
GAZETA DE MATEMÁTICA

JORNAL DOS CONCORRENTES AO EXAME DE APTIDÃO E DOS
ESTUDANTES DE MATEMÁTICA DAS ESCOLAS SUPERIORES

ANO XXXI

N.º 117-120

JANEIRO-DEZ. 1970

SUMÁRIO

Aspectos fundamentais da teoria geral
da Programação Matemática
por *Fernando de Jesus*

Sobre o Teorema dos Códigos
para um Canal sem Ruído
por *J. Marques Henriques*

Matrices partitionnées en blocs commutatifs
par *José Vitória*

Sobre um teorema de Lucas
relacionado com o pequeno teorema de Fermat
por *José Morgado*

Geometrização da Lógica de Proposições
(a n dimensões)
por *António José Mendes Silva*

Algumas propriedades da distribuição
binomial negativa
por *António Dorival Campos*
e *Euclides Custódio de Lima Filho*

Matemáticas superiores

Boletim bibliográfico

Panorama da Geodesia Contemporânea
por *F. Teixeira de Queiroz*

G A Z E T A D E M A T E M Á T I C A

EDITOR — *Gazeta de Matemática, Lda.*

ADMINISTRADOR — *A. Sá da Costa*

Sede e Administração — Rua Diário de Notícias, 134-1.º-Esq.º — Tel. 369449 — Lisboa-2

REDACÇÃO

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

OUTROS COMPONENTES

EM PORTUGAL:

Coimbra: L. Albuquerque; **Lisboa:** Almeida e Costa, A. Sá da Costa, J. J. Dionísio, J. Sebastião e Silva, J. Ribeiro Albuquerque, M. Teodora Alves, Fernando de Jesus, A. César de Freitas e Fernando Dias Agudo; **Porto:** Andrade Guimarães, Arala Chaves, Coimbra de Matos, Laureano Barros, L. Neves Real.

NO ESTRANGEIRO:

Argentina — *Buenos Aires:* António Monteiro, L. A. Santaló e Eduarde del Busto; *Mendoza:* F. Toranzos; *San Luis:* Manuel Balanzat; **Brasil** — *Belo Horizonte:* Cristovam dos Santos; *Recife:* Newton Maia, Ray Luís Gomes e José Morgado; *Rio de Janeiro:* Achille Bassi, Leopoldo Nachbin, Maria Laura Mousinho e Maurício Peixoto; *São Paulo:* Omar Catunda; **Espanha** — *Barcelona:* Francisco Sanvisens; *Madrid:* Sixto Ríos Garcia; **Itália** — *Roma:* Emma Castelnuovo; **França** — *Paris:* Paul Belgodère; A. Pereiro Gomes; **Suissa** — *Zürich:* H. Wermus; **Uruguay** — *Montevideo:* Rafael La Guardia; **U. S. A.** — *Pennsylvania:* Maria Pilar Ribeiro; **Venezuela** — J. Gallego Diaz.

Toda a colaboração enviada para publicação nesta revista deve ser dactilografada. A G. M. fornece separatas dos artigos publicados, mediante acordo prévio entre o Autor e a Redacção.

Publicações do CENTI (Centro de Tratamento da Informação)

Relatório Revisto sobre a Linguagem Algorítmica — ALGOL 60

Tradução de J. G. TEIXEIRA

Problemas de Matemática na Teoria dos Reactores Nucleares

J. G. TEIXEIRA

Natureza da Investigação Operacional

FERNANDO DE JESUS

Os sócios de S. P. M., assinantes de «Gazeta de Mat.» e de «Portugaliae Meth.», beneficiam para estas obras do desconto de 20%.

Aspectos fundamentais da teoria geral da programação matemática

por *Fernando de Jesus*

1. Introdução

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

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,

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

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

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

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 $x_j \geq 0$ ($j = 1, \dots, n$) 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 títu-

los 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 ($j = 1, \dots, n$) é a quantia a ser investida no j -ésimo título, então $\sum_{j=1}^n x_j = C$. O ren-

dimento total é $\sum_{j=1}^n r_j x_j$, o seu valor médio é

$$E = \sum_{j=1}^n \mu_j x_j \text{ e a sua variância é}$$

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

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

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

(σ constante), $\sum_{j=1}^n x_j = C$ e $x_j \geq 0$; ou mini-

mizar $\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$ sujeita às restrições

$$\sum_{j=1}^n \mu_j x_j = \mu \text{ (} \mu \text{ constante), } \sum_{j=1}^n x_j = C, x_j \geq 0.$$

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) ($j = 1, \dots, n$) 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 $a \geq 0, b \geq 0$.

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

$$C = \sum_{i=1}^n c_i X_i$$

satisfeitas as restrições

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

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

Balança comercial:

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

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

Duas hipóteses que também se admitem são as seguintes:

$$i) \quad a_{ij} \geq 0 \quad (i, j = 1, \dots, n), \quad \sum_{i=1}^n a_{ij} < 1 \\ (j = 1, \dots, n) \quad (\text{Propriedades de LEONTIEF}).$$

$$ii) \quad h_j = \gamma_j + \rho_j E_j \quad \text{com } \gamma_j > 0, \quad \rho_j < 0 \\ (j = 1, \dots, n).$$

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

fácilmente redutíveis a

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

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

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

$$\begin{aligned} c_j - \sum_{i=1}^m \lambda_i^* a_{ij} &= \mu_j^* \leq 0 \\ \mu_j^* x_j^* &= 0 \\ \sum_{j=1}^n a_{ij} x_j^* &= b_i, \end{aligned}$$

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

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

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

tem-se ($x_j \geq 0$)

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

e

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

donde resulta

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

Por outro lado, de acordo com o teorema I,

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

ou

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

donde vem

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

Podemos 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

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

$$\mu_j x_j = 0.$$

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

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

conduzem a 2^o 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 somente 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 $\mu_j \leq 0$, 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 simultaneamente 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.

Seja $X \subseteq R^n$, diz-se que X é conjunto convexo sse, com $0 \leq \theta \leq 1$ e $\bar{\theta} = 1 - \theta$,

$$1) \quad \forall x, y \in R^n \quad \theta x + \bar{\theta} y \in X.$$

Consideram-se convexas 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

$$2) \quad \forall x, y \in X \quad f(\theta x + \bar{\theta} y) \leq \theta f(x) + \bar{\theta} f(y).$$

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 obviamente uma função simultaneamente 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 ($0 \leq \theta \leq 1$). Então se $\hat{x} = \theta x + \bar{\theta} y$, tem-se

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

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

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

ou

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

Como consequência imediata da definição, é possível mostrar que

I. Se as funções f_j ($j = 1, \dots, k$) 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 $\lambda_j \geq 0$.

Pode demonstrar-se também que

II. f é convexa em X sse $[X, f] = \{(x, z) : f(x) \leq z\}$ 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 $f(x) \leq z$ e $f(x') \leq z'$ e

$$f(\theta x + \bar{\theta} x') \leq \theta f(x) + \bar{\theta} f(x') \leq \theta z + \bar{\theta} z'.$$

O ponto

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

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

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

e portanto

$$f(\theta x + \bar{\theta} y) \leq \theta f(x) + \bar{\theta} f(y).$$

Analogamente se mostraria que

II'. f é côncava em X sse $[X, f] = \{(x, z) : f(x) \geq z\}$ 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 \bar{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

$$f(\theta x^* + \bar{\theta} x^0) \leq \theta f(x^*) + \bar{\theta} f(x^0) < f(x^0).$$

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

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

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

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

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

A definição de convexidade e concavidade permite também mostrar facilmente 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 $x = \theta x^1 + \bar{\theta} x^2$. De facto,

$$\begin{aligned} f(x) &= f(\theta x^1 + \bar{\theta} x^2) \leq \theta f(x^1) + \bar{\theta} f(x^2) = \\ &= f(x^2) = f(x^1) \end{aligned}$$

e, como $f(x) \geq f(x^2)$, 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 int X .

Uma função f convexa ou côncava não é necessariamente 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

$$f(x) - f(x^*) \geq (\leq) \nabla f(x^*)(x - x^*),$$

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

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

$$f(\theta x + \bar{\theta} x^*) \leq \theta f(x) + \bar{\theta} f(x^*)$$

ou

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

que se pode escrever ainda na forma

$$f(x) - f(x^*) \geq \frac{f[\theta x + (1 - \theta)x^*] - f(x^*)}{\theta}.$$

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

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

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

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 $f(x) - f(x^*) \geq \nabla f(x^*)(x - x^*)$, a função f é convexa. Com efeito, de

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

escolhendo

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

obtém-se

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

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

e, notando que

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

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

vem

$$f(x^1) \geq f(\theta x^1 + \bar{\theta} x^2) + \bar{\theta} \nabla f(\theta x^1 + \bar{\theta} x^2)(x^1 - x^2)$$

$$f(x^2) \geq f(\theta x^1 + \bar{\theta} x^2) - \theta \nabla f(\theta x^1 + \bar{\theta} x^2)(x^1 - x^2)$$

ou

$$\begin{aligned} \theta f(x^1) &\geq \theta f(\theta x^1 + \bar{\theta} x^2) + \\ &+ \theta \bar{\theta} \nabla f(\theta x^1 + \bar{\theta} x^2)(x^1 - x^2) \\ \bar{\theta} f(x^2) &\geq \bar{\theta} f(\theta x^1 + \bar{\theta} x^2) - \\ &- \theta \bar{\theta} \nabla f(\theta x^1 + \bar{\theta} x^2)(x^1 - x^2). \end{aligned}$$

Somando ordenadamente estas desigualdades, vem finalmente

$$\theta f(x^1) + \bar{\theta} f(x^2) \geq f(\theta x^1 + \bar{\theta} x^2),$$

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 ($0 < \theta < 1$)

$$\begin{aligned} T &= \bar{\theta} f(x) + \theta f(y) - f(\bar{\theta} x + \theta y) \\ &= \bar{\theta} [f(x) - f(\bar{\theta} x + \theta y)] + \theta [f(y) - f(\bar{\theta} x + \theta y)]. \end{aligned}$$

Utilizando a fórmula de TAYLOR, vem

$$\begin{aligned} T &= \bar{\theta} \nabla f(\bar{\theta} x + \theta y) \theta (x - y) + \\ &+ \frac{\theta^2 \bar{\theta}}{2} (x^T - y^T) \nabla^2 f(\xi_1) (x - y) + \\ &+ \theta \bar{\theta} \nabla f(\bar{\theta} x + \theta y) (y - x) + \\ &+ \frac{\theta \bar{\theta}^2}{2} (y^T - x^T) \nabla^2 f(\xi_2) (y - x), \end{aligned}$$

simplificando-se a primeira e terceira parcelas.

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

$$T = \bar{\theta} f(x) + \theta f(y) - f(\bar{\theta} x + \theta y) \geq 0$$

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

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

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

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 $f(x) = x_1^2 + 2x_1x_2 + x_2^2$ é convexa pois

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

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

e, considerando $\left[\begin{array}{cc} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{array} \right]$, vem

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

EXEMPLO 2.4. A função $f(x) = -x_1^4 - 3x_1^2 + 5x_1 - 2x_1x_2 - x_2^2$ é côncava pois

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

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

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

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

resultando

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

e

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

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

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:

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

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

que, simplificadas, dão

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

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

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

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 $g(x) \leq 0$, onde $g(x)$ é convexa, é conjunto convexo.*

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

$$g(x) \leq 0$$

$$g(y) \leq 0$$

e

$$g(\theta x + \bar{\theta} y) \leq \theta g(x) + \bar{\theta} g(y) \leq 0,$$

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

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

XII. *O conjunto definido por $g_i(x) \leq 0$, onde $g_i(x)$ ($i = 1, \dots, m$) 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

$$g_i(x) \leq 0 \quad (i = 1, \dots, m)$$

$$x \geq 0.$$

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

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

$$\begin{aligned} G(x) &\leq 0 \\ x &\geq 0. \end{aligned}$$

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

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

o vector-linha $\lambda = [\lambda_1 \dots \lambda_m]$ 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 $G(\bar{x}) \leq 0$ e $\bar{x} \geq 0$ constitui *solução admissível* (ou *possível*); se, além de admissível, otimizar a função-objectivo, \bar{x} toma o nome de *solução óptima*. O conjunto das soluções admissíveis

constitui a *região admissível* (ou *conjunto admissível*) $C = \{x : G(x) \leq 0 \wedge x \geq 0\}$.

A restrição $g_i(x) \leq 0$ 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 necessariamente $g_i(\bar{x}) = 0$ 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 $g_{i_1}(\bar{x}) = \dots = g_{i_r}(\bar{x}) = \bar{x}_{j_1} = \dots = \bar{x}_{j_s} = 0$. 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

$$\nabla g_i(\bar{x}) dx \leq 0 \quad (i = i_1, \dots, i_r)$$

e

$$dx_j \geq 0 \quad (j = j_1, \dots, j_s),$$

então a direcção $dx = (dx_1, \dots, dx_n)$ 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

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

$$x_1 \geq 0, \quad x_2 \geq 0,$$

vê-se facilmente 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 $dx_1 \geq 0$ 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*.

