de um programa não-linear.

EXEMPLO 1. 1. Admitamos que um monopolista deseja maximizar o seu rendimento. O preço de venda p_i do i-ésimo produto depende da quantidade de todos os outros que são produzidos e suponhamos que

$$p_i = b_i - \sum_{j=1}^{n} a_{ij} x_j \quad (i = 1, \cdots, n).$$

Substituindo estes valores na função-rendimento $f(x) = \sum_{j=1}^{n} p_j x_j$, obtém-se uma função quadrática que deverá ser maximizada, sujeita às restrições $x_j \geqq 0$ $(j = 1, \cdots, n)$ e a certas restrições de capacidade.

EXEMPLO 1. 2. (Problema da escolha de uma carteira de títulos). Suponhamos que uma unidade económica deseja aplicar dada soma de numerário C na aquisição de títulos de rendimento. O objectivo a atingir pode ser o de maximizar o rendimento médio para uma variância fixada, ou o de minimizar a variância para um rendimento fixado. Se $x_j (j = 1, \cdots, n)$ é a quantia a ser investida no j-ésimo título, então $\sum_{j=1}^{n} x_j = C$. O rendimento total é $\sum_{j=1}^{n} r_j x_j$, o seu valor médio é $E = \sum_{j=1}^{n} \mu_j x_j$ e a sua variância é

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j,$$

onde $[a_{ij}]$ é a matriz das variâncias e covariâncias das variáveis aleatórias r_j de valor médio μ_j. O problema consiste então em maximizar $\sum_{j=1}^{n} \mu_j x_j$ sujeita às restrições

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \leqq \sigma$$

(σ constante), $\sum_{j=1}^{n} x_j = C$ e $x_j \geqq 0$; ou minimizar $\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$ sujeita às restrições $\sum_{j=1}^{n} \mu_j x_j = \mu$ (μ constante), $\sum_{j=1}^{n} x_j = C$, $x_j \geqq 0$.

EXEMPLO 1. 3. (Ajustamento linear com restrições). Consideremos o problema de ajustar uma recta $y = ax + b$ a um conjunto de dados observados (x_j, y_j) $(j = 1, \cdots, n)$ e que se impõem restrições aos coeficientes a e b. O problema pode tomar, por exemplo, a forma seguinte: achar a e b que minimizam a função $f(a, b) = \sum_{j=1}^{n} [y_j - (ax_j + b)]^2$ sujeita às restrições $a \geqq 0$, $b \geqq 0$.

EXEMPLO 1. 4 (Modelo não-linear de relações interindustriais). Admitamos que para cada um de n bens existe uma actividade de produção, uma actividade de importação e uma actividade de exportação. Sejam

X_j = quantidade produzida do j-ésimo bem
M_j = quantidade importada do j-ésimo bem
E_j = quantidade exportada do j-ésimo bem

Os parâmetros do modelo são os seguintes:

a_{ij} = coeficiente de produção para o i-ésimo bem utilizado na produção do j-ésimo (quantidade do i-ésimo bem utilizada na produção de uma unidade do j-ésimo bem)
w_j = coeficiente de produção para o trabalho utilizado na produção do j-ésimo bem
c_j = coeficiente de produção para o capital utilizado na produção do j-ésimo bem
g_j = preço unitário de importação do j-ésimo bem
h_j = preço unitário de exportação do j-ésimo bem
Y_j = procura final do j-ésimo bem
L = oferta de trabalho disponível
D = máximo défice admissível da balança comercial.

Todos os parâmetros são considerados não-negativos.

O problema de programação pode formular-se do modo seguinte:

Minimizar o capital total necessário

$$C = \sum_{i=1}^{n} c_j X_j$$

satisfeitas as restrições

Não-negatividade das variáveis: $X_j \geqq 0$, $M_j \geqq 0$, $E_j \geqq 0$ $(j = 1, \cdots, n)$

Produção: $X_j + M_j - E_j - \sum_{k=1}^{n} a_{jk} X_k \geqq Y_j$ $(j = 1, \cdots, n)$

Balança comercial:

$$D + \sum_{j=1}^{n} h_j E_j - \sum_{j=1}^{n} g_j M_j \geqq 0$$

Oferta de trabalho: $L - \sum_{j=1}^{n} w_j X_j \geqq 0$.

Duas hipóteses que também se admitem são as seguintes:

i) $a_{ij} \geqq 0$ $(i, j = 1, \cdots, n)$, $\sum_{i=1}^{n} a_{ij} < 1$ $(j = 1, \cdots, n)$ (Propriedades de LEONTIEF).

ii) $h_j = \gamma_j + \rho_j E_j$ com $\gamma_j > 0$, $\rho_j < 0$ $(j = 1, \cdots, n)$.

A primeira traduz que o «output» excede o «input» para todos os bens, condição necessária para planos de produção eficientes; a segunda hipótese admite-se frequentemente em economia: quanto maior for a quantidade exportada do j-ésimo bem, mais pequeno é o rendimento por unidade.

Em virtude de se ter $h_j = \gamma_j + \rho_j E_j$, é evidente que a não-linearidade do modelo resulta da restrição referente à balança comercial.

Restringindo a atenção aos programas em que $f(x)$ e $g_i(x)$ são contínuas e diferenciáveis até à ordem desejada, convém referir que uma das técnicas para abordar a resolução de um problema de programação matemática consiste em adoptar a *transformação em variáveis quadradas*. Esta tem por objectivo transformar as inequações em equações e portanto dar ao problema a forma de um problema de extremos condicionados por equações, susceptível de tratamento, por

exemplo, pelo clássico método dos multiplicadores de LAGRANGE.

Considerando o programa na forma

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

com as restrições

$$g_i(x) \leqq 0 \quad (i = 1, \cdots, m)$$
$$x_j \geqq 0 \quad (j = 1, \cdots, n),$$

introduzamos as variáveis quadradas $y_j^2 (j = 1, \cdots, n)$ e $z_i^2 (i = 1, \cdots, m)$. O problema pode então reduzir-se à forma

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

com as restrições

$$g_i(x) + z_i^2 = 0 \quad (i = 1, \cdots, m)$$
$$x_j - y_j^2 = 0 \quad (j = 1, \cdots, n).$$

Em geral, este método clássico depara na prática com numerosas dificuldades que provêm, fundamentalmente, da necessidade de analisar grande número de soluções possíveis.

É interessante estudar a aplicação do método proposto à resolução de um programa linear para ter ideia das dificuldades de ordem prática que surgem, mesmo no caso de estar presente a linearidade.

Como se sabe, mediante a introdução de convenientes variáveis auxiliares, todo o programa linear se pode reduzir à forma

$$\text{maximizar } f(x) = \sum_{j=1}^{n} c_j x_j$$

com as restrições

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \qquad (i = 1, \cdots, m)$$
$$x_j \geqq 0 \qquad (j = 1, \cdots, n).$$

Utilizando variáveis quadradas

$$y_j^2 (j = 1, \cdots, n)$$

o problema reduz-se a maximizar

$$f(x) = \sum_{j=1}^{n} c_j x_j$$

com as restrições

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \qquad (i = 1, \cdots, m)$$
$$x_j - y_j^2 = 0 \qquad (j = 1, \cdots, n).$$

O método dos multiplicadores de LAGRANGE conduz à maximização da função lagrangeana

$$1) \qquad F(x, y, \lambda, \mu) = f(x) -$$
$$- \sum_{i=1}^{m} \lambda_i \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right) - \sum_{j=1}^{n} \mu_j (x_j - y_j^2)$$

ou

$$1') \qquad F(x, y, \lambda, \mu) = \sum_{i=1}^{m} b_i \lambda_i +$$
$$+ \sum_{j=1}^{n} \left(c_j - \sum_{i=1}^{m} \lambda_i a_{ij} - \mu_j \right) x_j + \sum_{j=1}^{n} \mu_j y_j^2 .$$

As condições necessárias para o máximo de $F(x, y, \lambda, \mu)$ são as seguintes:

$$2) \begin{cases} \dfrac{\partial F}{\partial x_j} = c_j - \displaystyle\sum_{i=1}^{m} \lambda_i a_{ij} - \mu_j = 0 \quad (j = 1, \cdots, n) \\ \dfrac{\partial F}{\partial y_j} = 2 \mu_j y_j = 0 \quad (j = 1, \cdots, n) \\ \dfrac{\partial F}{\partial \lambda_i} = - \left(\displaystyle\sum_{j=1}^{n} a_{ij} x_j - b_i \right) = 0 \quad (i = 1, \cdots, m) \\ \dfrac{\partial F}{\partial \mu_j} = - (x_j - y_j^2) = 0 \quad (j = 1, \cdots, n), \end{cases}$$

fàcilmente redutíveis a

$$2') \quad \begin{cases} c_j - \sum_{i=1}^{m} \lambda_i a_{ij} = \mu_j \qquad (j = 1, \cdots, n) \\ \mu_j x_j = 0 \qquad (j = 1, \cdots, n) \\ \sum_{j=1}^{n} a_{ij} x_j = b_i \qquad (i = 1, \cdots, m). \end{cases}$$

Observando que $F(x, y, \lambda, \mu)$ só possui máximo livre se $\mu_j \leqq 0$, então pode enunciar-se a proposição seguinte:

I. *Se existem multiplicadores* $\lambda_i = \lambda_i^*$ *e* $\mu_j = \mu_j^*$ *e variáveis* $x_j = x_j^* \geqq 0$, *satisfazendo as condições*

$$c_j - \sum_{i=1}^{m} \lambda_i^* a_{ij} = \mu_j^* \leqq 0$$

$$\mu_j^* x_j^* = 0$$

$$\sum_{j=1}^{n} a_{ij} x_j^* = b_i,$$

então $(x_1^*, \cdots, x_n^*)$ *é solução maximizante do programa linear proposto.*

Considerando o sistema dual (λ_i não sujeito a restrições de sinal)

$$c_j - \sum_{i=1}^{m} \lambda_i a_{ij} \leqq 0,$$

tem-se $(x_j \geqq 0)$

$$c_j x_j - \sum_{i=1}^{m} \lambda_i a_{ij} x_j \leqq 0$$

e

$$\sum_{j=1}^{n} c_j x_j - \sum_{j=1}^{n} \sum_{i=1}^{m} \lambda_i a_{ij} x_j \leqq 0$$

donde resulta

$$3) \qquad \sum_{j=1}^{n} c_j x_j \leqq \sum_{i=1}^{m} b_i \lambda_i .$$

Por outro lado, de acordo com o teorema I,

$$\sum_{=1}^{m} \sum_{j=1}^{n} a_{ij} x_j^* \lambda_i^* = \sum_{i=1}^{m} b_i \lambda_i^*$$

ou

$$\sum_{j=1}^{n} (c_j - \mu_j^*) x_j^* = \sum_{i=1}^{m} b_i \lambda_i^*$$

donde vem

$$4) \qquad \sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i \lambda_i^* .$$

Pode pois concluir-se, atendendo a 3) e 4), que

II. *Resolvendo o problema de programação linear pelo método dos multiplicadores de Lagrange, verifica-se que, existindo os multiplicadores* λ_i *(sem restrição de sinal), eles devem minimizar a função*

$$\sum_{i=1}^{m} b_i \lambda_i ,$$

satisfazendo as condições

$$c_j - \sum_{i=1}^{m} \lambda_i a_{ij} = \mu_j \leqq 0$$

$$\mu_j x_j = 0 .$$

É evidente que se está de novo perante um problema de programação linear (programa dual). As condições

$$\mu_j x_j = 0 \quad (j = 1, \cdots, n)$$

conduzem a 2^n casos possíveis pois é necessário considerar quer $\mu_j = 0$ quer $x_j = 0$.

Verificou-se pois que *o método de resolução de um programa linear pelo método dos multiplicadores de Lagrange revela-se não-operacional porque é impossível na prática, quando n é grande, considerar todos os* 2^n *casos possíveis.*

O *método de simplex* para resolver um programa linear pode considerar-se como um processo sistemático para afastar a maior parte dos casos e considerar apenas um pequeno número. De facto, o método do simplex restringe o número de casos, considerando sòmente aqueles em que $n-m$ das variáveis são nulas, onde o determinante da matriz dos coeficientes das m variáveis restantes é não-nulo, e onde o valor bem definido destas m variáveis é positivo (caso de não-degenerescência). As condições $\mu_j x_j = 0$ indicam que $\mu_j = 0$ para $x_j > 0$ e isso determina de uma maneira unívoca λ_i e os μ_j restantes. Se todos os μ_j não satisfazem à restrição $\mu_j \leqq 0$, põe-se a solução de parte e examina-se novo caso na iteração seguinte, etc.

O interesse pelos problemas de programação não-linear tem aumentado simultâneamente com o crescente interesse pela programação linear. Em 1951, H. W. KUHN e A. W. TUCKER publicaram um importante artigo *Nonlinear Programming* onde estudaram condições necessárias e suficientes para a existência de soluções óptimas para problemas de programação e que constituíram ponto de partida para grande número de trabalhos sobre programação não-linear. Generalizações desta memória inicial apareceram feitas por vários autores num trabalho posterior intitulado *Studies in linear and nonlinear programming*, editado por K. J. ARROW, L. HURWICZ e H. UZAWA, e publicado em 1958.

Em 1954, A. CHARNES e C. LEMKE publicaram um método de aproximação para resolver problemas de minimização de uma função separável sujeita a restrições lineares quando cada uma das funções separáveis é convexa. Uma formulação alternativa deste problema foi dada por DANTZIG em 1956. Esta técnica foi generalizada em 1963 por C. MILLER por forma a incluir restrições separáveis. A partir de 1955 apareceram numerosas contribuições nos domínios da programação quadrática.

Pressupõe-se nesta exposição que o leitor conhece a teoria dos programas lineares e os métodos de optimização clássicos. O objectivo deste trabalho consistirá em apresentar os aspectos fundamentais da teoria moderna dos programas matemáticos.

2. Funções convexas e côncavas

As noções de convexidade e concavidade de uma função desempenham papel relevante em várias partes da teoria dos programas matemáticos e por isso é conveniente referir aqui os conceitos e proposições fundamentais relativos às funções convexas e côncavas.

Sendo $X \subseteq R^n$, diz-se que X é *conjunto convexo* sse, com $0 \leqq \theta \leqq 1$ e $\overline{\theta} = 1 - \theta$,

$$1) \qquad \forall\, x, y \in R^n \;\; \theta x + \overline{\theta} y \in X.$$

Consideram-se convexos o conjunto vazio, o conjunto constituído por um só elemento e a totalidade de R^n.

A função numérica f, definida no conjunto convexo X, é *convexa* sse

$$2) \quad \forall\, x, y \in X \; f(\theta x + \overline{\theta} y) \leqq \theta f(x) + \overline{\theta} f(y).$$

Se em 2) tem lugar apenas a desigualdade em sentido restrito quando $x \neq y$ e $0 < \theta < 1$, então f diz-se *convexa em sentido restrito*. Uma função f diz-se *côncava (côncava em sentido restrito)* sse $-f$ é convexa (convexa em sentido restrito).

EXEMPLO 2. 1. Chama-se *função linear afim* à função numérica do tipo $g(x)=cx+\alpha$ onde $c \in R^n$ e $\alpha \in R$: é òbviamente uma função simultâneamente convexa e côncava em R^n.

EXEMPLO 2. 2. A *forma quadrática* $g(x)= x^T A x$ semidefinida positiva (semidefinida negativa) é função convexa (côncava) sobre R^n; em particular, a forma quadrática definida positiva (definida negativa) é convexa (côncava) em sentido restrito.

Com efeito, para a forma quadrática semidefinida positiva, tomem-se dois pontos x e y e qualquer θ $(0 \leqq \theta \leqq 1)$. Então se $\hat{x} = \theta x + \overline{\theta} y$, tem-se

$$\begin{aligned}\hat{x}^T A \hat{x} &= (\theta x + \overline{\theta} y)^T A (\theta x + \overline{\theta} y) \\ &= [y + \theta (x - y)]^T A [y + \theta (x - y)] \\ &= y^T A y + 2\theta (x - y)^T A y + \\ &+ \theta^2 (x - y)^T A (x - y).\end{aligned}$$

Supondo $x^T A x \geqq 0$ para todo o x, então, para $0 \leqq \theta \leqq 1$, $\theta x^T A x \geqq \theta^2 x^T A x$ e pode escrever-se

$$\hat{x}^T A \hat{x} \leqq y^T A y + 2\theta (x - y)^T A y + \\ + \theta (x - y)^T A (x - y)$$

ou

$$\hat{x}^T A \hat{x} \leqq y^T A y + \theta (x - y)^T A y + \\ + \theta (x - y)^T A x \leqq \theta x^T A x + \theta y^T A y$$

Como consequência imediata da definição, é possível mostrar que

I. *Se as funções* f_j $(j = 1, \cdots, k)$ *são convexas (côncavas) no conjunto convexo* X *de* R^n, *então* $\sum_{j=1}^{k} \lambda_j f_j$ *é também função convexa (côncava) em* X *com* $\lambda_j \geqq 0$.

Pode demonstrar-se também que

II. f *é convexa em* X *sse* $[X, f] = \{(x, z) : f(x) \leqq z\}$ *é conjunto convexo de* R^{n+1}.

A condição é necessária. Com efeito, se f é convexa em X e se (x, z) e (x', z') pertencem a $[X, f]$ virá $f(x) \leqq z$ e $f(x') \leqq z'$ e

$$f(\theta x + \overline{\theta} x') \leqq \theta f(x) + \overline{\theta} f(x') \leqq \theta z + \overline{\theta} z'.$$

O ponto

$$\theta (x, z) + \overline{\theta} (x', z') = (\theta x + \overline{\theta} x', \theta z + \overline{\theta} z')$$

pertencerá então a $[X, f]$ o que prova a convexidade deste conjunto.

A condição é suficiente. De facto, se $[X, f]$ é convexo, com $(x, f(x))$ e $(y, f(y))$ pertencentes a $[X, f]$, vem

$$\theta (x, f(x)) + \overline{\theta} (y, f(y)) = (\theta x + \overline{\theta} y, \theta f(x) + \\ + \overline{\theta} f(y)) \in [X, f]$$

e portanto

$$f(\theta x + \overline{\theta} y) \leqq \theta f(x) + \overline{\theta} f(y).$$

Anàlogamente se mostraria que

II′. f *é côncava em* X *sse* $[X, f] = \{(x, z) : f(x) \geqq z\}$ *é conjunto convexo de* R^{n+1}.

Os teoremas II e II′ utilizam-se muitas vezes para caracterizar as funções convexas e côncavas. Eis agora uma propriedade extremamente importante para o estudo da programação matemática:

III. *Se* f *é função convexa (côncava) no conjunto fechado* X, *então qualquer mínimo (máximo) relativo de* f *em* X *é também mínimo (máximo) absoluto sobre* X.

Suponhamos f convexa no conjunto fechado X e admitamos que f assume um

mínimo relativo em $x^0 \in X$. Se f assume o mínimo absoluto em x^*, não se pode ter $f(x^0) > f(x^*)$. De facto, se assim acontecesse, a convexidade de f permitiria escrever

$$f(\theta x^* + \overline{\theta} x^0) \leqq \theta f(x^*) + \overline{\theta} f(x^0) < f(x^0).$$

Considerando qualquer $V_\varepsilon(x^0)$ com

$$\varepsilon < |x^* - x^0|$$

e tomando $0 < \theta < \dfrac{\varepsilon}{|x^* - x^0|}$, tem-se

$$x = \theta x^* + \overline{\theta} x^0 \in V_\varepsilon(x^0) \text{ e } f(x) < f(x^0).$$

Esta desigualdade contradiz a hipótese de f possuir mínimo relativo em x^0.

A definição de convexidade e concavidade permite também mostrar fàcilmente que

IV. *O conjunto de pontos de X onde a função convexa (côncava) f assume o seu mínimo (máximo) absoluto é conjunto convexo; se o extremo absoluto é assumido em dois pontos distintos, então também é assumido numa infinidade de pontos.*

O teorema é óbvio se o extremo absoluto é assumido num só ponto. Suponhamos então que o mínimo absoluto da função convexa f é assumido em dois pontos distintos x^1 e x^2. Mostra-se também que o mínimo é assumido em qualquer ponto $x = \theta x^1 + \overline{\theta} x^2$. De facto,

$$f(x) = f(\theta x^1 + \overline{\theta} x^2) \leqq \theta f(x^1) + \overline{\theta} f(x^2) = \\ = f(x^2) = f(x^1)$$

e, como $f(x) \geqq f(x^2)$, é evidente que $f(x) = f(x^2) = f(x^1)$.

V. *Se f é convexa (côncava) em sentido restrito, então o mínimo (máximo) absoluto de f é assumido num ponto único.*

Na definição de função convexa (côncava) não se exigiu que $f(x)$ fosse contínua ou diferenciável. Pode no entanto demonstrar-se que

VI. *Se f é convexa ou côncava em X e aí é limitada, então f é contínua no $\operatorname{int} X$.*

Uma função f convexa ou côncava não é necessàriamente diferenciável. As funções convexas (côncavas) diferenciáveis podem ser caracterizadas pela proposição seguinte:

VII. *Se X é conjunto convexo aberto e f é diferenciável em X, então f é convexa (côncava) em X sse*

$$f(x) - f(x^*) \geqq (\leqq) \nabla f(x^*)(x - x^*),$$

onde $\nabla f(x^*) = \left[\dfrac{\partial f}{\partial x_1}, \cdots, \dfrac{\partial f}{\partial x_n}\right]_{x^*}$.

Supondo f convexa em X, então, para $0 < \theta \leqq 1$, vem

$$f(\theta x + \overline{\theta} x^*) \leqq \theta f(x) + \overline{\theta} f(x^*)$$

ou

$$f[\theta x + (1 - \theta) x^*] \leqq \theta f(x) + (1 - \theta) f(x^*)$$

que se pode escrever ainda na forma

$$f(x) - f(x^*) \geqq \frac{f[x^* + \theta(x - x^*)] - f(x^*)}{\theta}.$$

Atendendo à diferenciabilidade (1) e fazendo tender θ para zero, vem imediatamente o resultado.

---

(1) Designemos por o_i uma função de escalares ou vectores com a seguinte propriedade:

$$\lim_{x \to 0} \frac{o_i(x)}{|x|^i} = 0.$$

Diz-se que f é diferenciável em x sse

$$f(x + \Delta x) = f(x) + \nabla f(x) \Delta x + o_1(\Delta x).$$

Reciprocamente, cumprida a desigualdade $f(x)-f(x^*)\geqq\nabla f(x^*)(x-x^*)$, a função f é convexa. Com efeito, de

$$f(x)-f(x^*)\geqq\nabla f(x^*)(x-x^*),$$

escolhendo

$$x^*=\theta x^1+\overline{\theta}x^2 \text{ e } x=x^1 \text{ ou } x^2,$$

obtém-se

$$f(x^1)\geqq f(\theta x^1+\overline{\theta}x^2)+$$
$$+\nabla f(\theta x^1+\overline{\theta}x^2)[x^1-(\theta x^1+\overline{\theta}x^2)]$$
$$f(x^2)\geqq f(\theta x^1+\overline{\theta}x^2)+$$
$$+\nabla f(\theta x^1+\overline{\theta}x^2)[x^2-(\theta x^1+\overline{\theta}x^2)]$$

e, notando que

$$x^1-(\theta x^1+\overline{\theta}x^2)=\overline{\theta}(x^1-x^2)$$
$$x^2-(\theta x^1+\overline{\theta}x^2)=-\theta(x^1-x^2),$$

vem

$$f(x^1)\geqq f(\theta x^1+\overline{\theta}x^2)+\overline{\theta}\nabla f(\theta x^1+\overline{\theta}x^2)(x^1-x^2)$$
$$f(x^2)\geqq f(\theta x^1+\overline{\theta}x^2)-\theta\nabla f(\theta x^1+\overline{\theta}x^2)(x^1-x^2)$$

ou

$$\theta f(x^1)\geqq\theta f(\theta x^1+\overline{\theta}x^2)+$$
$$+\theta\overline{\theta}\nabla f(\theta x^1+\overline{\theta}x^2)(x^1-x^2)$$
$$\overline{\theta}f(x^2)\geqq\overline{\theta}f(\theta x^1+\overline{\theta}x^2)-$$
$$-\theta\overline{\theta}\nabla f(\theta x^1+\overline{\theta}x^2)(x^1-x^2).$$

Somando ordenadamente estas desigualdades, vem finalmente

$$\theta f(x^1)+\overline{\theta}f(x^2)\geqq f(\theta x^1+\overline{\theta}x^2),$$

relação que exprime a convexidade de f.

VIII. *Se f é bi-diferenciável no conjunto convexo aberto X, ela é convexa (côncava) em X sse a forma quadrática*

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 f}{\partial x_i\,\partial x_j}\lambda_i\lambda_j$$

é semidefinida positiva (negativa) para todo o $x\in X$.

Observemos primeiramente que, sendo f bi-diferenciável em x, a fórmula de TAYLOR dá

$$f(x+\Delta x)=f(x)+\nabla f(x)\Delta x+$$
$$+\frac{1}{2}(\Delta x)^T\nabla^2 f(\xi)\Delta x$$

onde f tem a significação já conhecida e $\nabla^2 f(x)=\left[\frac{\partial^2 f}{\partial x_i\,\partial x_j}\right]$.

Consideremos então a expressão $(0<\theta<1)$

$$T=\overline{\theta}f(x)+\theta f(y)-f(\overline{\theta}x+\theta y)$$
$$=\overline{\theta}[f(x)-f(\overline{\theta}x+\theta y)]+\theta[f(y)-f(\overline{\theta}x+\theta y)].$$

Utilizando a fórmula de TAYLOR, vem

$$T=\overline{\theta}\nabla f(\overline{\theta}x+\theta y)\theta(x-y)+$$
$$+\frac{\theta^2\overline{\theta}}{2}(x^T-y^T)\nabla^2 f(\xi_1)(x-y)+$$
$$+\theta\overline{\theta}\nabla f(\overline{\theta}x+\theta y)(y-x)+$$
$$+\frac{\theta\overline{\theta}^2}{2}(y^T-x^T)\nabla^2 f(\xi_2)(y-x),$$

simplificando-se a primeira e terceira parcelas.

Sendo $\nabla^2 f(x)$ semidefinida positiva, vem

$$T=\overline{\theta}f(x)+\theta f(y)-f(\overline{\theta}x+\theta y)\geqq 0$$

o que exprime a convexidade de f.

Para provar a necessidade, observemos que, não sendo $\nabla^2 f(x)$ semidefinida positiva em x^0, existiria um vector u tal que

$u^T \nabla^2 f(x^0) u < 0$. Assim, para todo o x na vizinhança de x^0 ter-se-ia também

$$u^T \nabla^2 f(x) u < 0 .$$

Pondo $x = x^0$ e $y = x^0 + tu$, onde t é um escalar suficientemente pequeno, a expressão T viria negativa com esta escolha de x e y o que violaria a condição de convexidade de f.

Pode demonstrar-se também que

IX. *Se a forma quadrática*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j} \lambda_i \lambda_j$$

é definida positiva (negativa), f é convexa (côncava) em sentido restrito.

Sendo $|D_m|$ $(m \leqq n)$ o menor contido nas primeiras m linhas e primeiras m colunas da matriz hessiana

$$\left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right] \left(|D_n| = \left| \frac{\partial^2 f}{\partial x_i \partial x_j} \right| \right)$$

a teoria das formas quadráticas ensina que a forma é definida positiva quando e só quando $|D_m| > 0$ e é semidefinida quando $|D_m| \geqq 0$; é definida negativa quando e só quando $(-1)^m |D_m| > 0$ e semidefinida negativa quando e só quando $(-1)^m |D_m| \geqq 0$.

Este critério permite identificar na prática não só grande número de funções convexas e côncavas mas também definir regiões de convexidade e concavidade.

Exemplo 2. 3. A função $f(x) = x_1^2 + 2 x_1 x_2 + x_2^2$ é convexa pois

$$\frac{\partial f}{\partial x_1} = 2 x_1 + 2 x_2 \quad \frac{\partial^2 f}{\partial x_1^2} = 2 \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 2$$

$$\frac{\partial f}{\partial x_2} = 2 x_1 + 2 x_2 \quad \frac{\partial^2 f}{\partial x_2^2} = 2$$

e, considerando $\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$, vem

$$\frac{\partial^2 f}{\partial x_1^2} = 2 > 0 \text{ e } \frac{\partial^2 f}{\partial x_1^2} \frac{\partial^2 f}{\partial x_2^2} - \left(\frac{\partial^2 f}{\partial x_1 \partial x_2} \right)^2 = 0.$$

Exemplo 2. 4. A função $f(x) = -x_1^4 - 3 x_1^2 + 5 x_1 - 2 x_1 x_2 - x_2^2$ é côncava pois

$$\frac{\partial f}{\partial x_1} = -4 x_1^3 - 6 x_1 + 5 - 2 x_2$$

$$\frac{\partial^2 f}{\partial x_1^2} = -12 x_1^2 - 6$$

$$\frac{\partial f}{\partial x_2} = -2 x_1 - 2 x_2 \quad \frac{\partial^2 f}{\partial x^2} = -2$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = -2 ,$$

resultando

$$\frac{\partial^2 f}{\partial x_1^2} = -12 x_1^2 - 6 < 0$$

e

$$\frac{\partial^2 f}{\partial x_1^2} \frac{\partial^2 f}{\partial x_2^2} - \left(\frac{\partial^2 f}{\partial x_1 \partial x_2} \right)^2 = (-12 x_1^2 - 6)(-2) -$$

$$- (-2)^2 = 24 x_1^2 + 12 - 4 = 24 x_1^2 + 8 > 0 .$$

Trata-se até de função côncava em sentido restrito (teorema IX).

Exemplo 2. 5. Definir a região de concavidade para a função

$$f(x) = \frac{1}{2 \pi \sigma^2} \exp \left[- \frac{1}{2 \sigma^2} (x_1^2 + x_2^2) \right] .$$

As derivadas parciais de segunda ordem são

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{1}{2\pi\sigma^4}\left(\frac{x_1^2}{\sigma^2} - 1\right)\exp\left[-\frac{x_1^2 + x_2^2}{2\sigma^2}\right]$$

$$\frac{\partial^2 f}{\partial x_2^2} = \frac{1}{2\pi\sigma^4}\left(\frac{x_2^2}{\sigma^2} - 1\right)\exp\left[-\frac{x_1^2 + x_2^2}{2\sigma^2}\right]$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{1}{2\pi\sigma^4}\,\frac{x_1 x_2}{\sigma^2}\exp\left[-\frac{x_1^2 + x_2^2}{2\sigma^2}\right]$$

Para que a função seja côncava devem ser satisfeitas as seguintes desigualdades:

$$\frac{x_1^2}{\sigma^2} - 1 \leqq 0$$

$$\left(\frac{x_1^2}{\sigma^2} - 1\right)\left(\frac{x_2^2}{\sigma^2} - 1\right) - \left(\frac{x_1 x_2}{\sigma^2}\right)^2 \geqq 0$$

que, simplificadas, dão

$$x_1^2 \leqq \sigma^2 \quad \text{ou} \quad -\sigma \leqq x_1 \leqq \sigma$$

$x_1^2 + x_2^2 \leqq \sigma^2$ ou o círculo de raio σ e centro na origem.

A região de concavidade é pois o círculo $x_1^2 + x_2^2 \leqq \sigma^2$.

Para finalizar estas breves considerações sobre funções convexas e côncavas, referem-se dois teoremas que mostram como por meio de restrições sob a forma de desigualdades, envolvendo funções convexas, se pode definir uma região convexa.

XI. *O conjunto de pontos que satisfazem à restrição* $g(x) \leqq 0$, *onde* $g(x)$ *é convexa, é conjunto convexo.*

De facto, sendo x e y dois pontos que satisfazem a $g(x) \leqq 0$, vem

$$g(x) \leqq 0$$

$$g(y) \leqq 0$$

e

$$g(\theta x + \overline{\theta} y) \leqq \theta g(x) + \overline{\theta} g(y) \leqq 0,$$

o que mostra que $\theta x + \overline{\theta} y$ também satisfaz à restrição.

Como se sabe da teoria dos conjuntos convexos, a intersecção de conjuntos convexos é conjunto convexo e portanto

XII. *O conjunto definido por* $g_i(x) \leqq 0$, *onde* $g_i(x)$ $(i = 1, \cdots, m)$ *são convexas, é conjunto convexo.*

3. Teoremas fundamentais

Sem perda de generalidade, todo o problema de programação matemática se pode apresentar na forma

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

com as restrições

$$g_i(x) \leqq 0 \,(i = 1, \cdots, m)$$

$$x \geqq 0.$$

De facto, surgindo a necessidade de minimizar $f(x)$, proceder-se-á à maximização de $-f(x)$ (1); aparecendo restrições na forma $g_i(x) \geqq 0$, poderão transformar-se em $-g_i(x) \leqq 0$; notemos também que $g_i(x) = 0 \Leftrightarrow g_i(x) \leqq 0 \wedge -g_i(x) \leqq 0$, e, se x não está sujeito à condição de não-negatividade, poder-se-á escrever $x = x' - x''$ com $x' \geqq 0$ e $x'' \geqq 0$.

O objectivo deste parágrafo consiste no estudo de teoremas que exprimam condições necessárias e suficientes para a existência de solução óptima para um programa matemático.

---

(1) $\min f(x) = -\max(-f(x))$.

Raramente essas condições permitem obter o óptimo quando este existe mas essas proposições estão na base da construção de algoritmos iterativos e são fundamentais para a interpretação económica dos programas.

Na exposição que se segue, x é vector-coluna com n componentes. Tomando

$$G(x) = \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix},$$

o problema de programação pode apresentar-se na forma seguinte:

$$\max f(x)$$

com as restrições

$$G(x) \leqq 0$$
$$x \geqq 0.$$

Os gradientes de $f(x)$ e $g_i(x)$ representam-se por $\nabla f(x)$ e $\nabla g_i(x)$ e serão considerados vectores-linhas.