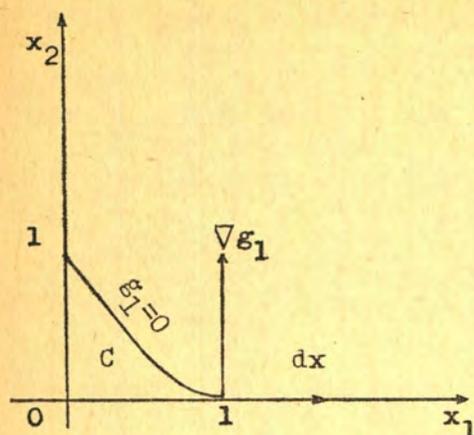


Fig. 1

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 $\bar{\lambda} \geq 0$ tal que*

$$\begin{aligned} \nabla f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x}) &\leq 0 \\ [\nabla f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x})] \bar{x} &= 0 \\ \bar{\lambda} G(\bar{x}) &= 0. \end{aligned}$$

Observemos que a proposição é obviamente 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

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

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

$$2) \quad \nabla f(\bar{x})(x - \bar{x}) \leq 0.$$

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

$$1') \quad \nabla g_i(\bar{x})x \leq \nabla g_i(\bar{x})\bar{x}$$

$$2') \quad \nabla f(\bar{x})x \leq \nabla f(\bar{x})\bar{x}.$$

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

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

com as restrições

$$4) \quad \nabla g_i(\bar{x})x \leq \nabla g_i(\bar{x})\bar{x} \quad \forall i \in \alpha.$$

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 $\lambda_i \geq 0$. O dual requer a minimização de

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

sujeita às restrições

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

Consequentemente, achar $\bar{x} = (\bar{x}_1, \dots, \bar{x}_n)$ que maximiza $f(x)$ reduz-se ao problema de encontrar $\lambda_i \geq 0$ 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

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

Como

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

e, pelo teorema da dualidade na programação linear,

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

é claro que

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

e também é evidente que

$$9) \quad \bar{\lambda} G(\bar{x}) = 0.$$

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 $\bar{\lambda} \geq 0$ e $G(\bar{x}) \leq 0$, vem

$$f(\bar{x}) \leq f(x) - \bar{\lambda} G(x)$$

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

$$\begin{aligned} f(\bar{x}) &\leq f(x) + \nabla f(\bar{x})(x - \bar{x}) \\ -G(\bar{x}) &\leq -G(x) - \nabla G(\bar{x})(x - \bar{x}) \end{aligned}$$

donde resulta

$$\begin{aligned} f(\bar{x}) &\leq f(x) + \nabla f(\bar{x})(x - \bar{x}) - \bar{\lambda} G(x) - \\ &\quad - \bar{\lambda} \nabla G(\bar{x})(x - \bar{x}) = f(\bar{x}) + \\ &\quad + [\nabla f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x})] x - \\ &\quad - [\nabla f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x})] \bar{x} \leq f(\bar{x}), \end{aligned}$$

isto é,

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

$$\begin{aligned} \nabla f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x}) &\leq 0 \\ [\nabla f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x})] \bar{x} &= 0 \\ \bar{\lambda} G(\bar{x}) &= 0. \end{aligned}$$

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

$$\bar{\lambda}_i \geq 0 \quad (i = 1, \dots, m)$$

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

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

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

ou também

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

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

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

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

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

com as restrições

$$\begin{aligned} G(x' - x'') &\leq 0 \\ x' \geq 0 \quad x'' &\geq 0. \end{aligned}$$

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

$$\begin{aligned} \nabla f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x}) &\leq 0 \\ -\nabla f(\bar{x}) + \bar{\lambda} \nabla G(\bar{x}) &\leq 0, \end{aligned}$$

equivalente a

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

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

$$\bar{\lambda} \geq 0$$

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

$$\bar{\lambda} G(\bar{x}) = 0.$$

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

$$\max f(x)$$

com as restrições

$$G(x) \leq 0$$

e portanto as condições de KUHN-TUCKER são

$$\bar{\lambda} \geq 0$$

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

$$\bar{\lambda} G(\bar{x}) = 0.$$

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

$$\max f(x)$$

com as restrições

$$G_p(x) \leq 0$$

$$G_q(x) = 0$$

podem considerar-se as restrições na forma

$$G_p(x) \leq 0$$

$$G_q(x) \leq 0$$

$$-G_q(x) \leq 0.$$

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

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

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

$$\bar{\lambda} \geq 0$$

$$\nabla f(\bar{x}) - \bar{\lambda}^p \nabla G_p(\bar{x}) - \bar{\mu}^q \nabla G_q(\bar{x}) +$$

$$+ \bar{\nu}^q \nabla G_q(\bar{x}) = 0$$

$$\bar{\lambda}^p G_p(\bar{x}) = 0$$

ou, simplesmente,

$$\nabla f(\bar{x}) - \bar{\lambda}^p \nabla G_p(\bar{x}) - \bar{\lambda}^q \nabla G_q(\bar{x}) = 0$$

$$\bar{\lambda}^p G_p(\bar{x}) = 0$$

com $\bar{\lambda}^p \geq 0$ e $\bar{\lambda}^q = \bar{\mu}^q - \bar{\nu}^q$ (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

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

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

Tomando a função lagrangeana

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

diz-se que o ponto

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

é ponto-sela de $F(x, \lambda)$ se

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

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:

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

$$\begin{aligned} \nabla_\lambda F(x, \lambda) &= \left[\frac{\partial F}{\partial \lambda_1}, \dots, \frac{\partial F}{\partial \lambda_m} \right] = \\ &= -[g_1, \dots, g_m] = -G^T(x) \end{aligned}$$

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

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

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

$$\nabla_x F(\bar{x}, \bar{\lambda}) \leq 0 \quad \nabla_x F(\bar{x}, \bar{\lambda}) \cdot \bar{x} = 0 \quad \bar{x} \geq 0$$

$$\nabla_\lambda F(\bar{x}, \bar{\lambda}) \geq 0 \quad \nabla_\lambda F(\bar{x}, \bar{\lambda}) \cdot \bar{\lambda} = 0 \quad \bar{\lambda} \geq 0.$$

Estas condições, juntamente com

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

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

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

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

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

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

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

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

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

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

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

Consequentemente, tem-se

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

desde que $\bar{x} \geq 0$ 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

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

As duas primeiras e a quarta dão

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

IV. *Uma condição necessária (suficiente) para que \bar{x} seja solução ótima 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

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

e portanto

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

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

o que dá a primeira condição necessária.

De

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

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 $\bar{\lambda} \geq 0$, $F(x, \bar{\lambda}) = f(x) - \bar{\lambda} G(x) \leq F(\bar{x}, \bar{\lambda}) + \nabla_x F(\bar{x}, \bar{\lambda})(x - \bar{x})$ e, como

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

(teorema III)

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

vem

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

Ainda pelo teorema III

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

e vem

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

Dado que

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

resulta

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

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 $G(x) \leq 0$ e $x \geq 0$ é que \bar{x} e $\bar{\lambda}$ ($\bar{\lambda} \geq 0$) 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)

$$G(x) \geq G(\bar{x}) + \nabla G(\bar{x})(x - \bar{x})$$

$$f(x) \leq f(\bar{x}) + \nabla f(\bar{x})(x - \bar{x})$$

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

$$\begin{aligned} F(x, \bar{\lambda}) = f(x) - \bar{\lambda} G(x) &\leq f(\bar{x}) - \bar{\lambda} G(\bar{x}) + \\ &+ [\nabla f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x})](x - \bar{x}) = F(\bar{x}, \bar{\lambda}) + \\ &+ \nabla_x F(\bar{x}, \bar{\lambda})(x - \bar{x}) \end{aligned}$$

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

$$\begin{aligned} F(\bar{x}, \lambda) &= f(\bar{x}) - \lambda G(\bar{x}) = f(\bar{x}) - \bar{\lambda} G(\bar{x}) - \\ &- \lambda G(\bar{x}) + \bar{\lambda} G(\bar{x}) = F(\bar{x}, \bar{\lambda}) - (\lambda - \bar{\lambda}) G(\bar{x}) = \\ &= F(\bar{x}, \bar{\lambda}) + \nabla_{\lambda} F(\bar{x}, \bar{\lambda}) (\lambda - \bar{\lambda})^T. \end{aligned}$$

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

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

$$x_j \geq 0 \quad (j = 1, \dots, n),$$

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

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

$$y_i \geq 0 \quad (i = 1, \dots, m).$$

A função lagrangeana para o primal é

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

e para o dual é

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

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

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

o que implica

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

Para os programas matemáticos em geral considera-se *problema primal* aquele que consiste em maximizar a função côncava $f(x)$ sujeita às restrições convexas $G(x) \leq 0$ e $x \geq 0$; 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

$$\lambda \geq 0.$$

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

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

possui inversa, então \bar{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

$$f(x_1, x_2) = \frac{(x_1 + 1)^5}{3} + x_2$$

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

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:

$$\begin{aligned} \text{maximizar } F(x_1, x_2) &= -f(x_1, x_2) = \\ &= -\frac{(x_1 + 1)^5}{3} - x_2 \end{aligned}$$

com as restrições

$$\begin{aligned} 1 - x_1 &\leq 0 \\ x_1 &\geq 0, \quad x_2 \geq 0. \end{aligned}$$

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

$$\begin{aligned} -(x_1 + 1)^2 + \lambda_1 &\leq 0 \\ -1 &\leq 0 \end{aligned}$$

$$\begin{aligned} [-(x_1 + 1)^2 + \lambda_1]x_1 + (-1)x_2 &= 0 \\ \lambda_1(1 - x_1) &= 0. \end{aligned}$$

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 $f(x_1, x_2) = 4x_1 + 6x_2 - x_1^5 - 2x_2^2$ com as restrições

$$\begin{aligned} x_1 + 3x_2 - 8 &\leq 0 \\ 5x_1 + 2x_2 - 14 &\leq 0 \\ x_1 &\geq 0, \quad x_2 \geq 0. \end{aligned}$$

Neste caso tem-se

$$\frac{\partial f}{\partial x_1} = 4 - 5x_1^4, \quad \frac{\partial^2 f}{\partial x_1^2} = -20x_1^3 \leq 0$$

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

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

e portanto a função é côncava pois

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

e

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

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

$$\begin{aligned} 4 - 3x_1^2 - \lambda_1 - 5\lambda_2 &\leq 0 \\ 6 - 4x_2 - 3\lambda_1 - 2\lambda_2 &\leq 0 \\ (4 - 3x_1^2 - \lambda_1 - 5\lambda_2)x_1 + \\ + (6 - 4x_2 - 3\lambda_1 - 2\lambda_2)x_2 &= 0 \end{aligned}$$

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

Há que considerar os casos seguintes:

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

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

$$4 - 3x_1^2 \leq 0 \Rightarrow x_1 \geq 2/\sqrt{3}$$

$$6 - 4x_2 \leq 0 \Rightarrow x_2 \geq 3/2.$$

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

$$4 - 3x_1^2 - 5\lambda_2 \leq 0$$

$$6 - 4x_2 - 2\lambda_2 \leq 0$$

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

$$\lambda_2(5x_1 + 2x_2 - 14) = 0.$$

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

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

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

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

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

$$4 - 3x_1^2 - \lambda_1 \leq 0$$

$$6 - 4x_2 - 3\lambda_1 \leq 0$$

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

$$\lambda_1(x_1 + 3x_2 - 8) = 0.$$

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

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

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

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

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

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

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

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

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

$$\begin{aligned} x_1 &\geq 5 \\ x_1 + x_2 &\geq 10 \\ x_1 + x_2 + x_3 &\geq 15 \\ x_j &\geq 0 \quad (j = 1, 2, 3). \end{aligned}$$

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 $(\lambda_i \geq 0)$

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

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

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 ótima.

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

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

A função $f(x_1, x_2)$ é côncava pois

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

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

$$\begin{aligned} 4 + x_2 - 2x_1 - \frac{\lambda_1}{2} &= 0 \\ 5 + x_1 - 2x_2 - \lambda_1 &= 0 \\ \lambda_1 \left(\frac{x_1}{2} + x_2 - k \right) &= 0. \end{aligned}$$

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

$$2e^{x_1} + x_1x_2 + x_2^2 \leq 10$$

$$x_1 \geq 0, \quad x_2 \geq 0$$

Observemos que

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

e portanto $f(x_1, x_2)$ é côncava. Também é fácil ver que $g(x_1, x_2) = 2e^{x_1} + x_1x_2 + x_2^2$ 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 $x_1 \geq 0$.