Há vantagem também em introduzir a matriz Jacobiana

$$\nabla G(x) = \left[\frac{\partial g_i}{\partial x_j}\right] \quad \begin{pmatrix} i = 1, \cdots, m \\ j = 1, \cdots, n \end{pmatrix},$$

o vector-linha $\lambda = [\lambda_1 \cdots \lambda_m]$ e a função lagrangeana

$$F(x, \lambda) = f(x) - \lambda G(x).$$

Adoptam-se também aqui as designações já conhecidas da teoria dos programas lineares. Assim, todo o ponto $\bar{x}$ tal que $G(\bar{x}) \leqq 0$ e $\bar{x} \geqq 0$ constitui *solução admissível* (ou *possível*); se, além de admissível, optimizar a função-objectivo, $\bar{x}$ toma o nome de *solução óptima*. O conjunto das soluções admissíveis constitui a *região admissível* (ou *conjunto admissível*) $C = \{x : G(x) \leqq 0 \wedge x \geqq 0\}$.

A restrição $g_i(x) \leqq 0$ diz-se *activa* no ponto $\bar{x}$ se $g_i(\bar{x}) = 0$; de contrário diz-se *inactiva*. Sendo $\bar{x}$ ponto fronteiro da região admissível, não se tem necessàriamente $g_i(\bar{x}) = 0$ para todo o i: são inactivas as restrições que definem a fronteira onde $\bar{x}$ não está.

Sendo $\bar{x}$ ponto fronteiro de C, admitamos que $g_{i_1}(\bar{x}) = \cdots = g_{i_r}(\bar{x}) = \bar{x}_{j_1} = \cdots = \bar{x}_{j_s} = 0$. Supõe-se que na região admissível é satisfeita a seguinte *condição de regularidade de Kuhn-Tucker*: a região admissível é tal que, se para todo o ponto fronteiro $\bar{x}$ e para todo o deslocamento dx suficientemente pequeno a partir de $\bar{x}$, se tem

$$\nabla g_i(\bar{x})\, dx \leqq 0 \; (i = i_1, \cdots, i_r)$$

e

$$dx_j \geqq 0 \; (j = j_1, \cdots, j_s),$$

então a direcção $dx = (dx_1, \cdots, dx_n)$ é tangencial a uma curva na região admissível, tendo $\bar{x}$ como ponto inicial (1).

Consideremos um exemplo em que não é satisfeita essa condição de regularidade. Tomando em R^2

$$g_1(x_1, x_2) = x_2 - (x_1 - 1)^2, \leqq 0$$
$$x_1 \geqq 0, \quad x_2 \geqq 0,$$

vê-se fàcilmente que no ponto fronteiro $\bar{x} = (1, 0)$ não é verificada a condição (fig. 1).

De facto, $\nabla g_1(\bar{x}) = (0, 1)$ e, para a direcção dx indicada pela seta, tem-se $\nabla g_1(\bar{x})\, dx = 0$ e $dx_i \geqq 0$ mas dx «não está» na região admissível.

---

(1) É costume dizer, na prática, que dx *está situado* na região admissível ou, talvez melhor, que dx é *direcção admissível*.

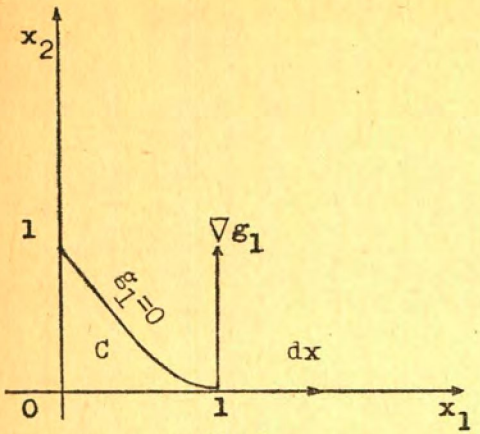

Fig. 1

Pode-se demonstrar agora o primeiro teorema fundamental:

I. *Satisfeita a condição de regularidade de Kuhn-Tucker, uma condição necessária para que a solução admissível* $\bar{x}$ *seja maximizante da função objectivo é que exista* $\bar{\lambda} \geqq 0$ *tal que*

$$\nabla f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x}) \leqq 0$$
$$[\nabla f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x})]\bar{x} = 0$$
$$\bar{\lambda} G(\bar{x}) = 0 .$$

Observemos que a proposição é òbviamente verdadeira quando $\bar{x}$ é maximizante interior pois nesse caso $\nabla f(\bar{x}) = 0$ e $G(\bar{x}) < 0$ (o que implica $\bar{\lambda} = 0$).

Sendo $\bar{x}$ fronteiro, designe $\alpha(\bar{x})$ o conjunto de índices para os quais $g_i(\bar{x}) = 0$.

Em virtude da condição de regularidade, os pontos x vizinhos de $\bar{x}$ tais que

1) $$\forall\, i \in \alpha \quad \nabla g_i(\bar{x})(x - \bar{x}) \leqq 0,$$
$$\bar{x}_j = 0 \Rightarrow x_j \geqq 0$$

são pontos interiores de C definindo direcções admissíveis $x - \bar{x}$ para as quais

2) $$\nabla f(\bar{x})\,(x - \bar{x}) \leqq 0 .$$

Ambas as desigualdades 1) e 2) são válidas se $\nabla f(\bar{x}) = 0$ ou $\nabla g_i(\bar{x}) = 0$. De 1) e 2) vem

1′) $$\nabla g_i(\bar{x})\, x \leqq \nabla g_i(\bar{x})\, \bar{x}$$

2′) $$\nabla f(\bar{x})\, x \leqq \nabla f(\bar{x})\, \bar{x} .$$

A relação 2′) implica que $\bar{x}$ é maximizante da função linear $\nabla f(\bar{x})x$. Consideremos então o problema de achar $x \geqq 0$ que maximiza

3) $$\nabla f(\bar{x})\, x$$

com as restrições

4) $$\nabla g_i(\bar{x})\, x \leqq \nabla g_i(\bar{x})\, \bar{x} \qquad \forall\, i \in \alpha .$$

Estamos em presença de um programa linear cuja solução é evidentemente $x = \bar{x}$. Tem um programa dual em que se podem utilizar para novas variáveis $\lambda_i \geqq 0$. O dual requer a minimização de

5) $$\sum_{i \in \alpha} \lambda_i [\nabla g_i(\bar{x})\, \bar{x}] ,$$

sujeita às restrições

6) $$\sum_{i \in \alpha} \lambda_i \nabla g_i(\bar{x}) \geqq \nabla f(\bar{x}) .$$

Consequentemente, achar $\bar{x} = (\bar{x}_1, \cdots, \bar{x}_n)$ que maximiza $f(x)$ reduz-se ao problema de encontrar $\lambda_i \geqq 0$ que resolvem o programa linear dual 5) e 6).

Tomando $\lambda_i = 0$ para $i \notin \alpha$, a desigualdade 6) pode escrever-se na forma

7) $$\nabla f(\bar{x}) - \lambda \nabla G(\bar{x}) \leqq 0 .$$

Como

$$[\nabla f(\bar{x}) - \bar{\lambda} \nabla (x)]\bar{x} = \nabla G f(\bar{x})\bar{x} - [\bar{\lambda} \nabla G(\bar{x})]\bar{x},$$

e, pelo teorema da dualidade na programação linear,

$$[\overline{\lambda}\nabla G(\overline{x})]\overline{x} = \nabla f(\overline{x})\overline{x},$$

é claro que

$$8) \qquad [\nabla f(\overline{x}) - \overline{\lambda}\nabla G(\overline{x})]\overline{x} = 0$$

e também é evidente que

$$9) \qquad \overline{\lambda} G(\overline{x}) = 0.$$

É fácil demonstrar que, sendo $f(x)$ côncava e $G(x)$ convexa, as condições do teorema I são também suficientes. De facto, atendendo a que $\overline{\lambda} \geqq 0$ e $G(x) \leqq 0$, vem

$$f(x) \leqq f(x) - \overline{\lambda} G(x)$$

e, como $f(x)$ e $-G(x)$ são côncavas (e diferenciáveis), vem

$$f(x) \leqq f(\overline{x}) + \nabla f(\overline{x})(x - \overline{x})$$
$$-G(x) \leqq -G(\overline{x}) - \nabla G(\overline{x})(x - \overline{x})$$

donde resulta

$$f(x) \leqq f(\overline{x}) + \nabla f(\overline{x})(x - \overline{x}) - \overline{\lambda} G(\overline{x}) - \overline{\lambda}\nabla G(\overline{x})(x - \overline{x}) = f(\overline{x}) + [\nabla f(\overline{x}) - \overline{\lambda}\nabla G(\overline{x})]x - [\nabla f(\overline{x}) - \overline{\lambda}\nabla G(\overline{x})]\overline{x} \leqq f(\overline{x}),$$

isto é,

II. *Sendo $f(x)$ côncava e $G(x)$ convexa, a solução admissível $\overline{x}$ é maximizante de $f(x)$ com as restrições $G(x) \leqq 0$ e $x \geqq 0$ sse existe $\overline{\lambda} \geqq 0$ tal que*

$$\nabla f(\overline{x}) - \overline{\lambda}\nabla G(\overline{x}) \leqq 0$$
$$[\nabla f(\overline{x}) - \overline{\lambda}\nabla G(\overline{x})]\overline{x} = 0$$
$$\overline{\lambda} G(\overline{x}) = 0.$$

As proposições que se demonstraram são teoremas de KUHN-TUCKER. Os escalares $\overline{\lambda}_i \geqq 0$ são conhecidos por *multiplicadores de Kuhn-Tucker*. As condições expressas nestes teoremas têm aparecido na literatura sob várias formas que interessa registar. Abandonando a notação matricial, tem-se:

$$\overline{\lambda}_i \geqq 0 \quad (i = 1, \cdots, m)$$

$$\left(\frac{\partial f}{\partial x_j}\right)_{\overline{x}} - \sum_{i=1}^{m} \overline{\lambda}_i \left(\frac{\partial g_i}{\partial x_j}\right)_{\overline{x}} \leqq 0 \quad (j = 1, , \cdots, n)$$

$$\sum_{j=1}^{n} \left[\left(\frac{\partial f}{\partial x_j}\right)_{\overline{x}} - \sum_{i=1}^{m} \overline{\lambda}_i \left(\frac{\partial g_i}{\partial x_j}\right)_{\overline{x}}\right] \overline{x}_j = 0$$

$$\sum_{i=1}^{m} \overline{\lambda}_i g_i(\overline{x}) = 0 \text{ (ou } \overline{\lambda}_i g_i(\overline{x}) = 0 \text{ para } i = 1, \cdots, m)$$

ou também

$$\overline{x}_j > 0 \Rightarrow \left(\frac{\partial f}{\partial x_j}\right)_{\overline{x}} - \sum_{i=1}^{m} \overline{\lambda}_i \left(\frac{\partial g_i}{\partial x_j}\right)_{\overline{x}} = 0$$

$$\overline{x}_j = 0 \Rightarrow \left(\frac{\partial f}{\partial x_j}\right)_{\overline{x}} - \sum_{i=1}^{m} \overline{\lambda}_i \left(\frac{\partial g_i}{\partial x_j}\right)_{\overline{x}} \leqq 0$$

$$g_i(\overline{x}) < 0 \Rightarrow \overline{\lambda}_i = 0 \quad (\text{ou } \overline{\lambda}_i > 0 \Rightarrow g_i(\overline{x}) = 0).$$

Quando num programa matemático as variáveis são livres, pode fazer-se a transformação $x = x' - x''$ com $x' \geqq 0$ e $x'' \geqq 0$. O programa apresenta-se então na forma

$$\text{maximizar } f(x' - x'')$$

com as restrições

$$G(x' - x'') \leqq 0$$
$$x' \geqq 0 \quad x'' \geqq 0.$$

A primeira condição de KUHN-TUCKER surge na forma ($\overline{\lambda} \geqq 0$)

$$\nabla f(\overline{x}) - \overline{\lambda}\nabla G(\overline{x}) \leqq 0$$
$$-\nabla f(\overline{x}) + \overline{\lambda}\nabla G(\overline{x}) \leqq 0,$$

equivalente a

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

A 2.ª condição é agora dispensável pois é automàticamente satisfeita. Tudo se resume pois às condições

$$\begin{gathered} \overline{\lambda} \geqq 0 \\ \nabla f(\overline{x}) - \overline{\lambda} \nabla G(\overline{x}) = 0 \\ \overline{\lambda} G(\overline{x}) = 0 . \end{gathered}$$

Observando que a restrição $x \geqq 0$ é equivalente a $x_j \geqq 0$ $(j = 1, \cdots, n)$ (ou $-x_j \leqq 0$), todo o programa matemático pode afinal ser apresentado na forma

$$\max f(x)$$

com as restrições

$$G(x) \leqq 0$$

e portanto as condições de KUHN-TUCKER são

$$\begin{gathered} \overline{\lambda} \geqq 0 \\ \nabla f(\overline{x}) - \overline{\lambda} \nabla G(\overline{x}) = 0 \\ \overline{\lambda} G(\overline{x}) = 0 . \end{gathered}$$

Quando aparecem restrições sob a forma de igualdades e desigualdades, isto é,

$$\max f(x)$$

com as restrições

$$\begin{aligned} G_p(x) &\leqq 0 \\ G_q(x) &= 0 \end{aligned}$$

podem considerar-se as restrições na forma

$$\begin{aligned} G_p(x) &\leqq 0 \\ G_q(x) &\leqq 0 \\ -G_q(x) &\leqq 0 . \end{aligned}$$

Tomando

$$G(x) = \begin{bmatrix} G_p(x) \\ G_q(x) \\ -G_q(x) \end{bmatrix},$$

$$\nabla G(x) = \begin{bmatrix} \nabla G_p(x) \\ \nabla G_q(x) \\ -\nabla G_q(x) \end{bmatrix}$$

e

$$\lambda = [\lambda^p \mu^q \nu^q] \geqq 0 ,$$

as condições de KUHN-TUCKER podem exprimir-se do modo seguinte:

$$\begin{gathered} \overline{\lambda} \geqq 0 \\ \nabla f(\overline{x}) - \overline{\lambda}^p \nabla G_p(\overline{x}) - \overline{\mu}^q \nabla G_q(\overline{x}) + \\ + \overline{\nu}^q \nabla G_q(\overline{x}) = 0 \\ \overline{\lambda}^p G_p(\overline{x}) = 0 \end{gathered}$$

ou, simplesmente,

$$\begin{gathered} \nabla f(\overline{x}) - \overline{\lambda}^p \nabla G_p(\overline{x}) - \overline{\lambda}^q G_q(\overline{x}) = 0 \\ \overline{\lambda}^p G_p(\overline{x}) = 0 \end{gathered}$$

com $\overline{\lambda}^p \geqq 0$ e $\overline{\lambda}^q = \overline{\mu}^q - \overline{\nu}^q$ (sem restrição de sinal).

Como caso particular interessante, surge o problema de extremos condicionados clássico:

$$\max f(x)$$

com as restrições

$$G(x) = 0 .$$

As condições de Kuhn-Tucker são agora

$$\nabla f(\overline{x}) - \overline{\lambda} \nabla G(\overline{x}) = 0$$

em que $\overline{\lambda}$ não está sujeito à condição de não-negatividade: as componentes de $\overline{\lambda}$ são os clássicos *multiplicadores de Lagrange*.

Tomando a função lagrangeana

$$F(x,\lambda) = f(x) - \lambda\, G(x),$$

diz-se que o ponto

$$(\bar{x},\bar{\lambda}) = (\bar{x}_1, \cdots, \bar{x}_n, \bar{\lambda}_1, \cdots, \bar{\lambda}_m)$$

é *ponto-sela* de $F(x,\lambda)$ se

$$F(x,\bar{\lambda}) \leqq F(\bar{x},\bar{\lambda}) \leqq F(\bar{x},\lambda).$$

O próximo objectivo consiste em mostrar que, em certas condições, a componente $\bar{x}$ do ponto-sela da função lagrangeana $F(x,\lambda)$ é precisamente a solução do programa.

Definamos os seguintes vectores-linhas:

$$\nabla_x F(x,\lambda) = \left[\frac{\partial F}{\partial x_1}, \cdots, \frac{\partial F}{\partial x_n}\right]$$

$$\nabla_\lambda F(x,\lambda) = \left[\frac{\partial F}{\partial \lambda_1}, \cdots, \frac{\partial F}{\partial \lambda_m}\right] =$$

$$= -[g_1, \cdots, g_m] = -G^T(x)$$

$$\nabla_x F(\bar{x},\lambda) = \nabla_x F(x,\lambda)|_{\bar{x}}$$

$$\nabla_\lambda F(x,\bar{\lambda}) = \nabla_\lambda F(x,\lambda)|_{\bar{\lambda}}.$$

III. *Uma condição necessária para que* $(\bar{x},\bar{\lambda})$ *seja ponto-sela de* $F(x,\lambda)$ *é que* $\bar{x}$ *e* $\bar{\lambda}$ *satisfaçam a*

$$\nabla_x F(\bar{x},\bar{\lambda}) \leqq 0 \qquad \nabla_x F(\bar{x},\bar{\lambda})\cdot\bar{x} = 0 \qquad \bar{x} \geqq 0$$

$$\nabla_\lambda F(\bar{x},\bar{\lambda}) \geqq 0 \qquad \nabla_\lambda F(\bar{x},\bar{\lambda})\cdot\bar{\lambda}^T = 0 \qquad \bar{\lambda} \geqq 0.$$

Estas condições, juntamente com

$$F(x,\bar{\lambda}) \leqq F(\bar{x},\bar{\lambda}) + \nabla_x F(\bar{x},\bar{\lambda})\cdot(x-\bar{x})$$

$$F(\bar{x},\lambda) \geqq F(\bar{x},\bar{\lambda}) + \nabla_\lambda F(\bar{x},\bar{\lambda})\cdot(\lambda-\bar{\lambda})^T$$

são suficientes para que $(\bar{x},\bar{\lambda})$ *seja ponto-sela com* $x \geqq 0$ *e* $\lambda \geqq 0$.

As duas primeiras condições são necessárias. Em relação à primeira, observemos que, para $x \geqq 0$ na região admissível,

$$\nabla_x F(\bar{x},\bar{\lambda})(x-\bar{x}) \leqq 0$$

pois $F(x,\bar{\lambda})$ tem máximo para $x=\bar{x}$. Outra maneira de exprimir esta desigualdade é a seguinte

$$\sum_{i=1}^{n}\left(\frac{\partial F}{\partial x_i}\right)_{\bar{x}} d\,x_i \leqq 0$$

com $d\,x_i = x_i - \bar{x}_i$ $(x_i \geqq 0)$. É possível escolher todos menos um dos x_i iguais a $\bar{x}_i$, seja x_k. Então a expressão dá

$$\left(\frac{\partial F}{\partial x_k}\right)_{\bar{x}}(x_k - \bar{x}_k) \leqq 0.$$

Se $\bar{x}_k = 0$, então $\left(\frac{\partial F}{\partial x_k}\right)_{\bar{x}} \leqq 0$ porque $x_k \geqq 0$; se $\bar{x}_k > 0$, então x_k pode ser escolhido por forma que $x_k > \bar{x}_k$ ou $\bar{x}_k > x_k \geqq 0$ e é evidente que a única maneira de satisfazer a desigualdade para ambas as escolhas de $x_k \geqq 0$ é que

$$\left(\frac{\partial F}{\partial x_k}\right)_{\bar{x}} = 0.$$

Consequentemente, tem-se

$$\nabla_x F(\bar{x},\bar{\lambda}) \leqq 0$$

desde que $\bar{x} \geqq 0$ maximize $F(x,\bar{\lambda})$. Por outro lado, em virtude do que se acaba de ver,

$$\nabla_x F(\bar{x},\bar{\lambda})\cdot\bar{x} = 0.$$

Para a segunda condição seguir-se-ia raciocínio análogo.

A demonstração da condição suficiente é óbvia. De facto, as três primeiras condições permitem concluir imediatamente que

$$F(x,\bar{\lambda}) \leqq F(\bar{x},\bar{\lambda}).$$

As duas primeiras e a quarta dão

$$F(\bar{x},\lambda)\geqq F(\bar{x},\bar{\lambda}).$$

IV. *Uma condição necessária (suficiente) para que* $\bar{x}$ *seja solução óptima do problema de programação matemática é que exista* $\bar{\lambda}$ *tal que* $(\bar{x},\bar{\lambda})$ *satisfaça às condições necessárias (primeiras três condições suficientes) do teorema III.*

Do teorema I resulta que no ponto $\bar{x}$ é

$$\nabla f(\bar{x})\leqq\bar{\lambda}\nabla G(\bar{x})$$

e portanto

$$\nabla_x F(\bar{x},\bar{\lambda})=\nabla f(\bar{x})-\bar{\lambda}\nabla G(\bar{x})\leqq 0$$

$$\nabla_x F(\bar{x},\bar{\lambda})\,\bar{x}=0,$$

o que dá a primeira condição necessária.

De

$$\nabla_\lambda F(\bar{x},\bar{\lambda})=-G^T(\bar{x})\geqq 0$$

e, ainda do Teorema I,

$$\nabla_\lambda F(\bar{x},\bar{\lambda})\,\bar{\lambda}^T=-G^T(\bar{x})\,\bar{\lambda}^T=0,$$

o que prova a segunda condição necessária.

Para provar a suficiência, obtém-se da terceira condição que, para $\bar{\lambda}\geqq 0$, $F(x,\bar{\lambda})=f(x)-\bar{\lambda}G(x)\leqq F(\bar{x},\bar{\lambda})+\nabla_x F(\bar{x},\bar{\lambda})(x-\bar{x})$ e, como

$$\nabla_x F(\bar{x},\bar{\lambda})\,x\leqq 0$$

$$\nabla_x F(\bar{x},\bar{\lambda})\,\bar{x}=0,$$

(teorema III)

vem

$$F(x,\bar{\lambda})\leqq F(\bar{x},\bar{\lambda})=f(\bar{x})-\bar{\lambda}G(\bar{x}).$$

Ainda pelo teorema III

$$-\bar{\lambda}G(\bar{x})=\nabla_\lambda F(\bar{x},\bar{\lambda})\,\bar{\lambda}^T=0,$$

e vem

$$F(x,\bar{\lambda})=f(x)-\bar{\lambda}G(x)\leqq f(\bar{x}).$$

Dado que

$$-\bar{\lambda}G(x)\geqq 0$$

resulta

$$f(x)\leqq f(\bar{x})$$

e que mostra que $\bar{x}$ é maximizante de $f(x)$.

O teorema seguinte, extremamente importante na teoria da programação, exprime que as condições de KUHN-TUCKER são também suficientes se $f(x)$ é côncava e $G(x)$ convexa (1). De facto,

V. *Sendo* $f(x)$ *côncava e* $G(x)$ *convexa, a condição necessária e suficiente para que* $\bar{x}$ *maximize* f *com as restrições* $G(x)\leqq 0$ *e* $x\geqq 0$ *é que* $\bar{x}$ *e* $\bar{\lambda}$ $(\bar{\lambda}\geqq 0)$ *constituam um ponto-sela para a função lagrangeana* $F(x,\lambda)$.

Como $G(x)$ é convexa e $f(x)$ é côncava, vem (teorema VII do n.º 2)

$$G(x)\geqq G(\bar{x})+\nabla G(\bar{x})\,(x-\bar{x})$$
$$f(x)\leqq f(\bar{x})+\nabla f(\bar{x})\,(x-\bar{x})$$

para todo $x,\bar{x}\geqq 0$. Para algum $\bar{\lambda}\geqq 0$,

$$F(x,\bar{\lambda})=f(x)-\bar{\lambda}G(x)\leqq f(\bar{x})-\bar{\lambda}G(\bar{x})+$$
$$+[\nabla f(\bar{x})-\bar{\lambda}\nabla G(\bar{x})]\,(x-\bar{x})=F(\bar{x},\bar{\lambda})+$$
$$+\nabla_x F(\bar{x},\bar{\lambda})\,(x-\bar{x})$$

e a terceira condição suficiente do teorema III é satisfeita.

(1) É uma nova versão do teorema II.

A quarta condição suficiente é cumprida em virtude da linearidade em λ; isto é

$$F(\overline{x}, \lambda) = f(\overline{x}) - \lambda G(\overline{x}) = f(\overline{x}) - \overline{\lambda} G(\overline{x}) - \\ - \lambda G(\overline{x}) + \overline{\lambda} G(\overline{x}) = F(\overline{x}, \overline{\lambda}) - (\lambda - \overline{\lambda}) G(\overline{x}) = \\ = F(\overline{x}, \overline{\lambda}) + \nabla_{\lambda} F(\overline{x}, \overline{\lambda}) (\lambda - \overline{\lambda})^T .$$

Em consequência do teorema IV, todas as condições do teorema III não necessárias e suficientes neste caso. Assim existe equivalência entre a *programação côncava* e o problema do valor-sela.

Desta proposição pode derivar-se o célebre teorema da dualidade da programação linear. De facto, pretendendo se

$$\text{maximizar} \sum_{j=1}^{n} c_j x_j$$

com as restrições

$$\sum_{j=1}^{n} a_{ij} x_j \leqq b_i \quad (i = 1, \cdots, m)$$

$$x_j \geqq 0 \quad (j = 1, \cdots, n),$$

o seu programa dual consiste em

$$\text{minimizar} \sum_{i=1}^{m} b_i y_i \left(\text{maximizar} - \sum_{i=1}^{m} b_i y_i \right)$$

com as restrições

$$\sum_{i=1}^{m} a_{ij} y_i \geqq c_j \quad (j = 1, \cdots, n)$$

$$y_i \geqq 0 \quad (i = 1, \cdots, m).$$

A função lagrangeana para o primal é

$$F_p(x, \lambda) = \sum_{j=1}^{n} c_j x_j - \sum_{i=1}^{m} \lambda_i \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right) \\ = \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m} b_i \lambda_i - \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij} \lambda_i x_j$$

e para o dual é

$$F_d(y, \mu) = - \sum_{i=1}^{m} b_i y_i - \sum_{j=1}^{n} \mu_j \left(c_j - \sum_{i=1}^{m} a_{ij} y_i \right) \\ = - \sum_{i=1}^{m} b_i y_i - \sum_{j=1}^{n} c_j \mu_j + \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij} \mu_j y_i .$$

Vê-se imediatamente que

$$F_p(x, \lambda) = - F_d(\lambda, x)$$

e, portanto, se uma função lagrangeana tem ponto-sela, a outra também tem e

$$F_p(\overline{x}, \overline{\lambda}) = - F_d(\overline{\lambda}, \overline{x}),$$

o que implica

$$\max \sum_{j=1}^{n} c_j x_j = - \max \sum_{i=1}^{m} b_i y_i = \min \sum_{i=1}^{m} b_i y_i .$$

Para os programas matemáticos em geral considera-se *problema primal* aquele que consiste em maximizar a função côncava $f(x)$ sujeita às restrições convexas $G(x) \leqq 0$ e $x \geqq 0$; o *problema dual* consiste em minimizar $F(x, \lambda) = f(x) - \lambda G(x)$ (ou maximizar $-F(x, \lambda)$) com as restrições

$$\nabla_x [f(x) - \lambda G(x)] = 0$$

e

$$\lambda \geqq 0 .$$

Existe um teorema, cuja demonstração será omitida, que estabelece o seguinte:

VI. *Se* $\overline{x}$ *maximiza* $f(x)$ *no problema primal, então existe* $\overline{\lambda} \geqq 0$ *tal que* $(\overline{x}, \overline{\lambda})$ *é solução do problema dual; recìprocamente, se* $(\overline{x}, \overline{\lambda})$ *minimiza* $F(x, \lambda)$ *no problema dual e se a matriz hessiana* $\nabla^2 F(\overline{x}, \overline{\lambda}) = \left[\dfrac{\partial^2 F}{\partial x_i \, \partial x_j} \right]_{(\overline{x}, \overline{\lambda})}$

possui inversa, então $\overline{x}$ *maximiza* $f(x)$ *no problema primal. Em ambos os casos*

$$\max_{x} f(x) = \min_{x,\lambda} F(x, \lambda).$$

Como já se afirmou, as proposições estudadas neste parágrafo são teoremas que, em geral, não se empregam directamente para achar a solução de um programa matemático. No entanto, a sua utilização permite muitas vezes a obtenção de informações sobre a existência do óptimo e até em certos casos simples a sua determinação. Estudemos alguns exemplos.

EXEMPLO 3. 1. Minimizar a função

$$f(x_1, x_2) = \frac{(x_1 + 1)^3}{3} + x_2$$

com as restrições $x_1 \geqq 1$ e $x_2 \geqq 0$.

Observemos que

$$\frac{\partial^2 f}{\partial x_1^2} = 2(x_1 + 1) > 0$$

e

$$\frac{\partial^2 f}{\partial x_1^2} \frac{\partial^2 f}{\partial x_2^2} - \left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 = 0$$

e portanto $f(x_1, x_2)$ é convexa. Podemos então reformular o problema do modo seguinte:

$$\text{maximizar } F(x_1, x_2) = -f(x_1, x_2) = -\frac{(x_1 + 1)^3}{3} - x_2$$

com as restrições

$$1 - x_1 \leqq 0$$

$$x_1 \geqq 0, \quad x_2 \geqq 0.$$

Tomando $\lambda_1 \geqq 0$, as condições de KUHN-TUCKER (necessárias e suficientes) dão

$$-(x_1 + 1)^2 + \lambda_1 \leqq 0$$

$$-1 \leqq 0$$

$$[-(x_1 + 1)^2 + \lambda_1] x_1 + (-1) x_2 = 0$$

$$\lambda_1 (1 - x_1) = 0.$$

Com $\lambda_1 = 0$, teria de ser $x_1 = 0$ (3.ª condição), o que é incompatível com a restrição.

Com $\lambda_1 > 0$, terá de ser $x_1 = 1$, o que implica $-(1+1)^2 + \lambda_1 = 0$ e $x_2 = 0$. Logo, $\lambda_1 = 4$, $x_1 = 1$ e $x_2 = 0$ satisfazem às condições e $(1, 0)$ é minimizante da função dada.

EXEMPLO 3. 2. Maximizar $f(x_1, x_2) = 4x_1 + 6x_2 - x_1^3 - 2x_2^2$ com as restrições

$$x_1 + 3x_2 - 8 \leqq 0$$

$$5x_1 + 2x_2 - 14 \leqq 0$$

$$x_1 \geqq 0, \quad x_2 \geqq 0.$$

Neste caso tem-se

$$\frac{\partial f}{\partial x_1} = 4 - 3x_1^2 \qquad \frac{\partial^2 f}{\partial x_1^2} = -6x_1 \leqq 0$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = 0$$

$$\frac{\partial f}{\partial x_2} = 6 - 4x_2 \qquad \frac{\partial^2 f}{\partial x_2^2} = -4 < 0$$

e portanto a função é côncava pois

$$\frac{\partial^2 f}{\partial x_1^2} \leqq 0$$

e

$$\frac{\partial^2 f}{\partial x_1^2} \frac{\partial^2 f}{\partial x_2^2} - \left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 \geqq 0.$$

Tomando $\lambda_1 \geqq 0$ e $\lambda_2 \geqq 0$, as condições de KUHN-TUCKER dão

$$4 - 3x_1^2 - \lambda_1 - 5\lambda_2 \leqq 0$$
$$6 - 4x_2 - 3\lambda_1 - 2\lambda_2 \leqq 0$$
$$(4 - 3x_1^2 - \lambda_1 - 5\lambda_2)x_1 +$$
$$+ (6 - 4x_2 - 3\lambda_1 - 2\lambda_2)x_2 = 0$$
$$\lambda_1(x_1 + 3x_2 - 8) + \lambda_2(5x_1 + 2x_2 - 14) = 0.$$

Há que considerar os casos seguintes:

λ_1	λ_2
0	0
0	$\neq 0$
$\neq 0$	0
$\neq 0$	$\neq 0$

a) $\lambda_1 = 0$, $\lambda_2 = 0$

$$4 - 3x_1^2 \leqq 0 \Rightarrow x_1 \geqq 2/\sqrt{3}$$

$$6 - 4x_2 \leqq 0 \Rightarrow x_2 \geqq 3/2.$$

A condição

$$(4 - 3x_1^2)x_1 + (6 - 4x_2)x_2 = 0$$

obriga a tomar $x_1 = 2/\sqrt{3}$ e $x_2 = 3/2$, ponto que, satisfazendo às restrições, é o maximizante procurado (aliás é maximizante interior).

b) $\lambda_1 = 0$, $\lambda_2 \neq 0$

$$4 - 3x_1^2 - 5\lambda_2 \leqq 0$$
$$6 - 4x_2 - 2\lambda_2 \leqq 0$$
$$(4 - 3x_1^2 - 5\lambda_2)x_1 + (6 - 4x_2 - 2\lambda_2)x_2 = 0$$
$$\lambda_2(5x_1 + 2x_2 - 14) = 0.$$

Terá de ser $5x_1 + 2x_2 - 14 = 0$, o que impede $x_1 = x_2 = 0$. Com $x_1 = 0$, $x_2 = 7$ viria $6 - 28 - 2\lambda_2 = 0$ o que é impossível; com $x_2 = 0$, $x_1 = 14/5$, viria $4 - 3(14/5)^2 - 5\lambda_2 = 0$ também impossível; com $x_1 > 0$, $x_2 > 0$ teria de ser

$$\begin{cases} 5x_1 + 2x_2 - 14 = 0 \\ 4 - 3x_1^2 - 5\lambda_2 = 0 \\ 6 - 4x_2 - 2\lambda_2 = 0 \end{cases}$$

$$\begin{cases} 5x_1 + 3 - \lambda_2 - 14 = 0 \\ 4 - 3x_1^2 - 5\lambda_2 = 0 \\ x_2 = \dfrac{3 - \lambda_2}{2} \end{cases} \qquad \begin{cases} x_1 = \dfrac{\lambda_2 + 11}{5} \\ x_2 = \dfrac{3 - \lambda_2}{2} \end{cases}$$

$$4 - 3\left(\frac{\lambda_2 + 11}{5}\right)^2 - 5\lambda_2 = 0 \Rightarrow -63 - 191\lambda_2 - 3\lambda_2^2 = 0$$ (impossível com $\lambda_2 > 0$).

c) $\lambda_1 \neq 0$, $\lambda_2 = 0$

$$4 - 3x_1^2 - \lambda_1 \leqq 0$$
$$6 - 4x_2 - 3\lambda_1 \leqq 0$$
$$x_1(4 - 3x_1^2 - \lambda_1) + x_2(6 - 4x_2 - 3\lambda_1) = 0$$
$$\lambda_1(x_1 + 3x_2 - 8) = 0.$$

Terá de ser $x_1 + 3x_2 - 8 = 0$ o que impede $x_1 = x_2 = 0$. Com $x_1 = 0$, $x_2 = 8/3$ vem $6 - \frac{32}{3} - 3\lambda_1 = 0$ o que é impossível para $\lambda_1 > 0$; com $x_1 = 8$, $x_2 = 0$, vem $4 - 3(8)^2 - \lambda_1 = 0$ também impossível; resta $x_1 > 0$, $x_2 > 0$ que implica

$$\begin{cases} x_1 + 3x_2 - 8 = 0 \\ 4 - 3x_1^2 - \lambda_1 = 0 \\ 6 - 4x_2 - 3\lambda_1 = 0 \end{cases} \qquad \begin{cases} x_1 = \dfrac{14 + 9\lambda_1}{4} \\ x_2 = \dfrac{6 - 3\lambda_1}{4} \end{cases}$$

$$4 - 3\left(\frac{14 + 9\lambda_1}{4}\right)^2 - \lambda_1 = 0$$ é impossível com $\lambda_1 > 0$.

d) $\lambda_1 \neq 0$, $\lambda_2 \neq 0$

$$\begin{cases} x_1 + 3x_2 - 8 = 0 \\ 5x_1 + 2x_2 - 14 = 0 \end{cases} \quad \begin{cases} x_1 = 2 \\ x_2 = 2 \end{cases}$$

$$\begin{cases} 4 - 3x_1^2 - \lambda_1 - 5\lambda_2 = 0 \\ 6 - 4x_2 - 3\lambda_1 - 2\lambda_2 = 0 \end{cases}$$

$$\begin{cases} -8 - \lambda_1 - 5\lambda_2 = 0 \\ -2 - 3\lambda_1 - 2\lambda_2 = 0 \end{cases} \quad \begin{cases} \lambda_1 + 5\lambda_2 = -8 \\ 3\lambda_1 + 2\lambda_2 = -2 \end{cases}$$

$$\begin{cases} -3\lambda_1 - 15\lambda_2 = 24 \\ 3\lambda_1 + 2\lambda_2 = -2 \end{cases} \quad \begin{cases} \cdots\cdots \\ -13\lambda_2 = 22 \end{cases} \text{ imp.}$$

Observemos que a análise dos casos *b*), *c*) e *d*) só se justifica para encontrar pontos que dão o mesmo máximo para $f(x_1, x_2)$. O máximo relativo para uma função côncava é também máximo absoluto.