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

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

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

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

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

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

ou seja

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

$$10 - 4x_2^5 - \lambda_1(x_1 + 2x_2) \leq 0$$

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

$$\lambda_1(2e^{x_1} + x_1x_2 + x_2^2 - 10) = 0$$

Com $\lambda_1 = 0$, teria de ser $\begin{cases} 8 - 4x_1 \leq 0 \\ 10 - 4x_2^5 \leq 0 \end{cases}$

ou $\begin{cases} x_1 \geq 2 \\ x_2 \geq \sqrt[5]{5/2} \end{cases}$, 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 $2e^{x_1} + x_1x_2 + x_2^2 - 10 = 0$. 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

$$\begin{cases} 8 - 4x_1 - \lambda_1(2e^{x_1} + x_2) = 0 \\ 10 - 4x_2^5 - \lambda_1(x_1 + 2x_2) = 0 \\ 2e^{x_1} + x_1x_2 + x_2^2 - 10 = 0 \end{cases}$$

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

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

$$x_1 \geq 0 \quad x_2 \geq 0.$$

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

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

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

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

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

ou

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

$$-\lambda_1 \leq 0$$

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

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

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

$$1) \quad g_i(x) \leq 0 \quad (i=1, \dots, m).$$

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

$$2) \quad \lambda_i \geq 0 \quad (i=1, \dots, m)$$

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

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

Em certas condições, se x^0 é solução óptima do programa matemático, então existe um vector $\lambda^0 = (\lambda_1^0, \dots, \lambda_m^0)$ 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 numericamente. 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

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

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

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

Admitindo que as $m+n$ equações 3) e 4') nas $m+n$ incógnitas $\lambda_1, \dots, \lambda_m, x_1, \dots, x_n$ 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,

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

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,

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

A resposta ao problema inicialmente proposto está dada. A rentabilidade 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 $x = (x_1, \dots, x_n)$. 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

$$g_i(x) \leq 0 \quad (i = 1, \dots, m) \\ x \geq 0.$$

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 $\lambda_1, \dots, \lambda_m$ (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:

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

isto é, \bar{x} é admissível.

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

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

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

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

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

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

$$g_i(x) \leq 0 \quad (i = 1, \dots, m)$$

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 $\lambda_i \geq 0$). O problema da indústria consiste agora em maximizar (em x , para um dado vector $\lambda \geq 0$) 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 $\lambda \geq 0$. 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 $\lambda_1^0 \geq 0, \dots, \lambda_m^0 \geq 0$.
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 $g_i(x^0)$ ($i = 1, \dots, m$):
 - 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 $0 \leq \lambda_i^1 < \lambda_i^0$.
 - 3.3 Se $g_i(x^0) \leq 0$ 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 $u(x_1, \dots, x_n)$ 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, $x \geq 0$, 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

$$px = M$$

$$x \geq 0$$

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

$$G(x) = 0$$

$$x \geq 0$$

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

reduzem-se a

$$\begin{aligned} \nabla f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x}) &\leq 0 \\ [f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x})] \bar{x} &= 0, \end{aligned}$$

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

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

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

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

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

$$\begin{aligned} G(x) &\leq 0 \\ -G(x) &\leq 0 \\ x &\geq 0. \end{aligned}$$

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

$$\begin{aligned} \nabla f(\bar{x}) - \bar{\lambda}_1 \nabla G(\bar{x}) + \bar{\lambda}_2 \nabla G(\bar{x}) &\leq 0 \\ [\nabla f(\bar{x}) - \bar{\lambda}_1 \nabla G(\bar{x}) + \bar{\lambda}_2 \nabla G(\bar{x})] \bar{x} &= 0 \end{aligned}$$

ou, simplesmente,

$$\begin{aligned} \nabla f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x}) &\leq 0 \\ [\nabla f(\bar{x}) - \bar{\lambda} \nabla G(\bar{x})] \bar{x} &= 0, \end{aligned}$$

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

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

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

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

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

Sejam j_1, \dots, j_s os bens consumidos, tem-se então

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

e para os bens não consumidos é

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

Para a função de utilidade

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

com

$$4x_1 + x_2 = 1,$$

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

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

que satisfazem às proposições enunciadas:

$$\begin{aligned} \left(\frac{\partial u}{\partial x_2} \right)_{(0,1)} / 1 &= 1 = \lambda \\ \left(\frac{\partial u}{\partial x_1} \right)_{(0,1)} / 4 &= 1/2 < \lambda. \end{aligned}$$

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

$$x \geq 0.$$

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 ótima, 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

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

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

$$\begin{aligned} Bx &\leq b \\ x &\geq 0 \end{aligned}$$

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

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

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

- 1) $\max \{f(x) : Bx \leq b \wedge x \geq 0\}$
- 2) $\max \{f(x) : Bx = b \quad x \geq 0\}$
- 3) $\max \{f(x) : Bx \leq b\}.$

Em 2) as inequações transformaram-se em equações mediante a introdução de variáveis auxiliares e em 3) as restrições $x_j \geq 0$ (se existem) foram integradas no sistema de restrições $Ax \leq b$.

Como se disse no n.º 1, não existem actualmente algoritmos eficientes para a pesquisa da solução ótima 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

$$\begin{aligned} Bx &= b \\ x &\geq 0, \end{aligned}$$

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

$$\nabla f(\bar{x}) - y^T \nabla G(\bar{x}) \leq 0 \quad (G(\bar{x}) = B\bar{x} - b = 0)$$

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

ou, fazendo

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

com

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

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

$$-v^T \bar{x} = 0$$

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 $x \geq 0$ e $v \geq 0$ 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 mini-

mizar 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 $x \geq 0$, $v \geq 0$, $z_1 \geq 0$, $z_2 \geq 0$ e com uma regra adicional: para $j = 1, \dots, n$, 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

$$x_1 + x_2 = 2$$

$$x_1 \geq 0, x_2 \geq 0.$$

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,

$$x_1 + x_2 + z_{11} = 2$$

$$2x_1 + y_1 - v_1 + z_{21} = 4$$

$$2x_2 + y_1 - v_2 + z_{22} = 2$$

$$v_1 x_1 + v_2 x_2 = 0.$$

A função-objectivo que deve ser minimizada é

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

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

$$\begin{aligned} &\text{minimizar } z_{11} + z_{21} + z_{22} \\ &(\text{maximizar } -z_{11} - z_{21} - z_{22}) \end{aligned}$$

com as restrições

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

$$\begin{aligned} x_1 \geq 0, x_2 \geq 0, u_1 \geq 0, u_2 \geq 0, v_1 \geq 0, \\ v_2 \geq 0, z_{11} \geq 0, z_{21} \geq 0, z_{22} \geq 0. \end{aligned}$$

Tomando a matriz

$$D = \begin{bmatrix} D_1 & D_2 & D_3 & D_4 & D_5 & D_6 & D_7 & D_8 & D_9 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 2 & 0 & 1 & -1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 1 & -1 & 0 & -1 & 0 & 0 & 1 \end{bmatrix}$$

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

Base	c_j	D	1	2	3	4	5	6	7	8	9
		0	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
$\leftarrow D_9$	-1	2	0	<u>2</u>	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	0

Os restantes quadros derivam-se da maneira habitual:

Base	c_j	D	1	2	3	4	5	6	7	8	9
		0	0	0	0	0	0	0	-1	-1	-1
$\leftarrow D_7$	-1	1	<u>1</u>	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	0	-3/2

Base	c_j	D	1	2	3	4	5	6	7	8	9
		0	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
$\leftarrow D_8$	-1	2	0	0	<u>2</u>	-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	0

Base	c_j	D	1	2	3	4	5	6	7	8	9
		0	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	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

$$\nabla f(\bar{x})\bar{x} = \max \{ \nabla f(\bar{x})w : w \geq 0 \wedge Bw \leq b \}$$

onde

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

O teorema da dualidade da programação linear justifica que $\max \{ \nabla f(\bar{x})w : w \geq 0 \wedge Bw \leq b \} = \min \{ u^T b : u \geq 0 \wedge u^T B \geq \nabla f(\bar{x}) \}$ onde $u = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}$. Então a condição neces-

sária e suficiente para que \bar{x} seja solução óptima é que

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

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

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

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

$$\begin{aligned} \bar{x} \geq 0, \bar{u} \geq 0, B\bar{x} \leq b, \nabla f(\bar{x}) \leq \bar{u}^T B, \\ F(\bar{x}, \bar{u}) = 0. \end{aligned}$$

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

$$\begin{aligned} x \geq 0, u \geq 0, v \geq 0, y \geq 0 \\ Bx + y = b \\ 2Ax + B^T u - v = c \\ v^T x + u^T y = 0, \end{aligned}$$

maximizar

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

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

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

com

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

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

$$\text{lugares, obtenha-se } w_1 = \begin{bmatrix} x \\ u \\ y \\ v \end{bmatrix}.$$

5. Calcula-se $v^T x + u^T y$.
- Se $v^T x + u^T y = 0$, o algoritmo termina.
 - 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 $F(w_2) = v^T x + u^T y$.
- Se $v^T x + u^T y = 0$, o algoritmo termina.
 - Se $v^T x + u^T y \neq 0$, e
 - Se $v^T x + u^T y$ não decresceu em relação à fase 5, determine-se λ tal que $F(w) = F[w_1 + \lambda(v_2 - w_1)]$ ($0 \leq \lambda \leq 1$) 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

$$\begin{aligned} x_1 + 2x_2 &\leq 4 \\ x_1 &\geq 0, \quad x_2 \geq 0. \end{aligned}$$

Para este problema é

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

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

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

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

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

vêm os coeficientes seguintes

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

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

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

$$D_5 : \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad D_4 : \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

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

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

Podê pois construir-se o quadro

	w_1	1	2	3	4	5	6	$w_1 = \begin{bmatrix} 0 \\ 0 \\ 4 \\ 4 \\ 0 \\ 2 \end{bmatrix}$
D_3	4	2	0	1	0	-1	0	x
D_4	4	1	2	0	1	0	0	u
D_6	2	4	-6	0	0	-2	1	y
								v

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

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

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

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

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	<u>4</u>	-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

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

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

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

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

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

Tem-se agora o quadro

	$\nabla F(w_2)$	w_2	0	0	-7/2	-3	-1/2	0
$\leftarrow D_3$	-7/2	3	0	<u>3</u>	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 facilmente:

	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

$$w_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \left. \begin{array}{l} x \\ u \\ y \\ v \end{array} \right\}$$

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

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 explicitamente 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

$$Bx = b$$

$$x \geq 0.$$

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 $\partial f / \partial z_k \leq 0$, tem o maior valor no ponto inicial. É claro que, na direcção de z_k , f decresce mais rapidamente e portanto a variação ao longo de z_k conduz mais rapidamente 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 $x_j \geq 0$. 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 decréscimo 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

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

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, \dots, z_m são variáveis não-básicas. Se $\partial f / \partial z_k < 0$ para algum k ($k = 1, \dots, n-m$), 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

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

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 o menor de todos. Segue-se a Fase 5a se z_k 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 , 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

$$\begin{aligned} x_1 + x_2 &\leq 2 \\ x_1 &\geq 0, \quad x_2 \geq 0. \end{aligned}$$

Tomando a variável auxiliar $x_3 \geq 0$, 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

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

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

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 $\frac{\partial f}{\partial x_1} = 0$ 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 proçeda-se à substituição em f . Vem

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

Tem-se agora

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

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

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

$$\frac{\partial f}{\partial x_3} = x_3 + 1 \quad \left(\frac{\partial f}{\partial x_3} \right)_{\substack{x_3=0 \\ u_1=0}} = 1 > 0$$

$$\frac{\partial f}{\partial u_2} = u_2 \quad \left(\frac{\partial f}{\partial u_2} \right)_{\substack{x_3=0 \\ u_1=0}} = 0$$

e portanto nenhuma variação em x_3 ou u_2 produz decréscimo em f e alcançamos 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 (possivelmente infinito) de óptimos locais, é apenas necessário encontrar um óptimo local pois este será necessariamente 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

$$A_i x \leq b_i,$$

consideremos um ponto arbitrário x^0 (interior ou fronteiro) pertencente ao conjunto admissível ($A_i x^0 \leq b_i$). 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

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

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 $g_i(x) \geq 0$ ($i=1, 2, \dots, m+n$) que incluem as condições de não-negatividade $x \geq 0$. 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 $g_i(x) \geq 0$. Notando que $1/g_i(x)$ tende para $+\infty$ quando $g_i(x) \rightarrow +0$, é evidente que o ponto x^* que minimiza $P(x; r)$ satisfaz a $g_i(x^*) > 0$ ($i=1, 2, \dots, m+n$). 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 $g_i(x) \geq 0$, 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

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

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 , tal que

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

Então, para $k > k_0$, vem

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

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 previamente escolhido. Para isso basta resolver o sistema de equações

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

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

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

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

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

A solução minimizante de $P(x_1, x_2; r)$ obtém-se facilmente neste caso. Trata-se de $x_1 = \sqrt{1 + \sqrt{r}}$, $x_2 = \sqrt{r}$. Quando $r \rightarrow 0$, vem a solução optimizante $x_1 = 1$, $x_2 = 0$ já encontrada anteriormente por meio das condições de KUHN-TUCKER (Exemplo 3.1).