EXEMPLO 3. 3. Minimizar $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 + 2(x_1 - 5) + 2(x_1 + x_2 - 10)$ com as restrições

$$x_1 \geqq 5$$
$$x_1 + x_2 \geqq 10$$
$$x_1 + x_2 + x_3 \geqq 15$$
$$x_j \geqq 0 \quad (j = 1, 2, 3).$$

O problema equivale a maximizar

$$F(x_1, x_2, x_3) = -f(x_1, x_2, x_3) = -x_1^2 - x_2^2 - x_3^2 - 2(x_1 - 5) - 2(x_1 + x_2 - 10),$$

que é côncava, com restrições convexas.

As condições de KUHN-TUCKER (necessárias e suficientes) dão ($\lambda_i \geqq 0$)

$$\begin{gathered} -2x_1 - 4 + \lambda_1 + \lambda_2 + \lambda_3 \leqq 0 \\ -2x_2 - 2 + \lambda_2 + \lambda_3 \leqq 0 \\ -2x_3 + \lambda_3 \leqq 0 \\ (-2x_1 - 4 + \lambda_1 + \lambda_2 + \lambda_3)x_1 + \\ + (-2x_2 - 2 + \lambda_2 + \lambda_3)x_2 + \\ + (-2x_3 + \lambda_3)x_3 = 0 \end{gathered}$$

$$\begin{gathered} \lambda_1(5 - x_1) + \lambda_2(10 - x_1 - x_2) + \\ + \lambda_3(15 - x_1 - x_2 - x_3) = 0. \end{gathered}$$

Aqui cada um dos λ_i pode ser zero ou diferente de zero e portanto, em princípio, é preciso examinar um total de $2^3 = 8$ casos. Consideremos o caso $\lambda_1 \neq 0$, $\lambda_2 \neq 0$, $\lambda_3 \neq 0$. É evidente que terão de ser satisfeitas as condições

$$\begin{cases} 5 - x_1 = 0 \\ 10 - x_1 - x_2 = 0 \\ 15 - x_1 - x_2 - x_3 = 0 \end{cases}$$

que dão $x_1 = 5$, $x_2 = 5$ e $x_3 = 5$. Como os x_j são não-nulos, terão de ser satisfeitas as condições

$$\begin{cases} -2x_1 - 4 + \lambda_1 + \lambda_2 + \lambda_3 = 0 \\ -2x_2 - 2 + \lambda_2 + \lambda_3 = 0 \\ -2x_3 + \lambda_3 = 0 \end{cases}$$

que dão $\lambda_1 = \lambda_2 = 2$ e $\lambda_3 = 10$. Está-se pois em presença da solução óptima.

EXEMPLO 3. 4. Maximizar $f(x_1, x_2) = 4x_1 + 5x_2 + x_1x_2 - x_1^2 - x_2^2 + 5$ com a restrição

$$\frac{x_1}{2} + x_2 \leqq k.$$

A função $f(x_1, x_2)$ é côncava pois

$$\frac{\partial^2 f}{\partial x_1^2} = -2 < 0, \quad \frac{\partial^2 f}{\partial x_1^2}\frac{\partial^2 f}{\partial x_2^2} - \left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 = 4 - 1 = 3 > 0.$$

As condições de KUHN-TUCKER dão ($\lambda_1 \geqq 0$)

$$4 + x_2 - 2x_1 - \frac{\lambda_1}{2} = 0$$

$$5 + x_1 - 2x_2 - \lambda_1 = 0$$

$$\lambda_1\left(\frac{x_1}{2} + x_2 - k\right) = 0.$$

Se $\lambda_1 = 0$, as duas primeiras condições dão $x_1 = 13/3$, $x_2 = 14/3$ que satisfazem ao problema se $k \geqq 41/6$ e não satisfazem se $k < 41/6$.

Com $\lambda_1 > 0$, terá de ser satisfeita a condição $\frac{x_1}{2} + x_2 - k = 0$ juntamente com as duas primeiras, isto é,

$$\begin{cases} 4 + x_2 - 2x_1 - \frac{\lambda_1}{2} = 0 \\ 5 + x_1 - 2x_2 - \lambda_1 = 0 \\ \frac{x_1}{2} + x_2 - k = 0. \end{cases}$$

Resolvendo este sistema em ordem a x_1, x_2 e λ_1, vem

$$\begin{cases} x_1 = \frac{3}{7} + \frac{4}{7}k \\ x_2 = -\frac{3}{14} + \frac{5}{7}k \\ \lambda_1 = \frac{41}{7} - \frac{6}{7}k \end{cases}$$

e, como $\lambda_1 > 0$, terá de ser $k < 41/6$.

Em resumo, com $k \geqq 41/6$, a solução óptima é

$$x_1 = 13/3, \quad x_2 = 14/3, \quad \lambda_1 = 0$$

e, com $k < 41/6$, vêm as soluções

$$x_1 = \frac{3}{7} + \frac{4}{7}k, \quad x_2 = -\frac{3}{14} + \frac{5}{7}k,$$

$$\lambda_1 = \frac{41}{7} - \frac{6}{7}k > 0.$$

EXEMPLO 3.5. Maximizar $f(x_1, x_2) = -8x_1 + 10x_2 - 2x_1^2 - x_2^4$ com as restrições

$$2e^{x_1} + x_1 x_2 + x_2^2 \leqq 10$$

$$x_1 \geqq 0, \quad x_2 \geqq 0$$

Observemos que

$$\frac{\partial^2 f}{\partial x_1^2} = -4 < 0 \quad \text{e} \quad \frac{\partial^2 f}{\partial x_1^2}\frac{\partial^2 f}{\partial x_2^2} - \left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 = 48x_2^2 \geqq 0$$

e portanto $f(x_1, x_2)$ é côncava. Também é fácil ver que $g(x_1, x_2) = 2e^{x_1} + x_1 x_2 + x_2^2$ é convexa na região admissível pois

$$\frac{\partial^2 g}{\partial x_1^2} = 2e^{x_1} > 0 \quad \text{e} \quad \frac{\partial^2 g}{\partial x_1^2}\frac{\partial^2 g}{\partial x_2^2} - \left(\frac{\partial^2 g}{\partial x_1 \partial x_2}\right)^2 = 4e^{x_1} - 1 > 0$$

para $x_1 \geqq 0$.

As condições de KUHN-TUCKER são pois necessárias e suficientes para a existência de máximo. Neste caso ter-se-á $(\lambda_1 \geqq 0)$

$$\frac{\partial f}{\partial x_1} - \lambda_1 \frac{\partial g}{\partial x_1} \leqq 0$$

$$\frac{\partial f}{\partial x_2} - \lambda_2 \frac{\partial g}{\partial x_2} \leqq 0$$

$$\left(\frac{\partial f}{\partial x_1} - \lambda_1 \frac{\partial g}{\partial x_1}\right) x_1 + \left(\frac{\partial f}{\partial x_2} - \lambda_1 \frac{\partial g}{\partial x_2}\right) x_2 = 0$$

$$\lambda_1 g(x_1, x_2) = 0$$

ou seja

$$8 - 4x_1 - \lambda_1(2e^{x_1} + x_2) \leqq 0$$

$$10 - 4x_2^3 - \lambda_1(x_1 + 2x_2) \leqq 0$$

$$[8 - 4x_1 - \lambda_1(2e^{x_1} + x_2)]x_1 + [10 - 4x_2^3 - \lambda_1(x_1 + 2x_2)]x_2 = 0$$

$$\lambda_1(2e^{x_1} + x_1 x_2 + x_2^2 - 10) = 0$$

Com $\lambda_1 = 0$, teria de ser $\begin{cases} 8 - 4x_1 \leqq 0 \\ 10 - 4x_2^5 \leqq 0 \end{cases}$ ou $\begin{cases} x_1 \geqq 2 \\ x_2 \geqq \sqrt[5]{5/2} \end{cases}$, o que implicaria $x_1 = 2$ e $x_2 = \sqrt[5]{5/2}$, pois tem de ser satisfeita a 3.ª condição. Ora o ponto $(2, \sqrt[5]{5/2})$ não satisfaz à restrição.

Terá pois de ser $\lambda_1 > 0$ o que implica $2e^{x_1} + x_1 x_2 + x_2^2 - 10 = 0$. Se fosse $x_1 = 0$, viria $x_2 = 2\sqrt{2}$ e teria de ser satisfeita a condição $10 - 4(2\sqrt{2})^5 - \lambda_1 4\sqrt{2} = 0$ ou $10 - 64\sqrt{2} - \lambda_1 4\sqrt{2} = 0$ que não é compatível com $\lambda_1 > 0$; se fosse $x_2 = 0$, viria $x_1 = \log 5$ e teria de ser satisfeita a condição $8 - 4\log 5 - 10\lambda_1 = 0$ o que é possível com $\lambda_1 = \frac{4 - 2\log 5}{5} > 0$ mas a condição $10 - 4x_2^5 - \lambda_1 (x_1 + 2x_2) = 0$ não é satisfeita pois $10 - \frac{4 - 2\log 5}{5} \log 5 > 0$. Então terá de ser $x_1 > 0$ e $x_2 > 0$, com $\lambda_1 > 0$, isto é, terá de ser satisfeito o sistema

$$\begin{cases} 8 - 4x_1 - \lambda_1 (2e^{x_1} + x_2) = 0 \\ 10 - 4x_2^5 - \lambda_1 (x_1 + 2x_2) = 0 \\ 2e^{x_1} + x_1 x_2 + x_2^2 - 10 = 0 \end{cases}$$

que infelizmente não é constituído por equações lineares. Eis um exemplo em que é difícil obter a solução óptima a partir das condições de KUHN-TUCKER.

A fim de se avaliar a importância da condição de regularidade de KUHN-TUCKER consideremos o seguinte problema:

EXEMPLO 3.6. Maximizar $f(x) = x_1$ com as restrições

$$g_1(x_1, x_2) = x_2 - (x_1 - 1)^2 \leqq 0$$

$$x_1 \geqq 0 \quad x_2 \geqq 0.$$

A região admissível apresentada já na fig. 1 mostra que o máximo de $f(x)$ é atingido para $x_1 = 1$, $x_2 = 0$. No entanto, as condições de KUHN-TUCKER não são satisfeitas nesse ponto pois nele, como se viu já (n.º 3), não é verificada a condição de regularidade.

De facto, teria de ser

$$\frac{\partial f}{\partial x_1} - \lambda_1 \frac{\partial g_1}{\partial x_1} \leqq 0$$

$$\frac{\partial f}{\partial x_2} - \lambda_1 \frac{\partial g_1}{\partial x_2} \leqq 0$$

$$x_1 \left(\frac{\partial f}{\partial x_1} - \lambda_1 \frac{\partial g_1}{\partial x_1} \right) + x_2 \left(\frac{\partial f}{\partial x_2} - \lambda_1 \frac{\partial g_1}{\partial x_2} \right) = 0$$

$$\lambda_1 g_1(x_1, x_2) = 0$$

ou

$$1 + \lambda_1 2(x_1 - 1) \leqq 0$$

$$-\lambda_1 \leqq 0$$

$$x_1 [1 + \lambda_1 2(x_1 - 1)] + x_2 (-\lambda_1) = 0$$

$$\lambda_1 [x_2 - (x_1 - 1)^2] = 0.$$

Com $\lambda_1 > 0$, viria $x_2 - (x_1 - 1)^2 = 0$ e $x_2 = 0$, isto é, o ponto $(1, 0)$ mas este não satisfaz às 1.ª e 3.ª condições.

Com $\lambda_1 = 0$, a 1.ª condição não é satisfeita.

4. Interpretação económica

4.1. Interpretação dos multiplicadores de Kuhn-Tucker

Consideremos o programa matemático na forma

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

com as restrições

1) $g_i(x) \leqq 0 \quad (i = 1, \cdots, m)$.

Como se viu, as condições de KUHN-TUCKER associadas a este programa podem apresentar-se do modo seguinte:

2) $\lambda_i \geqq 0 \quad (i - 1, \cdots, m)$

3) $\dfrac{\partial f}{\partial x_j} - \sum_{i=1}^{m} \lambda_i \dfrac{\partial g_i}{\partial x_j} = 0 \quad (j = 1, \cdots, n)$

4) $\lambda_i g_i(x) = 0 \quad (i = 1, \cdots, m)$.

Em certas condições, se x^0 é solução óptima do programa matemático, então existe um vector $\lambda^0 = (\lambda_1^0, \cdots, \lambda_m^0)$ tal que (x^0, λ^0) satisfaz 1), 2), 3) e 4). Sendo $f(x)$ côncava e $g_i(x)$ convexas, as condições também são suficientes.

Pode parecer à primeira vista que estas condições têm reduzido interesse prático para o gestor. Não é raro deparar com a opinião de que, no domínio da programação matemática, basta saber na prática formular o problema e resolvê-lo numèricamente. Ora, de facto, isto não é suficiente na prática da gestão e para o provar consideremos o problema seguinte:

Suponhamos que o gestor, por meio de um investimento adicional, substitui as restrições 1) por

1') $g_i(x) \leqq b_i \quad (i = 1, \cdots, m) \quad (b_i > 0)$

e admitamos que a passagem de 0 a um pequeno valor positivo b_i exige o investimento $\pi_i b_i$. Põe-se o problema de saber se este investimento é rendável ou não.

Para responder a esta questão consideremos o novo programa matemático em que 1) é substituído por 1') e portanto 4) por

4') $\lambda_i[g_i(x) - b_i] = 0 \quad (i = 1, \cdots, m)$.

Admitindo que as $m + n$ equações 3) e 4') nas $m + n$ incógnitas $\lambda_1, \cdots, \lambda_m, x_1, \cdots, x_n$ dão as funções $\lambda_i(b)$, $x_j(b)$ contínuas e deriváveis, seja $\varphi(b)$ o máximo de $f(x)$ (função de rendimento líquido) com as restrições 1'). Tomando o ponto $b = 0$, vem de 3)

$$df - \sum_{i=1}^{m} \lambda_i^0 \, dg_i = 0$$

e de 4')

$$d\lambda_i g_i(x^0) + \lambda_i^0 \, dg_i - \lambda_i^0 \, db_i = 0 .$$

Se $g_i(x^0) = g_i(x(0)) < 0$, então $g_i(x(b)) < b_i$ numa vizinhança de $b = 0$, de modo que $\lambda_i(b) = 0$ nessa vizinhança, o que implica $d\lambda_i g_i(x^0) = 0$; a mesma conclusão é verdadeira se $g_i(x^0) = 0$. Logo, tem-se

$$\lambda_i^0 \, dg_i = \lambda_i^0 \, db_i$$

e portanto

$$df = \sum_{i=1}^{m} \lambda_i^0 \, db_i ,$$

isto é,

5) $$\left(\frac{\partial \varphi}{\partial b_i}\right)_{b=0} = \lambda_i^0 .$$

Vê-se assim que o multiplicador λ_i^0 representa o valor marginal correspondente à relaxação da i-ésima restrição ($\lambda_i^0 = 0$ se $g_i(x^0) < 0$), supondo evidentemente que não existem quaisquer restrições para o ajustamento de x^0.

O acréscimo $\Delta\varphi$ do lucro máximo é, considerando apenas os termos de primeira ordem,

$$\Delta\varphi = \lambda_1^0 b_1 + \cdots + \lambda_m^0 b_m .$$

A resposta ao problema inicialmente proposto está dada. A rendabilidade do investi-

mento avalia-se comparando $\Delta\varphi$ com $\sum_{i=1}^{m} \pi_i b_i$.

Como se vê, os multiplicadores λ_i representam verdadeiros *preços contabilísticos (preços-sombra* ou *pseudo-preços)* dos vários recursos.

4. 2. Interpretação das condições de Kuhn-Tucker no âmbito da economia da empresa

Admitamos que uma empresa possui n actividades e representemos um programa por um vector de níveis de actividade $x = (x_1, \cdots, x_n)$. Fixemos a hipótese de que a empresa deseja maximizar uma função de rendimento líquido $f(x)$, supondo que existem m restrições sobre os níveis das actividades

$$g_i(x) \leqq 0 \qquad (i = 1, \cdots, m)$$
$$x \geqq 0 .$$

A interpretação económica das derivadas parciais de $f(x)$ e $g_i(x)$ é a seguinte:

$\frac{\partial f}{\partial x_j}$ = rendimento marginal da j-ésima actividade

$\frac{\partial g_i}{\partial x_j}$ = efeito marginal da j ésima actividade na i-ésima restrição.

Dispondo de um conjunto de valores não-negativos $\lambda_1, \cdots, \lambda_m$ (preços contabilísticos) associados às m restrições é evidente que

$$\sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j}$$

representa o custo imputado marginal da j-ésima actividade.

As condições de KUHN-TUCKER estabelecem que, sendo $\bar{x}$ a solução maximizante, existem valores imputados não-negativos $\bar{\lambda}_i$ tais que são cumpridas as condições seguintes:

1. $g_i(\bar{x}) \leqq 0$

   isto é, $\bar{x}$ é admissível.

2. $g_i(\bar{x}) < 0 \Rightarrow \bar{\lambda}_i = 0$

   isto é, se a i-ésima restrição é inactiva, então o valor imputado associado é zero.

3. $\left(\frac{\partial f}{\partial x_j}\right)_{\bar{x}} - \sum_{i=1}^{m} \bar{\lambda}_i \left(\frac{\partial g_i}{\partial x_j}\right)_{\bar{x}} \leqq 0$

   isto é, o rendimento líquido marginal da j-ésima actividade é inferior ou igual ao seu custo imputado.

4. $\left(\frac{\partial f}{\partial x_j}\right)_{\bar{x}} - \sum_{i=1}^{m} \bar{\lambda}_i \left(\frac{\partial g_i}{\partial x_j}\right)_{\bar{x}} < 0 \Rightarrow \bar{x}_j = 0$

   o que significa que, sendo o rendimento líquido marginal da j-ésima actividade inferior ao seu custo imputado marginal, a j-ésima actividade não é utilizada.

Os teoremas que envolvem a função lagrangeana também possuem interpretação interessante como se vai ver por meio de um exemplo.

Admitamos que uma indústria pretende maximizar o lucro $f(x)$ e que a produção satisfaz a certo conjunto C de restrições no espaço n-dimensional. Suponhamos que o governo pretende que algumas restrições suplementares

$$g_i(x) \leqq 0 \quad (i = 1, \cdots, m)$$

sejam satisfeitas. Para esse efeito pode ser criado um sistema de impostos e subsídios por forma a funcionar do modo seguinte: se a indústria viola alguma restrição, isto é, se $g_i(x) > 0$ para algum i, então é aplicado

um imposto $\lambda_i g_i(x)$; se a produção é tal que $g_i(x) < 0$ para algum i, então o governo paga um subsídio $-\lambda_i g_i(x)$ (supõe-se $\lambda_i \geqq 0$). O problema da indústria consiste agora em maximizar (em x, para um dado vector $\lambda \geqq 0$) a função lagrangeana

$$F(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x)$$

com $x \in C$, enquanto o problema do governo se reduz a minimizar (em λ) a mesma função $F(x, \lambda)$ sujeita a $\lambda \geqq 0$. A solução óptima é evidentemente o ponto-sela de $F(x, \lambda)$. Na prática, o diálogo entre o governo e a indústria pode estabelecer-se do modo seguinte:

1. O governo fixa um conjunto de taxas e subsídios $\lambda_1^0 \geqq 0, \cdots, \lambda_m^0 \geqq 0$.
2. A indústria maximiza $F(x, \lambda^0)$ com $x \in C$ e propõe ao Governo a solução óptima x^0.
3. O governo examina cada $g_i(x^0)$ $(i = 1, \cdots, m)$:
   - 3.1 Se $g_i(x^0) > 0$, então o imposto não é suficientemente elevado; λ_i^0 deve ser substituído por $\lambda_i^1 > \lambda_i^0$;
   - 3.2 Se $g_i(x^0) < 0$ e $\lambda_i^0 > 0$, então é possível reduzir o subsídio, substituindo λ_i^0 por $0 \leqq \lambda_i^1 < \lambda_i^0$.
   - 3.3 Se $g_i(x^0) \leqq 0$ e $\lambda_i^0 g_i(x^0) = 0$, então $\lambda_i^1 = \lambda_i^0$.

O diálogo repete-se então com λ^1 em vez de λ^0.

4.3. Contribuição da programação matemática para a reformulação das teorias matemáticas do consumo e da produção

Como se sabe, o problema fundamental na teoria matemática do consumo consiste em o indivíduo maximizar a sua função de utilidade $u(x_1, \cdots, x_n)$ com a restrição $\sum_{j=1}^{n} p_j x_j = M$. É esta a formulação clássica que conduz à célebre lei da igualdade das utilidades marginais ponderadas deduzida através do método dos multiplicadores de LAGRANGE. Mas, não impondo a restrição, $x \geqq 0$, o indivíduo arrisca-se a encontrar uma solução sem significado económico. Por exemplo, se

$$u = (1 + x_1)(1 + x_2)$$

e

$$4x_1 + x_2 = 1,$$

o ponto óptimo que se encontra é $x^* = (-1/4, 2)$ que dá $u^* = 9/4$. A solução é inaceitável sob o ponto de vista económico pois $x_1^* = -1/4 < 0$.

Ora, pondo o problema sob a forma (1)

$$\max u(x)$$

com as restrições

$$px = M$$

$$x \geqq 0$$

a teoria atrás desenvolvida permite chegar à solução satisfatória.

Notemos que as condições de KUHN-TUCKER para o problema de programação

$$\max f(x)$$

com as restrições

$$G(x) = 0$$

$$x \geqq 0$$

(1) Tomou-se $p = [p_1\, p_2 \cdots p_n]$.

reduzem-se a

$$\nabla f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x}) \leqq 0$$
$$[f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x})]\,\bar{x} = 0,$$

com $\bar{\lambda}$ sem restrição de sinal (1).

Para o caso presente é

$$G(x) = \sum_{j=1}^{n} p_j x_j - M = 0$$

e portanto vem

$$\left(\frac{\partial u}{\partial x_j}\right)_{\bar{x}} - \lambda p_j \leqq 0 \quad (j = 1, \cdots, n)$$

$$\sum_{j=1}^{n} \left[\left(\frac{\partial u}{\partial x_j}\right)_{\bar{x}} - \lambda p_j\right] \bar{x}_j = 0.$$

A interpretação económica de λ é aqui $\frac{d u^*}{d M}$ (utilidade marginal da moeda ou, mais pròpriamente, utilidade marginal do rendimento). As condições de KUHN-TUCKER estabelecem então que, sendo $\bar{x}$ a solução maximizante,

1. $\left(\frac{\partial u}{\partial x_j}\right)_{\bar{x}} \Big/ p_j \leqq \bar{\lambda}$

   isto é, a utilidade marginal ponderada do j-ésimo bem é inferior ou igual à utilidade marginal da moeda.

2. $\left(\frac{\partial u}{\partial x_j}\right)_{\bar{x}} \Big/ p_j < \bar{\lambda} \Rightarrow \bar{x}_j = 0$

   isto é, sendo a utilidade marginal ponderada do j-ésimo bem inferior à utilidade marginal da moeda, o j-ésimo bem não é consumido.

Sendo $j_1, \cdots, j_s$ os bens consumidos, tem-se então

$$\frac{\left(\frac{\partial u}{\partial x_{j_1}}\right)_{\bar{x}}}{p_{j_1}} = \frac{\left(\frac{\partial u}{\partial x_{j_2}}\right)_{\bar{x}}}{p_{j_2}} = \cdots = \frac{\left(\frac{\partial u}{\partial x_{j_s}}\right)_{\bar{x}}}{p_{j_s}} = \bar{\lambda}$$

e para os bens não consumidos é

$$\left(\frac{\partial u}{\partial x_j}\right)_{\bar{x}} \Big/ p_j \leqq \bar{\lambda}.$$

Para a função de utilidade

$$u = (1 + x_1)(1 + x_2),$$

com

$$4 x_1 + x_2 = 1,$$

as condições de KUHN-TUCKER fornecem a solução óptima $\bar{x} = (0, 1)$ para a qual se tem o conjunto de valores

$$\left(\frac{\partial u}{\partial x_1}\right)_{(0,1)} = 2, \qquad \left(\frac{\partial u}{\partial x_2}\right)_{(0,1)} = 1, \quad \lambda = 1$$

que satisfazem às proposições enunciadas:

$$\left(\frac{\partial u}{\partial x_2}\right)_{(0,1)} \Big/ 1 = 1 = \lambda$$

$$\left(\frac{\partial u}{\partial x_1}\right)_{(0,1)} \Big/ 4 = 1/2 < \lambda.$$

---

(1) O problema equivale a maximizar $f(x)$ com as restrições

$$G(x) \leqq 0$$
$$-G(x) \leqq 0$$
$$x \geqq 0.$$

Então, com $\bar{\lambda}_1 \geqq 0$ e $\bar{\lambda}_2 \geqq 0$, as condições de KUHN-TUCKER reduzem-se a

$$\nabla f(\bar{x}) - \bar{\lambda}_1 \nabla G(\bar{x}) + \bar{\lambda}_2 \nabla G(\bar{x}) \leqq 0$$
$$[\nabla f(\bar{x}) - \bar{\lambda}_1 \nabla G(\bar{x}) + \bar{\lambda}_2 \nabla G(\bar{x})]\,\bar{x} = 0$$

ou, simplesmente,

$$\nabla f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x}) \leqq 0$$
$$[\nabla f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x})]\,\bar{x} = 0,$$

com $\bar{\lambda} = \bar{\lambda}_1 - \bar{\lambda}_2$ sem restrição de sinal.

O problema fundamental da teoria matemática da produção também pode agora ser formalizado com maior rigor:

Maximizar a função de produção $f(x)$

com a restrição de custo

$$p\,x = C$$

e

$$x \geqq 0\,.$$

Os resultados são formalmente idênticos aos que se obtiveram para a teoria do consumo. O multiplicador λ é agora

$$\lambda = \frac{df^*}{dC}$$

que é o inverso do custo marginal.

5. Programação quadrática

Sob o ponto de vista da existência de algoritmos eficientes para a pesquisa da solução óptima, o caso especial da programação quadrática reveste-se de especial interesse e por isso será aqui discutido.

Todos os vectores-filas considerados nesta secção são tomados como vectores colunas, excluindo o gradiente de $f(x)$ que, tal como anteriormente, será o vector-linha

$$\left[\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_n}\right].$$

Entende-se por programação quadrática o problema que consiste em maximizar a função côncava

$$f(x) = c^T x - x^T A\,x$$

sujeita às restrições lineares

$$B\,x \leqq b$$
$$x \geqq 0$$

onde A é matriz simétrica de ordem n semi-definida positiva, B é matriz $m \times n$,

$$c = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix},\quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{ e } b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

Aliás o problema pode formular-se de três maneiras diferentes:

1) $\max\{f(x) : B\,x \leqq b \wedge x \geqq 0\}$

2) $\max\{f(x) : B\,x = b \quad x \geqq 0\}$

3) $\max\{f(x) : B\,x \leqq b\}$.

Em 2) as inequações transformaram-se em equações mediante a introdução de variáveis auxiliares e em 3) as restrições $x_j \geqq 0$ (se existem) foram integradas no sistema de restrições $A\,x \leqq b$.

Como se disse no n.º 1, não existem actualmente algoritmos eficientes para a pesquisa da solução óptima de um programa matemático geral. Os melhores métodos de cálculo conhecidos até hoje são os que permitem resolver um programa quadrático. Alguns deles serão expostos seguidamente.

5.1. Método de Wolfe

O método que se vai descrever é também conhecido por *método do simplex para a programação quadrática* e apoia-se nas condições de KUHN-TUCKER.

Adoptando a formulação

$$\max f(x) = c^T x - x^T A\,x$$

com as restrições

$$B\,x = b$$
$$x \geqq 0,$$

as condições de KUHN-TUCKER garantem

que $\bar{x}$ é solução óptima sse existe $y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$, não sujeito à condição de não-negatividade, tal que

$$\nabla f(\bar{x}) - y^T \nabla G(\bar{x}) \leqq 0 \qquad (G(\bar{x}) = B\bar{x} - b = 0)$$
$$[\nabla f(\bar{x}) - y^T \nabla G(\bar{x})]\bar{x} = 0$$

ou, fazendo

$$\nabla f(\bar{x}) - y^T \nabla G(\bar{x}) = -v^T,$$

com

$$v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \geqq 0,$$

$$c^T - 2\bar{x}^T A - y^T B = -v^T$$
$$-v^T \bar{x} = 0$$

que se pode escrever ainda na forma

$$2A\bar{x} + B^T y - v = c$$
$$v^T \bar{x} = 0.$$

O problema de programação quadrática pode pois resolver-se, procurando $x \geqq 0$ e $v \geqq 0$ que satisfaçam às restrições

$$2Ax + B^T y - v = c$$
$$Bx = b$$
$$v^T x = 0.$$

Escrevamos

$$Bx + E_1 z_1 = b$$
$$2Ax + B^T y - v + E_2 z_2 = c$$

onde z_1 é vector com m componentes e z_2 é vector com n componentes. As matrizes E_1 e E_2 são matrizes diagonais cujos elementos principais são ± 1 por forma que todas as componentes de z_1 e z_2 sejam não-negativas quando $x = y = v = 0$. Utiliza-se depois o método do simplex para minimizar a soma das componentes de z_1 mais a soma das componentes de z_2 (se houver solução óptima para o programa quadrático, esse mínimo é zero) com as restrições $x \geqq 0$, $v \geqq 0$, $z_1 \geqq 0$, $z_2 \geqq 0$ e com uma regra adicional: para $j = 1, \cdots, n$, se x_j está na base não se admite v_j; se v_j está na base não se admite x_j. Se existe solução óptima, as componentes de x da solução (se existe) resolvem o problema de programação quadrática.

EXEMPLO 5. 1. Maximizar $f(x) = 4x_1 + 2x_2 - x_1^2 - x_2^2$ com as restrições

$$x_1 + x_2 = 2$$
$$x_1 \geqq 0, \quad x_2 \geqq 0.$$

Neste caso, tem-se

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad B = [1 \;\; 1]$$

$$b = [2] \quad c = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

e portanto

$$[1 \;\; 1] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + z_{11} = 2$$

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} [y_1] - \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} z_{21} \\ z_{22} \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix},$$

isto é,

$$\begin{aligned} x_1 + x_2 + z_{11} &= 2 \\ 2x_1 + y_1 - v_1 + z_{21} &= 4 \\ 2x_2 + y_1 - v_2 + z_{22} &= 2 \\ v_1 x_1 + v_2 x_2 &= 0. \end{aligned}$$

A função-objectivo que deve ser minimizada é

$$z_{11} + z_{21} + z_{22}$$

com as restrições apresentadas e $x_1 \geqq 0$, $x_2 \geqq 0$, $v_1 \geqq 0$, $v_2 \geqq 0$, $z_{11} \geqq 0$, $z_{21} \geqq 0$, $z_{22} \geqq 0$ e y_1 sem restrição de sinal. Pondo $y_1 = u_1 - u_2$ $(u_1 \geqq 0,\ u_2 \geqq 0)$, o programa linear a resolver consiste em

$$\text{minimizar } z_{11} + z_{21} + z_{22}$$

$$(\text{maximizar } -z_{11} - z_{21} - z_{22})$$

com as restrições

$$\begin{aligned} x_1 + x_2 \qquad\qquad + z_{11} \qquad\quad &= 2 \\ 2x_1 \quad + u_1 - u_2 - v_1 \qquad + z_{21} \quad &= 4 \\ 2x_2 + u_1 - u_2 \quad - v_2 \qquad\quad + z_{22} &= 2 \\ v_1 x_1 + v_2 x_2 = 0 & \end{aligned}$$

$x_1 \geqq 0$, $x_2 \geqq 0$, $u_1 \geqq 0$, $u_2 \geqq 0$, $v_1 \geqq 0$, $v_2 \geqq 0$, $z_{11} \geqq 0$, $z_{21} \geqq 0$, $z_{22} \geqq 0$.