BIBLIOGRAFIA

- 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*, Addison-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*, Addison-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.

Sobre o Teorema dos Códigos para um Canal sem Ruído (*)

por J. Marques Henriques
Universidade Técnica de Lisboa

0. Sumário.

É examinado o teorema dos códigos para um canal sem ruído dum ponto de vista probabilístico, num caso especial, tomando a expansão relativa à base θ dum ponto do intervalo unitário, considerando-a como a realização do processo estocástico a ser codificado e relacionando-a com propriedades conhecidas do intervalo unitário, em particular no que se refere à dimensão de HAUSDORFF BESICOVITCH e ao teorema de SHANNON-McMILLAN-BREIMAN.

1. Generalidades e propriedades básicas.

Seja $\Omega = (0, 1]$ e \mathfrak{R} o conjunto dos borelianos de Ω . Para cada $\omega \in \Omega$ associemos-lhe a sua expansão infinita relativamente a uma base $\theta \geq 2$ (θ inteiro), ou seja,

$$\omega = \sum_{i=1}^{\infty} a_i(\omega)/\theta^i,$$

onde $a_i(\omega) = 0, 1, \dots, \theta - 1$. Então cada $a_i(\omega)$ é uma função mensurável relativamente a (Ω, \mathfrak{R}) (para detalhes relativamente à notação ver [4]).

Deste modo, se P for uma medida de probabilidade sobre \mathfrak{R} , $\{a_i(\omega) : i = 1, 2, \dots\}$ é um processo estocástico relativamente a $(\Omega, \mathfrak{R}, P)$, com espaço-amostra finito, $\Xi = \{0, 1, \dots, \theta - 1\}$, a que aqui chamamos *alfabeto*, sendo os elementos $0, 1, \dots, \theta - 1$ as suas *letras*.

Inversamente, pode provar-se que se a medida de probabilidade P for não-atômica (isto é, se com $P(A) > 0$ existir sempre um $B \subset A$ com $0 < P(B) < P(A)$), então qualquer processo estocástico com alfabeto Ξ pode representar-se desta forma.

Neste trabalho ocupar-nos-emos em particular daquelas medidas não-atômicas P , relativamente às quais o processo estocástico $\{a_i(\omega) : i = 1, 2, \dots\}$ é estacionário e ergódico.

Por definição o processo é *estacionário* se $P(\omega : a_1(\omega), \dots, a_n(\omega)) = P(\omega : a_{k+1}(\omega), \dots, a_{k+n}(\omega))$, com qualquer inteiro k . Quanto à *ergodicidade*, diremos que para a transformação T definida sobre Ω por $T\omega = \theta\omega \pmod{1}$ A é invariante se para $A \in \mathfrak{R}$, $T^{-1}A = A$, e que P é ergódica relativamente a T se para todo o $A \in \mathfrak{R}$ invariante $P(A) = 0$ ou $P(A) = 1$.

DEFINIÇÃO 1. Chamaremos *código* a qualquer função χ contínua e não decrescente em $[0, 1]$, tal que $\chi(0) = 0$ e $\chi(1) = 1$.

Em correspondência com cada ω obtemos então a expansão de $\chi(\omega)$ relativamente à

(*) Trabalho apresentado no «Simpósio sobre as Teorias da Informação e dos Sistemas», Lisboa, Setembro de 1970.

base θ , de tal modo que virá

$$\chi(\omega) = \sum_{i=1}^{\infty} b_i(\omega)/\theta^i.$$

Agora podemos interpretar o código χ como um esquema tal que a cada sucessão ou mensagem original $a = (a_1, a_2, \dots)$ de letras de Ξ lhe faz corresponder uma outra sucessão ou mensagem codificada $b = (b_1, b_2, \dots)$. Por uma questão de simplicidade só consideraremos códigos com o mesmo alfabeto Ξ para domínio e contradomínio, aos quais também chamaremos *input* e *output*.

Vamos assumir daqui em diante que todas as medidas de probabilidade P são não-atômicas; então o código χ é tal que com probabilidade 1 podemos determinar as primeiras n letras de b conhecido que seja um número finito de letras de a .

Em todo este artigo só nos ocuparemos de um canal sem ruído, isto é, um código ao qual estão associados um input e um output, cada um dos quais possui o alfabeto Ξ . Do ponto de vista matemático o canal sem ruído é simplesmente o alfabeto Ξ , ao qual se associam todas as possíveis mensagens. Se uma mensagem a é enviada através do canal, ela é recebida com perfeito rigor depois de codificada, b .

O teorema dos códigos para um canal sem ruído, de que é uma versão especial no nosso caso o Teorema 5 abaixo, pretende estabelecer condições segundo as quais uma mensagem, depois de codificada, $b = \chi(a)$, pode ser retransmitida do output para o input e aí reconstituída a mensagem inicial ([6] e [1]).

DEFINIÇÃO 2. Intervalo fundamental de ordem n de Ω é um subconjunto $\Delta_{k_1, \dots, k_n}^{(n)}$, tal que

$$\Delta_{k_1, \dots, k_n}^{(n)} = \{\omega \in \Omega : a_i(\omega) = k_i, \\ i = 1, \dots, n \text{ e } k_i \in \Xi\}.$$

É evidente que para qualquer n , $\Delta^{(n)}$ é um conjunto de \mathfrak{R} ; inversamente, os conjuntos $\Delta^{(n)}$ geram a σ -álgebra \mathfrak{R} .

No que se segue omitimos os índices k_1, \dots, k_n , visto não haver qualquer possibilidade de confusão, pois trata-se de elementos previamente escolhidos de Ξ . Também no caso de n ser bem definido escrevemos apenas $\Delta(\omega)$ ou Δ .

Da Definição 2 imediatamente se deduz que para todo o $n \geq 1$ e $\omega \in \Omega$ existe um e um só intervalo fundamental de ordem n que contém ω . Indicá-lo-emos por $\Delta^{(n)}(\omega)$ ou $\Delta(\omega)$ sempre que seja necessário individualizá-lo. Este é então o intervalo θ -ádico do tipo $(l/\theta^n, (l+1)/\theta^n]$, com $l = 0, 1, \dots, \theta^n - 1$ que contém ω , pois os θ^n intervalos de ordem n possuem todos o mesmo comprimento, pelo que se se designar por λ a clássica medida de LEBESGUE sobre Ω será $\lambda(\Delta) = \theta^{-n}$.

Mas agora, se os n primeiros símbolos da expansão de ω são conhecidos, também o é Δ , e é então evidente que $\chi(\omega) \in \chi(\Delta)$. Denotemos por $\{\chi(\Delta)\}$ o intervalo θ -ádico mais pequeno que contém $\chi(\Delta)$; deste modo o número de símbolos do alfabeto Ξ na expansão de $\chi(\omega)$ que pode ser determinado de maneira unívoca é exactamente a ordem do intervalo θ -ádico $\{\chi(\Delta)\}$ mas esta é $-\log_{\theta} \lambda \{\chi(\Delta)\}$, onde, tal como acima, λ é a medida de LEBESGUE. Isso é, aliás, óbvio, no caso de ser $\chi(\omega) \equiv \omega$, por vir $-\log_{\theta} \lambda(\Delta) = n$.

Deste modo, os primeiros n símbolos da expansão de ω determinam exactamente $-\log_{\theta} \lambda \{\chi(\Delta^{(n)}(\omega))\}$ símbolos na expansão de $\chi(\omega)$.

Assim, podemos medir a eficiência do código χ , considerando a relação entre o número de símbolos determinados na expansão de $\chi(\omega)$ e o número de símbolos da expansão de ω necessários para a determinação daquelas, a que chamamos a compressão originada

por χ , e que denotaremos por

$$C_n(\omega) := -n^{-1} \log_{\theta} \lambda \{ \chi(\Delta^{(n)}(\omega)) \}.$$

Para simplificar as deduções que vamos fazer substituiremos C_n por

$$D_n(\omega) := -n^{-1} \log_{\theta} \lambda(\chi(\Delta^{(n)}(\omega))).$$

Um código χ será *eficiente* se $C_n(\omega)$ for pequeno, no limite para n infinito. Assim, definimos

$$C_{\chi}(\omega) := \lim_{n \rightarrow \infty} C_n(\omega)$$

e

$$C_{\chi}^*(\omega) := \liminf_{n \rightarrow \infty} C_n(\omega)$$

no caso de existirem ambos os limites, e análogamente para $D_{\chi}(\omega)$ e $D_{\chi}^*(\omega)$.

Suponhamos agora que é dada uma medida de probabilidade P sobre (Ω, \mathfrak{R}) , tal que P é estacionária, não-atômica e ergódica. Se for $F(x) := P(0, x]$ a função de distribuição de P , então F é um código. Mostraremos abaixo que para este código é

$$P(\omega : D_F(\omega) = h) = 1,$$

onde h é a entropia relativa de $\{a_i\}$ em relação a P , ou seja, $h = \lim_{n \rightarrow \infty} \int_{\Omega} D_n(\omega) dP$ (a entropia relativa será simplesmente a entropia dividida por $\log \theta$).

Mostraremos ainda que para qualquer outro código χ é:

$$P(\omega : D_{\chi}^*(\omega) < h) = 0,$$

de modo que F possui a eficiência máxima, como aliás seria de esperar. Quer isto dizer que para um dado processo um código será optimal se considerado como função sobre Ω , for precisamente a função de distribuição deste processo.

2. O código F .

Se, tal como acima, designarmos por F a função de distribuição de P , então F é um código que goza da propriedade de que com $\omega \neq \omega'$,

$$P(\omega : F(\omega) = F(\omega')) = 0$$

e isto por P ser não-atômica; por outras palavras, a mensagem original pode, com probabilidade 1, deduzir-se da mensagem codificada.

TEOREMA 1. *Se P for não-atômica, estacionária e ergódica, então*

$$P(\omega : D_F(\omega) = h) = 1,$$

onde h é a entropia relativa de $\{a_i\}$ relativamente a P .

DEM.: Por ser P não-atômica, vem $\lambda(F(\Delta)) = P(\Delta)$ para qualquer intervalo Δ (isto é o mesmo que afirmar que se X for uma variável aleatória com função de distribuição $G(x)$, então $G(X)$ é uma variável aleatória uniformemente distribuída no intervalo unitário). Por conseguinte

$$D_n(\omega) = -n^{-1} \log_{\theta} P(\Delta^{(n)}(\omega)),$$

e a afirmação é agora uma consequência do teorema de SHANNON-McMILLAN-BREIMAN [1, pág. 197]. Q. E. D.

3. Alguns tópicos sobre a dimensão de Hausdorff-Besicovitch de subconjuntos de Ω .

Nesta Secção apresentaremos apenas a definição e propriedades mais importantes da dimensão de HAUSDORFF-BESICOVITCH de subconjuntos de Ω , que nos será de utilidade posterior.

Consideremos de novo o espaço de probabilidade $(\Omega, \mathfrak{R}, P)$ não-atômico e seja A um subconjunto qualquer de Ω (não necessariamente em \mathfrak{R}). Considere-se uma *cobertura* de A por conjuntos Z_i de \mathfrak{R} , em número finito ou infinito enumerável:

$$A \subset \bigcup_{i=1}^{\infty} Z_i, \quad Z_i \in \mathfrak{R}.$$

Agora, uma vez que o espaço de probabilidade é não atômico, será sempre possível, dado $\rho > 0$ encontrar conjuntos $\{Z_i\}$ em \mathfrak{R} que cobrem A , com $P(Z_i) < \rho$; chamaremos uma *cobertura* ρ de A . Então, fixado $\alpha > 0$ definimos para qualquer $A \subset \Omega$:

$$\Gamma_{\alpha}(A; \rho) := \inf \left\{ \sum_{i=1}^{\infty} P(Z_i)^{\alpha} \right\};$$

tomado o ínfimo para todas as coberturas ρ de A (com $P(Z_i) < \rho$; nada impede aqui que $\Gamma_{\alpha}(A; \rho)$ tome o valor $+\infty$),

Quando $\rho \downarrow 0$, $\Gamma_{\alpha}(A; \rho)$ é não decrescente, pois com $\rho' < \rho$ toda a cobertura ρ' de A é também uma cobertura ρ de A e deste modo o ínfimo é tomado para classes cada vez mais reduzidas de coberturas de A , e portanto o limite de $\Gamma_{\alpha}(A; \rho)$ para $\rho \downarrow 0$ existe. Então definimos

$$\Gamma_{\alpha}(A) := \lim_{\rho \downarrow 0} \Gamma_{\alpha}(A; \rho).$$

A $\Gamma_{\alpha}(A)$ chama-se por vezes a medida exterior α -dimensional de A . É evidente que $\Gamma_{\alpha}(\cdot)$ é monótona não decrescente: para $A, B \subset \Omega$, se $A \subset B$, então $\Gamma_{\alpha}(A) \leq \Gamma_{\alpha}(B)$.

E, dada uma sucessão de conjuntos A_n de Ω , podem escolher-se para cada um deles coberturas $\{Z_{ni}\}$, com $P(Z_{ni}) < \rho$, tais que

$$\sum_{i=1}^{\infty} P(Z_{ni})^{\alpha} < \Gamma_{\alpha}(A_n; \rho) + \varepsilon 2^{-n} \leq \Gamma_{\alpha}(A) + \varepsilon 2^{-n}.$$

Deste modo a totalidade dos Z_{ni} constitui uma cobertura ρ da união $\bigcup_n A_n$ tal que

$$\sum_n \sum_i P(Z_{ni})^{\alpha} < \sum_n \Gamma_{\alpha}(A_n) + \varepsilon$$

e assim vemos que $\Gamma_{\alpha}(\cdot)$ é uma função sub-aditiva, ou seja:

$$\Gamma_{\alpha}\left(\bigcup_n A_n\right) \leq \sum_n \Gamma_{\alpha}(A_n).$$

Quer dizer, tomada como função de A , $\Gamma_{\alpha}(A)$ é, de facto, uma medida exterior no sentido de CARATHÉODORY.

TEOREMA 2. Com $\alpha > 0$, $\Gamma_{\alpha}(A) < \infty$ implica $\Gamma_{\alpha+\varepsilon}(A) = 0$ para todos os $\varepsilon > 0$.

DEM. Se for $\{Z_i\}$ uma cobertura ρ de A para a qual

$$\sum_i P(Z_i)^{\alpha} \leq \Gamma_{\alpha}(A; \rho) + \varepsilon \leq \Gamma_{\alpha}(A) + 1 = K < \infty$$

(da Definição de $\Gamma_{\alpha}(A; \rho)$, uma tal cobertura existe sempre), então também

$$\begin{aligned} \Gamma_{\alpha+\varepsilon}(A; \rho) &\leq \sum_i P(Z_i)^{\alpha+\varepsilon} \leq \\ &\leq \rho^{\varepsilon} \sum_i P(Z_i)^{\alpha} < \rho^{\varepsilon} K. \end{aligned}$$

E, por ser $\varepsilon > 0$, fazendo $\rho \downarrow 0$ logo se vê que $\Gamma_{\alpha+\varepsilon}(A) = 0$. Q. E. D.

Anàlogamente se mostra que se $\Gamma_{\alpha}(A) > 0$ então para todos os ε tais que $0 < \varepsilon < \alpha$ vem $\Gamma_{\alpha-\varepsilon}(A) = \infty$.

Quer dizer: há um ponto α_0 , tal que $\Gamma_{\alpha}(A) = \infty$ para $\alpha < \alpha_0$ e $\Gamma_{\alpha}(A) = 0$ para $\alpha > \alpha_0$. O valor de $\Gamma_{\alpha}(A)$ em α_0 pode ser nulo, finito e positivo ou $+\infty$. A este número α_0 , bem determinado, chamamos

então a dimensão de HAUSDORFF-BESICOVITCH de A (relativamente à medida de probabilidade P):

DEF.

$$\begin{aligned} \dim A &:= \sup_P \{ \alpha : \Gamma_\alpha(A) = \infty \} = \\ &= \inf \{ \alpha : \Gamma_\alpha(A) = 0 \}. \end{aligned}$$

No caso de ser $\Gamma_\alpha(A) < \infty$ para todos os $\alpha > 0$, definiremos $\dim A$ como sendo 0 (omitiremos a indicação da medida de probabilidade P sempre que não haja qualquer motivo para dúvidas, o que aliás sucede sempre abaixo; no caso de $P = \lambda$, $\dim A$ coincide com a vulgar dimensão de HAUSDORFF de A).

Mostra-se com facilidade que $\dim \emptyset = 0$ (\emptyset é o conjunto vazio e $\dim \Omega = 1$. Também é válido o

TEOREMA 3. *A dimensão de Hausdorff-Besicovitch goza das propriedades seguintes:*

- a) $A, B \subset \Omega$ então $\dim A \leq \dim B$;
- b) $\dim A = 1$ para todos os $A \in \mathfrak{R}$ com probabilidade positiva;
- c) $0 \leq \dim A \leq 1$;
- d) $\dim \bigcup_{i=1}^{\infty} A_i = \sup_i \dim A_i$.

DEM. a)-c) deduzem-se imediatamente da Definição;

d) é ligeiramente mais complexa: se $\alpha > \sup \dim A_i$, então $\Gamma_\alpha(A_i) = 0$ ($i = 1, 2, \dots$) e por isso, pela subaditividade de $\Gamma_\alpha(\cdot)$ como medida exterior sobre Ω , também

$$\Gamma_\alpha \left(\bigcup_i A_i \right) = 0,$$

o que implica

$$\dim \bigcup_i A_i \leq \alpha;$$

agora, para provar a desigualdade no outro sentido, se $\alpha < \sup \dim A_i$, para i_0 convenientemente fixado é evidente que $\alpha < \dim A_{i_0}$ e deste modo $\Gamma_\alpha(A_{i_0}) = \infty$. Isto acarreta $\Gamma_\alpha \left(\bigcup_i A_i \right) = \infty$ (por a) acima) e por consequência $\dim \bigcup_i A_i \geq \alpha$. Q. E. D.

Um resultado importante, mas que pela extensão da demonstração não provaremos aqui (ver [2] e [4]) é o seguinte:

TEOREMA 4. *Se P for uma medida de probabilidade não atômica sobre \mathfrak{R} , e Π uma outra medida sobre \mathfrak{R} com $\Pi(\Omega) < \infty$, então para o conjunto*

$$A := \left\{ \omega : \liminf_{n \rightarrow \infty} \frac{\log \Pi(\Delta^{(n)}(\omega))}{\log P(\Delta^{(n)}(\omega))} \leq \delta \right\}$$

vem $\dim A \leq \delta$.

4. O código geral.

Nesta secção vamos mostrar que o código F da Secção 2 é optimal no sentido de que para nenhum outro código χ se obtém uma compressão inferior a h .

TEOREMA 5. *Sendo P não-atômica, estacionária e ergódica e χ um código qualquer, então*

$$P(\omega : D_\chi^*(\omega) < h) = 0.$$

DEM. Considere-se a medida de probabilidade Π sobre \mathfrak{R} , cuja função de distribuição é χ . Como χ é contínua, Π é não-atômica e para qualquer intervalo Δ , $\Pi(\Delta) = \lambda(\chi(\Delta))$. Assim,

$$D_n(\omega) = -\frac{1}{n} \log_{\theta} \Pi(\Delta^{(n)}(\omega)) = \\ = \frac{\log_{\theta} \Pi(\Delta^{(n)}(\omega))}{\log_{\theta} P(\Delta^{(n)}(\omega))} \left[-\frac{1}{n} \log_{\theta} P(\Delta^{(n)}(\omega)) \right].$$

Mas, de acordo com o Teorema 1, o segundo factor do lado direito converge para h , a menos de um conjunto de medida P nula. Por isso,

$$D^*(\omega) = h \liminf_{n \rightarrow \infty} \frac{\log_{\theta} \Pi(\Delta^{(n)}(\omega))}{\log_{\theta} P(\Delta^{(n)}(\omega))} [P].$$

Ora, aplicando o Teorema 4 com $\delta = \beta/h$ e $\beta < h$ vem

$$\dim \left(\omega : \liminf_{n \rightarrow \infty} \frac{\log_{\theta} \Pi(\Delta^{(n)}(\omega))}{\log_{\theta} P(\Delta^{(n)}(\omega))} \leq \frac{\beta}{h} \right) < 1.$$

Mas como qualquer conjunto de medida P positiva possui dimensão 1, então isso acarreta que com $\beta < h$

$$P \left(\omega : \liminf_{n \rightarrow \infty} \frac{\log_{\theta} \Pi(\Delta^{(n)}(\omega))}{\log_{\theta} P(\Delta^{(n)}(\omega))} \leq \frac{\beta}{h} \right) = 0,$$

ou seja

$$P(\omega : D_{\chi}^*(\omega) \leq \beta) = 0.$$

Q. E. D.

Antes, na Secção 1, tínhamos assumido que $\chi(0) = 0$ e $\chi(1) = 1$. Contudo podemos relaxar ligeiramente esta hipótese fazendo apenas $0 \leq \chi(0) \leq \chi(1) \leq 1$ e definindo $\Pi(0, x] := \chi(x) - \chi(0)$. Neste caso será ainda $\Pi(\Omega) < \infty$, embora não uma medida de probabilidade, e a demonstração permanece inalterável.

Como $P(\omega : D_{\chi}^*(\omega) > h) = 1$, pelo Lema de FATOU vem

$$\liminf_{n \rightarrow \infty} \int_{\Omega} D_n(\omega) dP \geq h,$$

o que quer dizer que h é não só a compressão média, mas também a compressão mínima, de acordo com o Teorema 5. Este resultado é devido a KHINCHIN [5, pág. 24], embora partindo de definições diferentes e vindo aí o limite superior em vez do limite inferior.

Como exemplo de aplicação consideremos o conjunto de CANTOR, isto é, com $\theta = 3$

$$K := \{ \omega : a_n(\omega) = i; \quad i = 0, 2; \quad n = 1, 2, \dots \}$$

ao qual corresponde a medida de probabilidade singular

$$P(\omega : a_n(\omega) = i) = p_i,$$

com $p_0 = p_2 = 1/2$ e $p_1 = 0$. Mostra-se com facilidade que $\lambda(K) = 0$. Neste caso o código optimal consiste em substituir na expansão de qualquer $\omega \in K$ os dígitos 2 por dígitos 1, considerando as mensagens constituídas por 0's e 1's como expansões binárias tomadas relativamente à base 3. A compressão que assim se obtém é $\log 2 / \log 3$, eventualmente o mesmo valor que para a dimensão de HAUSDORFF-BESICOVITCH deste mesmo conjunto, tomada relativamente à medida de LEBESGUE.

Um exemplo clássico é devido a SHANNON : com $\theta = 4$, façamos $P(\omega : a_n(\omega) = i) = p_i$ com $p_0 = 1/2$, $p_1 = 1/4$, e $p_2 = p_3 = 1/8$. Então substituindo cada letra numa mensagem a codificar por um conjunto de 0's e de 1's, de acordo com o código $0 \rightarrow 0$, $1 \rightarrow 10$, $2 \rightarrow 110$, $3 \rightarrow 111$, vem esta mensagem substituída por uma outra de dígitos binários. Estes símbolos são agora agrupados dois a dois e traduzidos para a base $\theta = 4$ de acordo com a tabela $00 \rightarrow 0$, $01 \rightarrow 1$, $10 \rightarrow 2$ e $11 \rightarrow 3$. Quer dizer, qualquer $\omega = (a_1(\omega), a_2(\omega), \dots)$ vem codificado $\chi(\omega) = (b_1(\omega), b_2(\omega), \dots)$, onde χ é a função de

de distribuição de P , definida acima. Então o código χ é optimal, obtendo-se a compressão média de $7/8$ [6, pág. 63-4].

BIBLIOGRAFIA

- [1] ASH, ROBERT B. (1965). *Information Theory*. John Wiley & Sons, New York — London.
- [2] BILLINGSLEY, PATRICK P. (1960). Hausdorff dimension in probability theory, *Ill. J. Math.* **4** 187-209.
- [3] ——— (1965). *Ergodic Theory and Information*. John Wiley & Sons, New York—London.
- [4] HENRIQUES, J. MARQUES. (1967). *Hausdorff-Besicovitch Dimension for Probability Product Spaces*, Master's Paper, Department of Statistics, The University of Chicago, Chicago, Ill.
- [5] KHINCHIN, A. YA. (1957). *Mathematical Foundations of Information Theory*. Dover Publications, New York.
- [6] SHANNON, CLAUDE E. (1949). *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana, Ill. (with Warren Weaver).

Matrices partitionnées en blocs commutatifs

par José Vitória (1)

Luourenço Marques (Moçambique)

Résumé

En travaillant sur des matrices partitionnées en blocs commutatifs, on obtient des résultats formellement semblables à la règle de Cramer et au théorème de Rouché; et on donne, à partir de là, la solution de l'équation $\mathbf{A}X = \lambda X$, où λ est une valeur propre de \mathbf{A} et où X est soit un vecteur soit une matrice.