Tomando a matriz

$$D = \begin{array}{c} \begin{matrix} D_1 & D_2 & D_3 & D_4 & D_5 & D_6 & D_7 & D_8 & D_9 \end{matrix} \\ \left[\begin{matrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 2 & 0 & 1 & -1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 1 & -1 & 0 & -1 & 0 & 0 & 1 \end{matrix}\right] \end{array}$$

obtém-se fàcilmente a solução básica inicial $z_{11} = 2$, $z_{21} = 4$, $z_{22} = 2$ e o primeiro quadro do simplex é

		D	1	2	3	4	5	6	7	8	9
Base	c_j		0	0	0	0	0	0	−1	−1	−1
D_7	−1	2	1	1	0	0	0	0	1	0	0
D_8	−1	4	2	0	1	−1	−1	0	0	1	0
←D_9	−1	2	0	[2]	1	−1	0	−1	0	0	1
z_j		−8	−3	−3	−2	2	1	1	−1	−1	−1
$c_j - z_j$			3	3 ↑	2	−2	−1	−1	0	0	0

Os restantes quadros derivam-se da maneira habitual:

		D	1	2	3	4	5	6	7	8	9
Base	c_j		0	0	0	0	0	0	−1	−1	−1
←D_7	−1	1	[1]	0	−1/2	1/2	0	1/2	1	0	−1/2
D_8	−1	4	2	0	1	−1	−1	0	0	1	0
D_2	0	1	0	1	1/2	−1/2	0	−1/2	0	0	1/2
z_j		−5	−3	0	−1/2	1/2	1	−1/2	−1	−1	1/2
$c_j - z_j$			3 ↑	0	1/2	−1/2	−1	1/2	0	0	−3/2

		D	1	2	3	4	5	6	7	8	9
Base	c_j		0	0	0	0	0	0	−1	−1	−1
D_1	0	1	1	0	−1/2	1/2	0	1/2	1	0	−1/2
←D_8	−1	2	0	0	[2]	−2	−1	−1	−2	1	1
D_2	0	1	0	1	1/2	−1/2	0	−1/2	0	0	1/2
z_j		−2	0	0	−2	2	1	1	2	−1	−1
$c_j - z_j$			0	0	2 ↑	−2	−1	−1	−3	0	0

		D	1	2	3	4	5	6	7	8	9
Base	c_j		0	0	0	0	0	0	−1	−1	−1
D_1	0	3/2	1	0	0	0	−1/4	1/4	1/2	1/4	1/4
D_8	0	1	0	0	1	−1	−1/2	−1/2	−1	1/2	1/2
D_2	0	1/2	0	1	0	0	1/4	−1/4	1/2	−1/4	1/4
z_j		0	0	0	0	0	0	0	0	0	0
$c_j - z_j$			0	0	0	0	0	0	−1	−1	−1

Atingiu-se o mínimo de $z_{11} + z_{21} + z_{22}$ e portanto o máximo da função-objectivo obtém-se para $x_1 = 3/2$, $x_2 = 1/2$.

5.2 Método de Frank e Wolfe

Sabe-se que para o problema de programação quadrática as condições de KUHN-

TUCKER são necessárias e suficientes para a existência de óptimo.

De acordo com a demonstração do teorema I do n.º 3, $\bar{x}$ é solução do problema de programação quadrática sse

$$\nabla f(\bar{x})\,\bar{x} = \max\{\nabla f(\bar{x})\,w : w \geqq 0 \wedge B\,w \leqq b\}$$

onde

$$w = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}.$$

O teorema da dualidade da programação linear justifica que $\max\{\nabla f(\bar{x})\,w : w \geqq 0 \wedge B\,w \leqq b\} = \min\{u^T b : u \geqq 0 \wedge u^T B \geqq \nabla f(\bar{x})$ onde $u = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}$. Então a condição necessária e suficiente para que $\bar{x}$ seja solução óptima é que

$$\max\{\nabla f(\bar{x})\,\bar{x} - u^T b : u \geqq 0 \wedge u^T B \geqq \nabla f(\bar{x})\} = 0$$

porque

$$\nabla f(\bar{x})\,\bar{x} - \min u^T b = 0$$

e

$$\nabla f(\bar{x})\,\bar{x} - \min u^T b = \max\{\nabla f(\bar{x})\,\bar{x} - u^T b\}.$$

Seja

$$\begin{aligned} F(x,u) &= \nabla f(x)\,x - u^T b \\ &= (c^T - 2\,x^T A)\,x - u^T b \\ &= c^T x - u^T b - 2\,x^T A\,x. \end{aligned}$$

Em virtude do exposto, é fácil concluir que $\bar{x}$ é solução óptima do problema de programação quadrática sse existe $\bar{u}$ tal que

$$\bar{x} \geqq 0,\ \bar{u} \geqq 0,\ B\,\bar{x} \leqq b,\ \nabla f(\bar{x}) \leqq u^T B,$$
$$F(\bar{x},\bar{u}) = 0.$$

O problema equivale pois a maximizar $F(x,u)$ (cujo máximo é zero) sujeita às restrições apresentadas. O algoritmo de FRANK e WOLFE utiliza o método do simplex da programação linear. Introduzindo variáveis auxiliares $v = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}$ e $y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$, pode-se reformular o problema do modo seguinte:

Dadas as restrições

$$x \geqq 0,\ u \geqq 0,\ v \geqq 0,\ y \geqq 0$$
$$B\,x + y = b$$
$$2\,A\,x + B^T u - v = c$$
$$v^T x + u^T y = 0,$$

maximizar

$$\begin{aligned} F(x,u) &= c^T x - u^T b - 2\,x^T A\,x = \\ &= -\,v^T x - u^T y. \end{aligned}$$

O valor máximo é zero, se existir solução óptima para o programa quadrático.

As fases do algoritmo de FRANK e WOLFE são as seguintes:

1. Constrói-se a equação

$$D\,w = \begin{bmatrix} b \\ c \end{bmatrix} = d,$$

com

$$D = \begin{bmatrix} B & 0 & \mathrm{I} & 0 \\ 2\,A & B^T & 0 & -\mathrm{I} \end{bmatrix}$$

e

$$w = \begin{bmatrix} x \\ u \\ y \\ v \end{bmatrix}.$$

2. Escolhe-se uma solução admissível básica inicial como no método do simplex.
3. Exprime-se d e todas as colunas de D em termos da base escolhida.

4. Seja w_1^* o vector constituído pelos coeficientes da combinação linear de d em termos dos vectores da base. Utilizando os coeficientes de w_1^* nas posições indicadas pelos índices dos vectores da base e zeros nos outros lugares, obtenha-se $w_1 = \begin{bmatrix} x \\ u \\ y \\ v \end{bmatrix}$.

5. Calcula-se $v^T x + u^T y$.

   a) Se $v^T x + u^T y = 0$, o algoritmo termina.

   b) Se $v^T x + u^T y \neq 0$, o algoritmo procede de acordo com as fases seguintes.

6. Seja $\nabla F(w)$ o gradiente de F tomando as derivadas parciais de F em relação às componentes de w, na ordem em que aparecem em w.

7. Calcula-se $\nabla F(w_1)$ e usa-se $\nabla F(w_1)$ como vector dos coeficientes de custo no método do simplex para obter a nova base.

8. Obtém-se w_2 a partir do vector d expresso como combinação linear da nova base.

9. Calcula-se $F(w_2) = v^T x + u^T y$.

   a) Se $v^T x + u^T y = 0$, o algoritmo termina.

   b) Se $v^T x + u^T y \neq 0$, e

   c) Se $v^T x + u^T y$ não decresceu em relação à fase 5, determine-se λ tal que $F(w) = F[w_1 + \lambda(w_2 - w_1)]$ $(0 \leqq \lambda \leqq 1)$ é minimizada. Obtém-se assim o novo w para o qual ∇F é calculado, achando-se os coeficientes de custo. Quando c) está completada volta-se à fase 5.

EXEMPLO 5. 2. Maximizar $f(x) = 4x_1 + 6x_2 - x_1^2 - 3x_2^2$ com as restrições

$$x_1 + 2x_2 \leqq 4$$
$$x_1 \geqq 0, \quad x_2 \geqq 0.$$

Para este problema é

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \quad B = [1 \;\; 2] \quad b = [4] \quad c = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$$

$$D\,w = \begin{array}{c} \begin{matrix} D_1 & D_2 & D_3 & D_4 & D_5 & D_6 \end{matrix} \\ \begin{bmatrix} 1 & 2 & 0 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 & -1 & 0 \\ 0 & 6 & 2 & 0 & 0 & -1 \end{bmatrix} \end{array} \begin{array}{c} w \\ \begin{bmatrix} x_1 \\ x_2 \\ u \\ y \\ v_1 \\ v_2 \end{bmatrix} \end{array} =$$

$$= \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix}.$$

Tomemos, por exemplo, a base inicial D_3, D_4 e D_6 e exprimam-se os restantes vectores de D como combinações lineares destes. Como

$$[D_3 \;\; D_4 \;\; D_6]^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 2 & -1 \end{bmatrix},$$

vêm os coeficientes seguintes

$$d = [D_3 \;\; D_4 \;\; D_6]^{-1} \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 2 \end{bmatrix}$$

$$D_1 : [D_3 \;\; D_4 \;\; D_6]^{-1} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix}$$

$$D_2 : [D_3 \;\; D_4 \;\; D_6]^{-1} \begin{bmatrix} 2 \\ 0 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ -6 \end{bmatrix}$$

$$D_3 : \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad D_4 : \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$D_5 : [D_3\, D_4\, D_5]^{-1} \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ -2 \end{bmatrix}$$

$$D_6 : \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Pode pois construir-se o quadro

	w_1	1	2	3	4	5	6
D_3	4	2	0	1	0	−1	0
D_4	4	1	2	0	1	0	0
D_6	2	4	−6	0	0	−2	1

$$w_1 = \begin{bmatrix} 0 \\ 0 \\ 4 \\ 4 \\ 0 \\ 2 \end{bmatrix} \begin{matrix} \} x \\ u \\ y \\ \} v \end{matrix}$$

Notemos que se obtém w_1 exprimindo d como composição linear dos vectores da base. Tem-se $F(w_1) = (-v^T x - u^T y)_{w_1} = -[0\ 2]\begin{bmatrix} 0 \\ 0 \end{bmatrix} - [4]\,[4] = -16$ e portanto temos de calcular $\nabla F(w_1)$. Ora neste caso $F(w) = -v_1 x_1 - v_2 x_2 - u_1 y_1$ e portanto

$$\frac{\partial F}{\partial x_1} = -v_1 = 0 \qquad \frac{\partial F}{\partial y_1} = -u_1 = -4$$

$$\frac{\partial F}{\partial x_2} = -v_2 = -2 \qquad \frac{\partial F}{\partial v_1} = -x_1 = 0$$

$$\frac{\partial F}{\partial u_1} = -y_1 = -4 \qquad \frac{\partial F}{\partial v_2} = -x_2 = 0.$$

Adoptando um quadro análogo ao que se utiliza no método do simplex para a programação linear, vem

	$\nabla F(w_1)$	w_1	0	−2	−4	−4	0	0
D_3	−4	4	2	0	1	0	−1	0
D_4	−4	4	1	2	0	1	0	0
$\leftarrow D_6$	0	2	[4]	−6	0	0	−2	1
z_j			−12	−8	−4	−4	4	0
$F_j - z_j$			12	6	0	0	−4	0
			↑					

Como a maior diferença positiva é 12, isto indica que se deve introduzir na base o vector D_1. Para determinar o vector que deve ser retirado calcula-se

$$\min\left(\frac{\text{componente da solução básica } w_1}{\text{correspondente componente positiva de } D_1}\right) =$$

$$= \min\left(\frac{4}{2}, \frac{4}{1}, \frac{2}{4}\right) = 1/2$$

e portanto o vector a retirar é D_6. O novo quadro é

	w_2	1	2	3	4	5	6
D_3	3	0	3	1	0	0	−1/2
D_4	7/2	0	7/2	0	1	1/2	−1/4
D_1	1/2	1	−3/2	0	0	−1/2	1/4

$$w_2 = \begin{bmatrix} 1/2 \\ 0 \\ 3 \\ 7/2 \\ 0 \\ 0 \end{bmatrix} \begin{matrix} \} x \\ u \\ y \\ \} v \end{matrix}$$

$$F(w_2) = -[0\ 0]\begin{bmatrix} 1/2 \\ 0 \end{bmatrix} - [3]\,[7/2] = -21/2$$

$$\frac{\partial F}{\partial x_1} = 0 \qquad \frac{\partial F}{\partial y_1} = -3$$

$$\frac{\partial F}{\partial x_2} = 0 \qquad \frac{\partial F}{\partial v_1} = -1/2$$

$$\frac{\partial F}{\partial u_1} = -7/2 \qquad \frac{\partial F}{\partial v_2} = 0.$$

Tem-se agora o quadro

	$\nabla F(w_2)$	w_2	0	0	−7/2	−3	−1/2	0
← D_3	−7/2	3	0	[3]	1	0	0	−1/2
D_4	−3	7/2	0	7/2	0	1	1/2	−1/4
D_1	0	1/2	1	−3/2	0	0	−1/2	1/4
z_j			0	−21	−7/2	−3	−3/2	5/2
$F_j - z_j$			0	21	0	0	2	−5/2
				↑				

O vector a introduzir na base é D_2 e o vector a retirar é D_3 ou D_4. Admita-se que é D_3. O novo quadro obtém-se fàcilmente:

	w_3	1	2	3	4	5	6
D_2	1	0	1	1/3	0	0	−1/6
D_4	0	0	0	−7/6	1	1/2	1/3
D_1	2	1	0	1/2	0	−1/2	0

$$w_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{matrix} \}x \\ \\ u \\ y \\ \}v \\ \end{matrix}$$

$$F(w_3) = -[0\,0]\begin{bmatrix} 2 \\ 1 \end{bmatrix} - [0][0] = 0$$

Logo, atingiu-se a solução optimizante dada por $x_1 = 2$ e $x_2 = 1$.

5. 3. Método de Beale

O método de BEALE para resolver um problema de programação quadrática não utiliza explìcitamente as condições de KUHN-TUCKER.

Consideremos o problema de minimizar a função convexa

$$f(x) = a + c^T x + x^T A x \quad (a \text{ constante})$$

com as restrições lineares

$$Bx = b$$

$$x \geqq 0\,.$$

É fácil ver que $f(x)$ se pode escrever na forma

$$f(x) = a + \sum_{j=1}^{n} c_j x_j + \frac{1}{2} \sum_{j=1}^{n} x_j \frac{\partial f}{\partial x_j}\,.$$

O método de BEALE, que se descreve seguidamente de modo intuitivo, apresenta muitas semelhanças com o método do simplex para a programação linear. O processo para achar uma solução admissível básica é o mesmo porque as restrições são lineares. Suponhamos determinada uma solução admissível básica não-degenerada (m variáveis básicas positivas e $n - m$ variáveis não-básicas nulas). Como mais adiante são introduzidas novas variáveis não-básicas u_k, que não são x, designemos por z_k uma variável não-básica.

O processo começa por incrementar uma das variáveis não-básicas até que um dos três casos seguintes se dê:

1. A função f começa a crescer. Isto acontece se $\partial f/\partial z_k$ se torna nula e depois positiva, onde z_k é a variável não-básica para a qual $|\partial f/\partial z_k|$, entre todas as $\partial f/\partial z_k \leqq 0$, tem o maior valor no ponto inicial. É claro que, na direcção de z_k, f decresce mais ràpidamente e portanto a variação ao longo de z_k conduz mais ràpidamente ao min f.

2. Uma das restrições é violada quando z_k cresce.

3. Quando z_k cresce, uma das variáveis básicas, que originalmente tinha um

valor positivo, torna-se negativa, violando a condição $x_j \geqq 0$. Esta condição é essencialmente a mesma que (2).

Consoante o caso, introduz-se um novo conjunto de variáveis não-básicas, segundo um processo que se descreve adiante. Prosseguindo deste modo, atinge-se um ponto onde nenhuma variação posterior do valor de uma variável não-básica produz decrescimento do valor de f. *Como f é convexa, o mínimo absoluto é atingido.*

Seja x_i uma variável básica e z_j uma variável não-básica. Então, usando as restrições lineares, pode escrever-se

$$1) \quad x_i = \alpha_{i0} + \sum_{j=1}^{n-m} \alpha_{ij} z_j \quad (i = 1, 2, \cdots, m)$$

onde $z_j = x_{m+j}$. A função f também se pode exprimir em termos das variáveis x_j de acordo com as expressões 1):

$$2) \quad f = \sum_{i=0}^{n-m} \sum_{j=0}^{n-m} \beta_{ij} z_i z_j$$

onde $z_0 = 1$ e $z_1, \cdots, z_m$ são variáveis não-básicas. Se $\partial f / \partial z_k < 0$ para algum $k (k = 1, \cdots, n-m)$, então um pequeno aumento de z_k com $z_i = 0$ $(i \neq k)$ reduzirá f. É claro que é conveniente aumentar z_k até que (1) algum x_k se anule e decresça para valores negativos ou (2) $\partial f / \partial z_k$ se anule e passe a tomar valores positivos. No caso (1), o conjunto de variáveis não-básicas é mudado, substituindo x_k por z_k, e exprime-se f em termos das novas variáveis não-básicas; no caso (2), como $\partial f / \partial z_k$ é uma função linear dos z_j, introduz-se

$$u_k = \frac{1}{2} \frac{\partial f}{\partial z_k}$$

como variável não-básica. Esta variável u_k é semelhante a qualquer outra variável não-básica; a diferença reside em que u_k pode tomar valores positivos e negativos (por esse facto é u_k uma *variável livre*).