Sumário

Trabalhando com matrizes fraccionadas em blocos comutativos, obtemos resultados formalmente semelhantes à regra de Cramer e ao teorema de Rouché; e damos, a partir daí, a solução da equação $\mathbf{A}X = \lambda X$, em que λ é um valor próprio de \mathbf{A} e em que X é quer um vector quer uma matriz.

I. Introduction, définitions, notations

Dans la pratique, quand on a à résoudre des systèmes d'équations différentielles aux dérivées partielles par la méthode des différences finies, la matrice du système est partitionnée en blocs commutatifs et symétriques [5].

Dans cette étude, nous nous intéressons à certains systèmes d'équations linéaires (matricielles), où la matrice du système est partitionnée en blocs commutatifs, qui surgissent, par exemple, quand on a affaire à la résolution simultanée de systèmes d'équations li-

néaires (algébriques) qui diffèrent seulement par les colonnes des termes indépendants. On n'aura pas besoin de calculer le déterminant de la matrice donnée, mais simplement de savoir si une certaine matrice de l'ordre de ses sous-matrices est non-singulière.

\mathbf{K} désigne le corps des nombres réels ou complexes (\mathbf{R} ou \mathbf{C});

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = (a_{ij}) \in \mathbf{M}_{n,n}(\mathbf{K})$$

désigne la matrice de type (n, n) d'éléments a_{ij} appartenant à \mathbf{K} ;

$$\begin{vmatrix} x_1 \\ \vdots \\ x_n \end{vmatrix} = (x_j) = X \in \mathbf{K}^n \quad \text{et} \quad \begin{vmatrix} b_1 \\ \vdots \\ b_m \end{vmatrix} = (b_i) = B \in \mathbf{K}^m$$

vecteurs d'éléments de \mathbf{K} ;

$$D := \det \begin{vmatrix} a_{11} & \cdots & a_{1R} \\ \cdots & \cdots & \cdots \\ a_{R1} & \cdots & a_{RR} \end{vmatrix},$$

déterminant principal; $R \leq n$, rang de la matrice \mathbf{A} ;

$$D_i := \det \begin{vmatrix} a_{11} & \cdots & a_{1,i-1} & \boxed{b_1} & a_{1,i+1} & \cdots & a_{1R} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{R1} & \cdots & a_{R,i-1} & \boxed{b_R} & a_{R,i+1} & \cdots & a_{RR} \end{vmatrix}$$

($i = 1, \dots, R$);

(1) Boursier de la Fondation Calouste Gulbenkian, Lisboa, Portugal.

$$C_j := \det \left| \begin{array}{ccc|c} a_{11} & \cdots & a_{1R} & b_1 \\ \cdots & \cdots & \cdots & \vdots \\ a_{R1} & \cdots & a_{RR} & b_R \\ \hline a_{j1} & \cdots & a_{jR} & b_j \end{array} \right|, \quad (j=R+1, \dots, n),$$

déterminant caractéristique; (si $R=n$, C n'est pas défini);

$$A = \begin{vmatrix} A_{11} & \cdots & A_{1n} \\ \cdots & \cdots & \cdots \\ A_{m1} & \cdots & A_{mn} \end{vmatrix} = (A_{pq}),$$

($p=1, \dots, m$; $q=1, \dots, n$), où $A_{pq} \in M_{s,s}(\mathbb{K})$ et $A_{pq} \cdot A_{\mu\nu} = A_{\mu\nu} \cdot A_{pq}$, c'est-à-dire, on considère les sous-matrices A_{ij} appartenant à un anneau commutatif (relativement à la multiplication et à l'addition) avec élément unité;

$$\Delta := \text{dev} \begin{vmatrix} A_{11} & \cdots & A_{1r} \\ \cdots & \cdots & \cdots \\ A_{r1} & \cdots & A_{rr} \end{vmatrix}$$

désigne la matrice d'ordre s que l'on obtient en développant le déterminant de (A_{ij}) , ($i, j=1, \dots, r$; $r \leq \min(m, n)$) et en considérant les A_{ij} comme éléments (scalaires);

$$\Delta_i := \widetilde{\text{dev}} \begin{vmatrix} A_{11} \cdots A_{1,i-1} & \begin{array}{c} \overline{B_1} \\ \cdots \\ \overline{B_r} \end{array} & A_{1,i+1} \cdots A_{1r} \\ \cdots & \cdots & \cdots \\ A_{r1} \cdots A_{r,i-1} & \begin{array}{c} \overline{B_1} \\ \cdots \\ \overline{B_r} \end{array} & A_{r,i+1} \cdots A_{rr} \end{vmatrix}$$

($i=1, \dots, r$), désigne la matrice de type (r, s) que l'on obtient en développant le déterminant suivant la «colonne» i (colonne de blocs) et en considérant les A_{pq} , ($p=1, \dots, r$; $q=1, \dots, i-1, i+1, \dots, r$), et les matrices B_j^i , ($j=1, \dots, r$) de type (s, t) comme éléments, mais où les facteurs B_j^i sont toujours à droite; (pour que l'on puisse effectuer les multiplications...);

$$\xi_j := \text{Dev} \begin{vmatrix} A_{11} \cdots A_{1r} & B_1 \\ \cdots & \vdots \\ A_{r1} \cdots A_{rr} & B_r \\ \hline A_{j1} \cdots A_{jr} & B_j \end{vmatrix}, \quad (j=r+1, \dots, n),$$

désigne la matrice obtenue en développant le déterminant suivant la dernière «colonne», les A_{pq} , ($p=1, \dots, r$; $q=1, \dots, r$), et les B_i^j , ($i=1, \dots, r$; $j=1, \dots, r$) (ceux-ci apparaîtront à droite) considérés comme éléments.

On définit le «co-facteur» de «l'élément» en position (i, j) comme étant le déterminant de la matrice que l'on obtient en rayant la «ligne» i et la «colonne» j de la matrice en considération, multiplié par $(-1)^{i+j}$.

Désignons par Δ_{ij} , ($i, j=1, \dots, n$), les matrices obtenues en développant les «co-facteurs» des matrices A_{ij} .

EXEMPLE

$$A = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix}; \quad A_{ij} \in M_{s,s}(\mathbb{K})$$

et commutatives; B_i matrices de type (s, t) , ($i=1, 2, 3$); Supposons $r=2$.

$$\Delta := \text{dev} \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = A_{11} A_{22} - A_{21} A_{12} \in M_{s,s}(\mathbb{K});$$

$$\Delta_1 := \widetilde{\text{dev}} \begin{vmatrix} B_1 & A_{12} \\ B_2 & A_{22} \end{vmatrix} = A_{22} B_1 - A_{12} B_2 \in M_{s,t}(\mathbb{K});$$

$$\Delta_2 := \widetilde{\text{dev}} \begin{vmatrix} A_{11} & B_1 \\ A_{21} & B_2 \end{vmatrix} = A_{11} B_2 - A_{21} B_1 \in M_{s,t}(\mathbb{K});$$

$$\xi_5 := \text{Dev} \begin{vmatrix} A_{11} & A_{12} & B_1 & 1 \\ A_{21} & A_{22} & B_2 & 2 \\ \hline A_{31} & A_{32} & B_3 & 3 \end{vmatrix} = \text{dev} \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \cdot (-1)^{5+5} B_3 +$$

$$\begin{aligned}
 & + \operatorname{dev} \begin{vmatrix} A_{11} & A_{12} \\ A_{31} & A_{32} \end{vmatrix} \cdot (-1)^{2+5} B_2 + \\
 & + \operatorname{dev} \begin{vmatrix} A_{21} & A_{22} \\ A_{51} & A_{52} \end{vmatrix} \cdot (-1)^{1+5} B_1 = \\
 & = (A_{11} A_{22} - A_{21} A_{12}) B_3 - \\
 & - (A_{11} A_{32} - A_{31} A_{12}) B_2 + \\
 & + (A_{21} A_{32} - A_{31} A_{22}) B_1 \in \mathbf{M}_{s,t}(\mathbf{K}); \\
 \text{«co-facteur» de } A_{12} & : =
 \end{aligned}$$

$$= (-1)^{1+2} \det \begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix};$$

$$\Delta_{12} : = \operatorname{dev} (\text{«co-facteur» de } A_{12}) : =$$

$$= \operatorname{dev} (-1)^{1+2} \begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix} =$$

$$= (-1)^{1+2} \operatorname{dev} \begin{vmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{vmatrix};$$

II. Résultats intermédiaires (Connus)

Considérons le système

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

On peut énoncer [4, 1], si le rang de la matrice du système est $r = m = n$:

A. Règle de Cramer (MAC-LAURIN, BÉZOUT [1]).

Tout système de n équations à n inconnues, où le déterminant des coefficients est différent de zéro, admet un système de solutions uniques, données par les formules suivantes:

$$x_i = D^{-1} D_i, \quad (i = 1, 2, \dots, n).$$

Lorsque $r = m < n$, le système est possible indéterminé et les r inconnues princi-

pales sont fonctions linéaires des $n - r$ inconnues non-principales.

Lorsque $r < m$, le système est possible si et seulement si $C_j = 0$, ($j = r + 1, \dots, m$).

Et l'on peut énoncer [4, 7]:

B. Théorème de Rouché (CAPPELLI [3]).

Un système d'équations linéaires est possible si et seulement si il n'y a pas de déterminants caractéristiques ou tous s'annulent.

On a, alors, r étant le rang de la matrice du système:

$$x_i = D^{-1} D_i \quad (i = 1, \dots, r),$$

$n - r$ inconnues arbitraires et

$$C_j = 0, \quad (j = r + 1, \dots, m).$$

Il nous faut un théorème hybride d'abord (1) énoncé par INGRAHAM [6] et récemment généralisé par BRENNER [2]:

C. (Ingraham)

Si $\mathbf{A} = (A_{ij})$, ($i, j = 1, 2, \dots, n$), et A_{ij} sont des matrices carrées de type (s, s) permutable entre elles et si Δ est la matrice obtenue en développant le déterminant de \mathbf{A} en prenant les A_{ij} comme éléments, alors $\det \mathbf{A} = \det \Delta$.

REMARQUE. INGRAHAM a démontré ce théorème sous une double hypothèse: toutes les sous-matrices A_{ij} sont commutatives et le corps des coefficients est aussi commutatif. BRENNER l'a généralisé pour le cas où les matrices ne sont pas commutatives.

Maintenant, considérons l'équation $A x_1 = \lambda_1 x_1$, où $A \in \mathbf{M}_{n,n}(\mathbf{K})$, x_1 est un vecteur propre de A correspondant à la valeur propre simple λ_1 . λ_1 étant simple, la matrice

(1) OSTROWSKI [8] se réfère aussi à Schur.

$(A - \lambda_1 I)$ admet, au moins, un mineur d'ordre $n - 1$ différent de zéro. Sans perdre de la généralité on peut supposer que c'est le déterminant de la sous-matrice principale d'ordre $n - 1$ de $(A - \lambda_1 I)$.

De la théorie des équations algébriques, on sait que l'on peut avoir [11, p. 67]:

D.

$$x_1 = \begin{vmatrix} x_1^1 & c_{n1} \\ x_2^1 & c_{n2} \\ \vdots & \vdots \\ x_n^1 & c_{nn} \end{vmatrix}$$

où c_{ni} ; désigne le co-facteur de l'élément de $(A - \lambda_1 I)$ dans la position

$$(n, i), (i = 1, 2, \dots, n).$$

III. Résultats (formellement) semblables à A. et B.

Considérons le système d'équations matricielles linéaires

$$(2) \quad \sum_{k=1}^n A_{ik} X_k = B_i, (i = 1, 2, \dots, m),$$

où

$$A_{pq} \in M_{s,t}(\mathbf{K}), (p=1, 2, \dots, m; q=1, 2, \dots, n)$$

sont des éléments connus de l'anneau commutatif $M_{s,t}(\mathbf{K})$, (relativement à l'addition et à la multiplication), avec élément unité;

$$B_i \in M_{s,t}(\mathbf{K}), (i = 1, 2, \dots, m)$$

sont connus et

$$X_k \in M_{s,t}(\mathbf{K}), (k = 1, 2, \dots, n)$$

sont inconnus;

a) Soit d'abord $m = n = r$.

En pré-multipliant les deux membres de (2) par $\Delta_{ij}, (i, j = 1, \dots, n)$ et en sommant, vient

$$\sum_{i=1}^n \sum_{k=1}^n \Delta_{ij} A_{ik} X_k = \sum_{i=1}^n \Delta_{ij} B_i, (j=1, 2, \dots, n),$$

et, par définition de Δ , de Δ_{ij} et de Δ_j , ($j = 1, \dots, n$), on obtient [9]

$$\Delta X_j = \Delta_j, (j = 1, \dots, n).$$

D'où $X_j = \Delta^{-1} \Delta_j$, ($j = 1, 2, \dots, n$), si et seulement si Δ est non-singulière.