Enfim, as fases do algoritmo são as seguintes:

Fase 1 — Acha-se uma solução admissível básica.

Fase 2 — Exprime-se f em termos das variáveis não-básicas.

Fase 3 — Calcula-se

$$\partial f / \partial z_k \quad (k = 1, \cdots, n-m)$$

e escolhe-se z_k por forma que $|\partial f / \partial z_k|$ é o maior entre todos os k para os quais $\partial f / \partial z_k < 0$.

Fase 4 — Deram-se três condições para o acréscimo de z_k que vão ser agora aplicadas. Resolve-se em ordem a z_k a equação $\partial f / \partial z_k = 0$, que é linear em z_k e portanto conduz a um só valor de z_k, mantendo nulas as outras variáveis não-básicas. Resolve-se também em ordem a z_k cada uma das equações

$$3) \quad x_i = \alpha_{i0} + \sum_{j=1}^{n-m} \alpha_{ij} z_j,$$

onde

$$z_j = 0 \quad (j \neq k)$$

e

$$x_i = 0.$$

Comparam-se os diferentes valores de z_k obtidos destes cálculos. Seja z_{k_0} o menor de todos. Segue-se a Fase 5a se z_{k_0} não se obtém das equações 3); caso contrário segue-se a Fase 5b.

Fase 5a — Põe-se $u_k = \frac{1}{2} \frac{\partial f}{\partial z_k}$ e resolve-se em ordem a z_k (esta expressão

é linear em z_k). Substitui-se depois z_k por essa expressão. Note-se que nesta fase $u_k = 0$ porque $\partial f/\partial z_k = 0$.

Fase 5b — Tendo-se determinado z_{k_0} como o menor dos valores correspondentes a uma certa equação em 3), substitui-se z_k em f pelo x_i nessa equação e obtém-se uma nova expressão para f nas variáveis não-básicas. Note-se que, nesta fase, z_k é básica e igual a $-\alpha_{i0}/\alpha_{ik}$.

Fase 6 — Calcula-se $\partial f/\partial z_k$, onde z_k é qualquer variável não-básica. Calcula-se $|\partial f/\partial z_k|$ para todos os z_k tais que $\partial f/\partial z_k < 0$ ou $z_k = u_k$.

Distinguem-se dois casos:

Caso 1. Se $|\partial f/\partial z_k|$ é máximo para $z_k = u_k$, então u_k deve decrescer ou crescer consoante $\partial f/\partial u_k$ é positiva ou negativa e segue-se a fase 5 usando a nova solução admissível com o mínimo valor de z_k acabado de calcular.

Caso 2. Se z_h não é u_k, então retoma-se a fase 4. O algoritmo termina quando a variação de qualquer variável não-básica já não produz uma diminuição do valor de f.

EXEMPLO 5. 3.

minimizar $f = x_1^2 + x_2^2 - 4x_1 - 2x_2 + 5$

com as restrições

$$x_1 + x_2 \leqq 2$$

$$x_1 \geqq 0, \quad x_2 \geqq 0.$$

Tomando a variável auxiliar $x_3 \geqq 0$, a desigualdade transforma-se na equação $x_1 + x_2 + x_3 = 2$. Uma solução básica imediata é $x_1 = x_2 = 0$, $x_3 = 2$. Tem-se

$$x_3 = 2 - x_1 - x_2$$

$$\frac{\partial f}{\partial x_1} = 2x_1 - 4 \qquad \left(\frac{\partial f}{\partial x_1}\right)_{\substack{x_1=0\\x_2=0\\x_3=2}} = -4$$

$$\frac{\partial f}{\partial x_2} = 2x_2 - 2 \qquad \left(\frac{\partial f}{\partial x_2}\right)_{\substack{x_1=0\\x_2=0\\x_3=2}} = -2$$

e portanto deve ser aumentada a variável não básica x_1.

Tomando $x_3 = 2 - x_1 - x_2$, vem $0 = 2 - -x_1$ ou $x_1 = 2$; de $\dfrac{\partial f}{\partial x_1} = 0$ vem também $x_1 = 2$. Como os valores são iguais, tome-se, por exemplo, a equação $x_3 = 2 - x_1 - x_2$ que, resolvida em ordem a x_1 (nova variável básica), dá

$$x_1 = 2 - x_2 - x_3$$

e proceda-se à substituição em f. Vem

$$f = 2x_2^2 + x_3^2 + 2x_2x_3 - 2x_2 + 1.$$

Tem-se agora

$$\frac{\partial f}{\partial x_2} = 4x_2 + 2x_3 - 2 \qquad \left(\frac{\partial f}{\partial x_2}\right)_{\substack{x_2=0\\x_3=0}} = -2$$

$$\frac{\partial f}{\partial x_3} = 2x_3 + 2x_2 \qquad \left(\frac{\partial f}{\partial x_3}\right)_{\substack{x_2=0\\x_3=0}} = 0$$

e portanto deve-se aumentar x_2. Como de

$$x_1 = 2 - x_2 - x_3$$

vem

$$0 = 2 - x_2 - 0,$$

ou

$$x_2 = 2,$$

e de

$$\frac{\partial f}{\partial x_2} = 0$$

vem

$$4x_2 + 0 - 2 = 0,$$

ou

$$x_2 = \frac{1}{2},$$

a variável x_2 só poderá aumentar até $1/2$. Fazendo

$$u_2 = \frac{1}{2}\frac{\partial f}{\partial x_2} = 2x_2 + x_3 - 1,$$

vem

$$x_2 = \frac{u_2}{2} - \frac{x_3}{2} + \frac{1}{2}$$

e

$$f = \frac{u_2^2}{2} + \frac{x_3^2}{2} + x_3 + \frac{1}{2}.$$

Agora

$$\frac{\partial f}{\partial x_3} = x_3 + 1 \qquad \left(\frac{\partial f}{\partial x_3}\right)_{\substack{x_3=0\\u_2=0}} = 1 > 0$$

$$\frac{\partial f}{\partial u_2} = u_2 \qquad \left(\frac{\partial f}{\partial u_2}\right)_{\substack{x_3=0\\u_2=0}} = 0$$

e portanto nenhuma variação em x_3 ou u_2 produz decrescimento em f e alcançámos o mínimo.

Portanto a solução minimizante é $x_3 = 0$, $x_2 = 1/2$, $x_1 = 3/2$ e o mínimo é $f = 1/2$.

6. Programação côncava geral

Como já se disse no n.º 3, o problema da programação côncava (1) consiste em maximizar a função côncava $f(x)$, supondo que todas as funções $g_i(x)$ são convexas. Poucos progressos se têm feito no que se refere à pesquisa de algoritmos eficientes para resolver o problema geral da programação matemática mas, no caso especial da programação côncava, existem já alguns processos, na maior parte iterativos, para a pesquisa da solução óptima.

Quando as funções $g_i(x)$ são convexas, o teorema XII do n.º 2 ensina que o conjunto admissível é convexo. Esta propriedade e a concavidade de $f(x)$ implicam que qualquer óptimo local é também óptimo absoluto (teorema III do n.º 2). Logo, em vez de procurar e comparar um grande número (possìvelmente infinito) de óptimos locais, é apenas necessário encontrar um óptimo local pois este será necessàriamente óptimo absoluto.

Grande número de algoritmos apoia-se no gradiente $\nabla f(x)$ da função-objectivo e são por isso mesmo chamados *métodos do gradiente*. Eles aproveitam o facto de a direcção de $\nabla f(x)$ ser a direcção de máximo crescimento de $f(x)$ para, a partir de um ponto arbitrário pertencente ao conjunto admissível, gerarem uma sucessão de pontos convergente para a solução óptima.

Comparando os diferentes métodos do gradiente, verifica-se que eles possuem certas semelhanças. Se, por exemplo, o problema consiste em maximizar a função côncava $f(x)$ sujeita às restrições lineares

$$A_i x \leqq b_i,$$

consideremos um ponto arbitrário x^0 (interior ou fronteiro) pertencente ao conjunto admissível ($A_i x^0 \leqq b_i$). Para passar do ponto x^k ao ponto x^{k+1} determina-se uma direcção s^k em x^k tal que, para um pequeno $\lambda > 0$,

(1) Alguns autores chamam-lhe programação convexa. De facto ela também se pode definir como a minimização de uma função convexa, supondo que as funções $g_i(x)$ são convexas.

$x^k + \lambda s^k$ pertence ainda ao conjunto admissível. Para isso é necessário e suficiente que

$$A_i s^k \leqq 0 \text{ para todo o } i \text{ com } A_i x^k = b_i .$$

A direcção s_k diz-se *admissível*. A função $f(x)$ tem de aumentar ao longo de $x^k + \lambda s^k$ e portanto deverá obedecer à condição

$$\nabla f(x^k) s^k > 0 .$$

Toda a direcção s^k que satisfaz a esta desigualdade diz-se *utilizável*. Os diferentes métodos dão regras para determinar s^k de entre o conjunto de todas as direcções s admissíveis e utilizáveis em x^k. Achada a direcção s^k mais favorável, o cálculo de λ_k faz-se sempre da mesma maneira para todos os métodos. Determina-se o valor λ' para o qual $x^k + \lambda' s^k$ não pertence ao conjunto admissível, assim como o valor λ'' para o qual $f(x)$ é óptimo sobre $x^k + \lambda s_k$. Então λ_k é dado por

$$\lambda_k = \min(\lambda', \lambda'')$$

e vem

$$x^{k+1} = x^k + \lambda_k s^k .$$

Entre os métodos do gradiente figuram os de FRISCH, ROSEN e ZOUTENDIJK.

Método recente muito promissor deve-se a C. W. CARROL. Aperfeiçoado por FIACCO e MCCORMICK, é vulgarmente conhecido por *método sequencial* ou *método SUMT* («sequential unconstrained minimization technique»).

Consideremos o problema de minimizar a função convexa $f(x)$ sujeita às restrições $g_i(x) \geqq 0$ $(i = 1, 2, \cdots, m+n)$ que incluem as condições de não-negatividade $x \geqq 0$. Admite-se que as funções $g_i(x)$ são côncavas.

O processo iterativo socorre-se da função

$$P(x; r) = f(x) + r \sum_{i=1}^{m+n} \frac{1}{g_i(x)}$$

onde $r > 0$ é um parâmetro perturbador.

A função $P(x; r)$ deve ser minimizada para diferentes valores de r sujeita às restrições $g_i(x) \geqq 0$. Notando que $1/g_i(x)$ tende para $+\infty$ quando $g_i(x) \to +0$, é evidente que o ponto x^* que miniminiza $P(x; r)$ satisfaz a $g_i(x^*) > 0$ $(i = 1, 2, \cdots, m+n)$. Logo o termo

$$r \sum_{i=1}^{m+n} \frac{1}{g_i(x)}$$

impede que x^* seja ponto fronteiro do conjunto admissível. Este facto é importantíssimo pois reduz o problema da minimização de $P(x; r)$, sujeita às restrições $g_i(x) \geqq 0$, à pesquisa de um minimizante interior, o que é muito mais fácil. É claro que ainda pode ser difícil resolver este último problema e usualmente é necessário utilizar um método de aproximações sucessivas mas pelo menos o problema pode ser resolvido pelos métodos clássicos.

O objectivo último do método sequencial é achar uma solução admissível que minimize $f(x)$. Ora não é difícil mostrar que

$$I. \lim_{r_k = 0} \min_x P(x; r_k) = \min_x f(x) \equiv v_0 .$$

Note-se que, sendo $g(x)$ côncava, então $1/g(x)$ é convexa para todo o x tal que $g(x) > 0$. Logo, para r fixo, $P(x; r)$ possui um único mínimo em relação a x. Também é claro que, para $r_a < r_b$,

$$\min_x P(x; r_a) < \min_x P(x; r_b) .$$

Para provar o teorema basta mostrar que

$$\forall \varepsilon > 0, \exists k_0 : k > k_0 \Rightarrow |\min_x P(x; r_k) - v_0| < \varepsilon .$$

Ora, sendo $v_0 = \min_x f(x)$, tem-se

$$f(x) < v_0 + \varepsilon/2$$

para x suficientemente vizinho do minimizante. Se x^* satisfaz a esta condição, escolha-se r_{k_0} tal que

$$\frac{r_{k_0}}{\min_i g_i(x^*)} < \varepsilon/2\,(m+n)\,.$$

Então, para $k > k_0$, vem

$$\min_x P(x\,;r_k) < \min_x P(x\,;r_{k_0}) < v_0 + \\ + \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = v_0 + \varepsilon$$

e também é verdade que

$$\min_x P(x\,;r_k) > v_0 - \varepsilon$$

porque $r_k \sum_{i=1}^{m+n} \frac{1}{g_i(x)} > 0$. Está pois completada a demonstração.

Na prática, começa-se por determinar o ponto x que minimiza $P(x\,;r_0)$ para um pequeno valor de r_0 prèviamente escolhido. Para isso basta resolver o sistema de equações

$$\frac{\partial P}{\partial x_j} = 0 \qquad (j = 1\,,\cdots,n)\,.$$

Toma-se depois $r_1 < r_0$ e determina-se novamente o mínimo de $P(x\,;r_1)$. Repete-se este processo para valores sucessivamente mais pequenos de r até se obter uma sucessão de pontos que justifique a extrapolação para a solução óptima.

EXEMPLO 6. 1. Resolver o problema 3. 1 utilizando o método sequencial.

Tem-se

$$P(x_1\,,x_2\,;r) = \frac{(x_1+1)^3}{3} + x_2 + \\ + \frac{r}{x_1 - 1} + \frac{r}{x_2}$$

$$\frac{\partial P}{\partial x_1} = (x_1+1)^2 - \frac{r}{(x_1-1)^2}$$

$$\frac{\partial P}{\partial x_2} = 1 - \frac{r}{x_2^2}\,.$$

A solução minimizante de $P(x_1\,,x_2\,;r)$ obtém-se fàcilmente neste caso. Trata-se de $x_1 = \sqrt{1+\sqrt{r}}$ $x_2 = \sqrt{r}$. Quando $r \to 0$, vem a solução optimizante $x_1 = 1$, $x_2 = 0$ já encontrada anteriormente por meio das condições de KUHN-TUCKER (Exemplo 3. 1).

BIBLIOGRÁFIA

ABADIE, J. (editor): *Nonlinear programming*, North-Holland Publishing Company, Amsterdam, 1967.

BEALE, E. M. L.: «Nonlinear programming» in *Digital computer user's handbook* (edit. M. Klerer e G. A. Korn). McGraw-Hill Book Company, New York, 1967.

BERGE, C. (e A. Ghouila-Houri): *Programmes, jeux et réseaux de transport*, Dunod, Paris, 1962.

BRACKEN, J. (e G. P. McCormick): *Selected applications of nonlinear programming*, John Wiley & Sons, Inc., New York, 1968.

DANTZIG, G.: *Linear programming and extensions* (tradução francesa do original americano). Dunod, Paris, 1966.

DORFMAN, R. (P. A. Samuelson e R. M. Solow): *Linear programming and economic analysis*. McGraw-Hill, New York, 1958.

FIACCO, A. V. (e G. P. McCormick): *Nonlinear programming*, John Wiley and Sons, Inc., New York, 1968.

GUE, R. L. (e M. E. Thomas): *Mathematical methods in operations research*. The MacMillan Company, New York, 1968.

HADLEY, G.: *Nonlinear and dynamic programming*, Addinson-Wesley Publishing Company, Reading, 1964.

HILLIER, F. S. (e G. J. Lieberman): *Introduction to operations research*, Holden-Day, Inc., San Francisco, 1967.

KARLIN, S.: *Mathematical methods and theory in games, programming, and economics* (Vol. 1), Pergamon Press, London, 1959.

KÜNZI, H. P. (e W. Krelle): *Nonlinear Programming* (tradução inglesa do original alemão). Blaisdell Publishing Company, Waltham, 1966.

KÜNZI, H. P. (H. G. Tzschach e C. A. Zehnder): *Numerical methods of mathematical optimization* (tradução americana do original alemão), Academic Press, New York, 1968.

LANCASTER, K.: *Mathematical economics*, The MacMillan Company, New York, 1968.

MANGASARIAN, O. L.: *Nonlinear programming*, McGraw-Hill Book Company, New York, 1969.

SAATY, T. L.: *Mathematical methods of operations research*, McGraw-Hill, New York, 1959.

———: *Nonlinear mathematics*, McGraw-Hill, New York, 1964.

TEICHROEW, D.: *An introduction to management science--deterministic models*, John Wiley & Sons, New York, 1964.

VADJA, S.: *Mathematical programming*, Addinson-Wesley Publishing Company, London, 1961.

———: *Readings in mathematical programming*, Pitman, London, 1962.

ZANGWILL, W. I.: *Nonlinear programming: a unified approach*. Prentice Hall, Englewood Cliffs, N. J., 1969.