Toujours pour $m = n = r$, le système (2) admet une solution unique si et seulement si sa matrice $\begin{vmatrix} A_{11} \cdots A_{1n} \\ \dots \dots \dots \\ A_{n1} \cdots A_{nn} \end{vmatrix}$ est non-singulière.

Donc, vu le théorème de Ingraham, on peut énoncer :

A 1. Le système (2) admet une unique solution si et seulement si la matrice Δ est non-singulière, et la solution est donnée par

$$X_j = \Delta^{-1} \Delta_j, (j = 1, 2, \dots, n).$$

b) Considérons maintenant $r = m < n$

Le système (2) peut s'écrire

$$(3) \quad \sum_{k=1}^m A_{ik} X_k = B'_i, (i = 1, \dots, m)$$

avec

$$B'_i = B_i - \sum_{j=m+1}^n A_{ij} X_j, (i = 1, 2, \dots, m),$$

où les matrices X_j , ($j = m + 1, \dots, n$), sont considérées arbitraires [10].

La matrice de (2) est

$$\begin{vmatrix} A_{11} \cdots A_{1m} \cdots A_{1n} \\ \dots \dots \dots \dots \dots \dots \dots \\ A_{m1} \cdots A_{mm} \cdots A_{mn} \end{vmatrix},$$

celle de (3) est

$$A' = \begin{vmatrix} A_{11} & \cdots & A_{1m} \\ \dots & \dots & \dots \\ A_{m1} & \cdots & A_{mm} \end{vmatrix}.$$

Notons par Δ' et Δ'_i les correspondants de Δ et Δ_i (pour A' et B'_i). L'on peut dire:

A 2. Le système (3) admet la solution $X_j = (\Delta')^{-1} \Delta'_j$, ($j = 1, 2, \dots, m$), si et seulement si la matrice Δ' est non-singulière. La solution dépend évidemment des $n - m = n - r$ matrices arbitraires restantes.

c) Supposons finalement $m > n = r$. Considérons le système (2) et aussi le système

$$(4) \begin{cases} A_{11} X_1 + \cdots + A_{1n} X_n = B_1 \\ \dots \\ A_{n1} X_1 + \cdots + A_{nn} X_n = B_n \\ A_{n+1,1} X_1 + \cdots + A_{n+1,n} X_n + X'_{n+1} = B_{n+1} \\ A_{n+2,1} X_1 + \cdots + A_{n+2,n} X_n + X'_{n+2} = B_{n+2} \\ \dots \\ A_{m1} X_1 + \cdots + A_{mn} X_n + X'_m = B_m \end{cases}$$

La matrice du système (2) est

$$\begin{vmatrix} A_{11} & \cdots & A_{1n} & \cdots & A_{1m} \\ \dots & \dots & \dots & \dots & \dots \\ A_{m1} & \cdots & A_{mn} & \cdots & A_{mm} \end{vmatrix}$$

et celle du système (4) est

$$A'' = \begin{vmatrix} A_{11} & \cdots & A_{1n} & \vdots & 0 & 0 \cdots 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ A_{n1} & \cdots & A_{nn} & \vdots & 0 & 0 \cdots 0 \\ \hline A_{n+1,1} & \cdots & A_{n+1,n} & I & 0 \cdots 0 \\ A_{n+2,1} & \cdots & A_{n+2,n} & 0 & I \cdots 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ A_{m1} & \cdots & A_{m,n} & 0 & 0 \cdots I \end{vmatrix}$$

où I est l'identité de $M_{s,s}(K)$.

Le système (4) admet une unique solution

si et seulement si la matrice $\Delta'' := \text{dev } A''$ est non-singulière. La solution est donnée par les formules

$$(5) X'_j = (\Delta'')^{-1} \Delta''_j, (j = 1, 2, \dots, m),$$

où les matrices Δ''_j sont définies par rapport au système (4). De (5) on voit que $X'_j = 0$, ($j = n + 1, \dots, m$) si et seulement si $\Delta''_j = 0$, ($j = n + 1, \dots, m$). Mais, lorsque $X'_j = 0$, ($j = n + 1, \dots, m$), $X'_i = X_i$, ($i = 1, 2, \dots, n$). On remarquera que, dans le cas du système (4) [10]:

$$\Delta'' = \text{dev} \begin{vmatrix} A_{11} & \cdots & A_{1n} \\ \dots & \dots & \dots \\ A_{n1} & \cdots & A_{nn} \end{vmatrix};$$

$$\Delta''_j = \widetilde{\text{dev}} \begin{vmatrix} A_{11} & \cdots & A_{1,j-1} & \overline{B_1} & A_{1,j+1} & \cdots & A_{1n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ A_{n1} & \cdots & A_{n,j-1} & \overline{B_n} & A_{n,j+1} & \cdots & A_{nn} \end{vmatrix},$$

($j = 1, 2, \dots, n$), et

$$\Delta''_k = \text{Dev} \begin{vmatrix} A_{11} & \cdots & A_{1n} & B_1 \\ \dots & \dots & \dots & \dots \\ A_{n1} & \cdots & A_{nn} & B_n \\ \dots & \dots & \dots & \dots \\ A_{k1} & \cdots & A_{kn} & B_k \end{vmatrix} := \xi''_k, (k = n + 1, \dots, m).$$

Donc on peut énoncer:

B 1. Soit Δ'' une matrice non-singulière de type (s, s) . Le système (2) admet une unique solution donnée par

$$X_i = (\Delta'')^{-1} \Delta''_i, (i = 1, 2, \dots, n = r),$$

si et seulement si $\xi''_k = 0$, ($k = n + 1, \dots, m$);

c'est-à-dire,

Un système d'équations linéaires matricielles est possible si et seulement s'il n'y a pas de matrices « caractéristiques » ou toutes s'annulent.

[Matrices « caractéristiques », par ressemblance formelle avec déterminants caractéristiques].

On peut appeler r le « rang généralisé » de la matrice du système.

IV. Résultats (formellement) semblables à D.

On s'intéresse à la résolution de l'équation $\mathbf{A}X = \lambda X$ où

$$\begin{vmatrix} A_{11} & \dots & A_{1n} \\ \dots & \dots & \dots \\ A_{n1} & \dots & A_{nn} \end{vmatrix} = (A_{ij}) = \mathbf{A} \in M_{ns, ns}(\mathbf{K})$$

est une matrice carrée d'éléments commutatifs $A_{ij} \in M_{s, s}(\mathbf{K})$; dans un cas,

$$\begin{vmatrix} x_1 \\ \vdots \\ x_{ns} \end{vmatrix} = X = \begin{vmatrix} X_1 \\ \vdots \\ X_n \end{vmatrix},$$

où $X_i \in M_{s, 1}(\mathbf{K})$, ($i = 1, \dots, n$), sont des vecteurs inconnus non tous nuls; dans un autre cas,

$$X = \begin{vmatrix} X_1 \\ \vdots \\ X_n \end{vmatrix}, X \in M_{s, s}(\mathbf{K}), (i = 1, \dots, n),$$

sont aussi inconnus et non tous nuls.

L'équation matricielle $\mathbf{A}X = \lambda X$, λ étant un nombre réel ou complexe quelconque, peut s'écrire de la forme suivante:

$$(6) \begin{cases} (A_{11} - \lambda I)X_1 + A_{12}X_2 + \dots + A_{1n}X_n = 0 \\ A_{21}X_1 + (A_{22} - \lambda I)X_2 + \dots + A_{2n}X_n = 0 \\ \dots \\ A_{n1}X_1 + A_{n2}X_2 + \dots + (A_{nn} - \lambda I)X_n = 0 \end{cases}$$

(I matrice identité).

On pose

$$\Delta_n^\lambda := \text{dev} \begin{vmatrix} A_{11} & A_{12} & \dots & A_{1,n-1} & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2,n-1} & A_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ A_{n-1,1} & A_{n-1,2} & \dots & A_{n-1,n-1} & A_{n-1,n} \\ A_{n1} & A_{n2} & \dots & A_{n,n-1} & A_{nn} \end{vmatrix} := \text{dev } \mathbf{A}_n^\lambda;$$

$$\delta_j^\lambda := \widetilde{\text{dev}} \begin{vmatrix} A_{11} & \dots & A_{1,j-1} & \boxed{-A_{1n} \quad X_n} & \dots & A_{1,n-1} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ A_{n-1,1} & \dots & A_{n-1,j-1} & \boxed{-A_{n-1,n} \quad X_n} & \dots & A_{n-1,n-1} \end{vmatrix},$$

($j = 1, 2, \dots, n-1$);

$$\xi_n^\lambda := \text{Dev} \begin{vmatrix} A_{11} & \dots & A_{1,n-1} & \boxed{-A_{1n} \quad X_n} \\ \dots & \dots & \dots & \dots \\ A_{n-1,1} & \dots & A_{n-1,n-1} & \boxed{-A_{n-1,n} \quad X_n} \\ A_{n1} & \dots & A_{n,n-1} & \boxed{-A_{nn} \quad X_n} \end{vmatrix},$$

avec $\Lambda_{ii} = A_{ii} - \lambda I$, ($i = 1, 2, \dots, n$).

Par C_{ni} , ($i = 1, 2, \dots, n$) on désigne la matrice obtenue par développement des « co-facteurs » des « éléments » de \mathbf{A}_n^λ en position (n, i) , ($i = 1, 2, \dots, n$).

a) Cas où l'inconnue est un vecteur

On a à résoudre l'équation 7) $\mathbf{A}X = \lambda X$, où λ est une valeur propre de \mathbf{A} . λ étant une valeur propre de \mathbf{A} , le vecteur X est différent de zéro, par conséquent la matrice \mathbf{A}_n^λ est singulière. (Si et seulement si \mathbf{A}_n^λ était non-singulière, on aurait la solution unique $X = 0 \in \mathbf{K}^{ns}$).

Par C , Δ_n^λ est singulière. Sans perdre de la généralité on suppose que $\Delta_{n-1}^\lambda := \text{dev } \mathbf{A}_{n-1}^\lambda$, où \mathbf{A}_{n-1}^λ est la sous-matrice principale d'ordre $n-1$.

On vérifie aisément que

$$\delta_j^\lambda = C_{nj} X_n, (j = 1, 2, \dots, n-1)$$

et

$$\xi_n^\lambda = -\Delta_n^\lambda X_n.$$

Comme X_n est arbitraire (A. 2), $\xi_n^\lambda = 0$

(matrice) si et seulement si $\Delta_n^\lambda = 0$. Et, vu le résultat B 1., on peut énoncer :

D 1. La solution de l'équation 7) est donnée par $X_j = (\Delta_{n-1}^\lambda)^{-1} C_{nj} X_n$, ($j=1, 2, \dots, n-1$), si et seulement si la matrice Δ_{n-1}^λ est non-singulière et si et seulement $\Delta_n^\lambda = 0$.

$X_n \in M_{s,1}(K) \in K^s$ étant arbitraire, on peut poser $X_n = \Delta_{n-1}^\lambda K$, où $K \in K^s$ est un vecteur arbitraire différent de zéro, et l'on obtient

$$\begin{aligned} X_j &= (\Delta_{n-1}^\lambda)^{-1} (C_{nj} \Delta_{n-1}^\lambda) K = \\ &= ((\Delta_{n-1}^\lambda)^{-1} \Delta_{n-1}^\lambda) C_{nj} K = C_{nj} K, \\ &\quad (j = 1, \dots, n-1); \end{aligned}$$

Comme $\Delta_{n-1}^\lambda = C_{nn}$, vient $X_j = C_{nj} K$, ($j = 1, 2, \dots, n$).

Et l'on a :

D 2.

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} C_{n1} K \\ C_{n2} K \\ \vdots \\ C_{nn} K \end{pmatrix} \in K^{ns}.$$

b) Cas où l'inconnue est une matrice

Considérons la matrice A du début de ce paragraphe et soit λ un nombre quelconque; soit $Y^T = [Y_1 Y_2 \dots Y_n]$ une matrice non nulle où $Y_i \in M_{r,s}(K)$, ($i=1, 2, \dots, n$), sont inconnus. Soit l'équation

$$8) \quad A Y = \lambda Y.$$

Comme avant, on peut affirmer :

D 3. Si et seulement si la matrice Δ_{n-1}^λ est non-singulière, et si et seulement si $\Delta_n^\lambda = 0$, la solution de 8) est donnée par

$$Y_j = (\Delta_{n-1}^\lambda)^{-1} C_{nj} Y_n, (j = 1, 2, \dots, n-1)$$

où Y_n est une matrice arbitraire.

On peut poser $Y_n = \Delta_{n-1}^\lambda = C_{nn}$, donc $Y_j = C_{nj}$, ($j = 1, 2, \dots, n$) et, alors, on a :

D 4.

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} C_{n1} \\ C_{n2} \\ \vdots \\ C_{nn} \end{pmatrix}.$$

V. Exemple numérique.

Les deux systèmes suivants ont la même matrice et seules les colonnes des termes indépendants diffèrent.

$$\begin{cases} x + y + 2z + 2w = 1 \\ y + 2w = 2 \\ 3x + 3y + 4z + 4w = 3 \\ 3y + 4w = 4 \end{cases}, \quad \begin{cases} \alpha + \beta + 2\gamma + 2\delta = 3 \\ \beta + 2\delta = 1 \\ 3\alpha + 3\beta + 4\gamma + 4\delta = 2 \\ 3\beta + 4\delta = 4 \end{cases}.$$

On peut résoudre simultanément ceux deux systèmes, en considérant le système

$$\begin{cases} A_{11} X_1 + A_{12} X_2 = B_1 \\ A_{21} X_1 + A_{22} X_2 = B_2 \end{cases} \quad \text{où } A = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix}$$

dont les sous-matrices sont carrées et commutatives ;

$$X_1 = \begin{vmatrix} x & \alpha \\ y & \beta \end{vmatrix}, \quad X_2 = \begin{vmatrix} z & \gamma \\ w & \delta \end{vmatrix};$$

et

$$B_1 = \begin{vmatrix} 1 & 3 \\ 2 & 1 \end{vmatrix}, \quad B_2 = \begin{vmatrix} 3 & 2 \\ 4 & 4 \end{vmatrix}.$$

On aura

$$\Delta := \text{dev } \mathbf{A} = \begin{vmatrix} -2 & -4 \\ 0 & -2 \end{vmatrix} \Rightarrow$$

$$\Rightarrow \Delta^{-1} = \begin{vmatrix} -\frac{1}{2} & 1 \\ 0 & -\frac{1}{2} \end{vmatrix};$$

$$\Delta_1 := \widetilde{\text{dev}} \left| \begin{array}{cc|cc} 1 & 3 & 2 & 2 \\ 2 & 1 & B_1 & 0 & 2 \\ \hline 3 & 2 & 4 & 4 \\ 4 & 4 & B_2 & 0 & 4 \end{array} \right| =$$

$$= \begin{vmatrix} 4 & 4 \\ 0 & 4 \end{vmatrix} \cdot \begin{vmatrix} 1 & 3 \\ 2 & 1 \end{vmatrix} - \begin{vmatrix} 2 & 2 \\ 0 & 2 \end{vmatrix} \cdot \begin{vmatrix} 3 & 2 \\ 4 & 4 \end{vmatrix} =$$

$$= \begin{vmatrix} -2 & 4 \\ 0 & -4 \end{vmatrix};$$

$$\Delta_2 := \widetilde{\text{dev}} \left| \begin{array}{cc|cc} 1 & 1 & 1 & 3 \\ 0 & 1 & 2 & 1 \\ \hline 3 & 3 & 3 & 2 \\ 0 & 3 & 4 & 4 \end{array} \right| =$$

$$= \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} \cdot \begin{vmatrix} 3 & 2 \\ 4 & 4 \end{vmatrix} - \begin{vmatrix} 3 & 3 \\ 0 & 3 \end{vmatrix} \cdot \begin{vmatrix} 1 & 3 \\ 2 & 1 \end{vmatrix} =$$

$$= \begin{vmatrix} -2 & -6 \\ -2 & 1 \end{vmatrix};$$

$$X_1 = \Delta^{-1} \Delta_1 = \begin{vmatrix} 1 & -6 \\ 0 & 2 \end{vmatrix}, X_2 = \Delta^{-1} \Delta_2 =$$

$$= \begin{vmatrix} -1 & 4 \\ 1 & -\frac{1}{2} \end{vmatrix}.$$

Considérons, maintenant, simultanément, les deux systèmes

$$i) \begin{cases} x + y + 2z + 2w = 1 \\ y + 2w = 2 \\ 3x + 3y + 4z + 4w = 3 \\ 3y + 4w = 4 \\ 2x + 2y + 4z + 4w = 5 \\ 2y + 4w = 6 \end{cases}$$

$$ii) \begin{cases} \alpha + \beta + 2\gamma + 2\delta = 1 \\ \beta + 2\delta = 2 \\ 3\alpha + 3\beta + 4\gamma + 4\delta = 3 \\ 3\beta + 4\delta = 4 \\ 2\alpha + 2\beta + 4\gamma + 4\delta = 2 \\ 2\beta + 4\delta = 4 \end{cases}$$

On aura

$$\xi_5 := \text{Dev} \left| \begin{array}{cc|c} A_{11} & A_{12} & B_1 \\ A_{21} & A_{22} & B_2 \\ \hline A_{31} & A_{32} & B_3 \end{array} \right| :=$$

$$= \text{Dev} \begin{array}{c} \begin{array}{cc|cc} 1 & 2 & 3 & \\ \hline 1 & 1 & 2 & 2 \\ 0 & 1 & 0 & 2 \end{array} \begin{array}{cc|cc} 1 & 1 & 1 & 3 \\ \hline 3 & 3 & 3 & 2 \\ 0 & 3 & 4 & 4 \end{array} \end{array} \begin{array}{l} 1 \\ 2 \\ 3 \end{array} =$$

$$= \text{dev} \begin{vmatrix} 3 & 3 & 4 & 4 \\ 0 & 3 & 0 & 4 \\ \hline 2 & 2 & 4 & 4 \\ 0 & 2 & 0 & 4 \end{vmatrix} \cdot (-1)^{1+5} \begin{vmatrix} 1 & 1 \\ 2 & 2 \end{vmatrix} +$$

$$+ \text{dev} \begin{vmatrix} 1 & 1 & 2 & 2 \\ 0 & 1 & 0 & 2 \\ \hline 2 & 2 & 4 & 4 \\ 0 & 2 & 0 & 4 \end{vmatrix} \cdot (-1)^{2+5} \cdot \begin{vmatrix} 3 & 3 \\ 4 & 4 \end{vmatrix} +$$

$$+ \text{dev} \begin{vmatrix} 1 & 1 & 2 & 2 \\ 0 & 1 & 0 & 2 \\ \hline 3 & 3 & 4 & 4 \\ 0 & 3 & 0 & 4 \end{vmatrix} \cdot (-1)^{3+5} \cdot \begin{vmatrix} 5 & 2 \\ 6 & 4 \end{vmatrix} =$$

$$= \begin{vmatrix} -14 & 0 \\ -4 & 0 \end{vmatrix} \neq 0.$$

On peut remarquer, néanmoins, en regardant la 2.^e colonne de ξ_5 , que le système \tilde{u}) est possible.

VI. Remarque sur l'égalité des déterminants de certaines matrices.

On montre que

E. *Les matrices*

$$A = \begin{vmatrix} A_{11} & \cdots & A_{1n} \\ \cdot & \cdot & \cdot \\ A_{n1} & \cdots & A_{nn} \end{vmatrix} \quad \text{et} \quad B = \begin{vmatrix} A_{11}^T & \cdots & A_{1n}^T \\ \cdot & \cdot & \cdot \\ A_{n1}^T & \cdots & A_{nn}^T \end{vmatrix}$$

ont les mêmes déterminants. (A_{ij}^T est la transposée de A_{ij} . Les sous-matrices A_{ij} sont commutatives).

En effet, posant $\alpha := \text{dev } B$ et $\Delta := \text{dev } A$, vu le théorème de Ingraham, et vu que l'on vérifie aisément que $\alpha^T = \Delta$, on a

$$\det A = \det \Delta = \det \alpha^T = \det \alpha = \det B.$$

RÉFÉRENCES

- [1] BOYER, C. B., «*A history of mathematics*», J. Willey, N. York, 1968.
- [2] BRENNER, J. L., «*Applications of the Dieudonné determinant*», *Lin. Alg. Appl.* **1**, **4** (1968), 511-536.
- [3] CHERUBINO, S. «*Calcolo delle matrici*», Ed. Cremonese, Roma, 1957.
- [4] DIAS AGUDO, F. R., «*Introdução à Álgebra Linear e Geometria Analítica*», Vol. I, Lisboa, 1964.
- [5] HOUSEHOLDER, A. S., «*The approximative solution of matrix problems*», *J. A. C. M.* **5** (1958), 204-241.
- [6] INGRAHAM, M. H., «*A note on determinants*», *Bull. A. M. S.* **43** (1937); 579-580.
- [7] LARRIEU, J. «*Principes d'Algèbre Linéaire*», Dunod, Paris, 1965.
- [8] OSTROWSKI, A., «*On some metrical properties of operator matrices and matrices partitioned into blocks*», *J. Math. Anal. Appl.* **2** (1961), 161-209.
- [9] VITÓRIA, J., «*On certain systems of matricial linear equations*», *Rev. E. G. U. M., Ser. I-Ciências Mat.*, Vol. IV, L. Marques, 1967.
- [10] ———, «*On certain systems of linear matricial equations (II)*», *Revista de Ciências Matemáticas*, **1** (1969), 9-14; (Univ. L. Marques).
- [11] WILKINSON, J. H., «*The algebraic eigenvalue problem*», Oxford, 1967.

NOTA DE AULA

Sobre um teorema de Lucas relacionado com o pequeno teorema de Fermat

por José Morgado

Instituto de Matemática, Universidade Federal de Pernambuco, Brasil

1. O pequeno teorema de FERMAT estabelece que, para todo inteiro racional a não divisível pelo inteiro racional primo positivo p , é válida a congruência

$$a^{p-1} \equiv 1 \pmod{p}.$$

No entanto, como é bem sabido, do facto de a congruência

$$a^{n-1} \equiv 1 \pmod{n}$$

ser válida para algum inteiro racional a (necessariamente primo com n), não se pode concluir que o inteiro racional positivo n seja primo. Assim, como foi observado por F. SARRUS em 1819 [1], tem-se

$$2^{340} \equiv 1 \pmod{341},$$

apesar de $341 (= 11 \cdot 31)$ não ser primo.

Em 1876, ÉDOUARD LUCAS obteve o seguinte resultado [2]:

Se n é um inteiro racional positivo tal que, para algum inteiro racional a , se tem

$$a^{n-1} \equiv 1 \pmod{n}$$

e, para todo inteiro racional t satisfazendo à condição $0 < t < n - 1$, se tem

$$a^t \not\equiv 1 \pmod{n},$$

então n é primo.

Posteriormente, em 1891, E. LUCAS [3] melhorou este resultado, estabelecendo que:

Se a congruência

$$a^x \equiv 1 \pmod{n}$$

é válida para $x = n - 1$, mas não é válida se x é divisor próprio de $n - 1$, então n é primo.

Nesta nota, formulamos um análogo a este teorema de LUCAS, para anéis comutativos.

2. Seja A um anel comutativo e seja P um ideal próprio de A .

Recordemos que P se diz um ideal primo de A , se é satisfeita a seguinte condição:

Se $x, y \in A$ e $xy \in P$, então $x \in P$ ou $y \in P$.

Recordemos ainda que, se A é um anel comutativo com elemento um, então o ideal P de A é um ideal primo, se e só se o anel quociente A/P é um domínio de integridade.

Um análogo ao pequeno teorema de FERMAT para anéis comutativos pode enunciar-se assim:

TEOREMA 1. *Se A é um anel comutativo com elemento um e P é um ideal primo de A*

tal que o anel quociente A/P tem precisamente n elementos, então tem-se

$$(a + P)^{n-1} = 1 + P$$

ou, equivalentemente,

$$a^{n-1} \equiv 1 \pmod{P},$$

para todo elemento a de A que não pertença a P .

Dem.: Com efeito, da circunstância de A/P ser um domínio de integridade finito, conclui-se que A/P é um corpo e, por isso, os seus elementos não nulos constituem um grupo multiplicativo cuja ordem é $n - 1$.

Para todo elemento $a \in A$ tal que $a \notin P$ tem-se, por consequência,

$$a^{n-1} + P = (a + P)^{n-1} = 1 + P,$$

como se pretendia.

O pequeno teorema de FERMAT, na sua forma original, é justamente um caso particular deste teorema, a saber, o caso em que o anel A é o anel Z dos inteiros racionais. Então um ideal primo P de A é o ideal principal gerado por um inteiro racional positivo primo p e o anel quociente A/P tem exactamente p elementos.

O teorema anterior mostra que o pequeno teorema de FERMAT é válido no anel dos inteiros de GAUSS $G = \{a + ib : a, b \in Z\}$ ([4], pp. 19-21) (em anéis de inteiros algébricos ver [4], pp. 109-110).

3. Vamos agora estabelecer, para anéis comutativos, um teorema análogo ao teorema de LUCAS, tal como foi formulado em [3].

TEOREMA 2. *Seja A um anel comutativo com elemento um e seja P um ideal de A tal*

que o anel quociente A/P tem n elementos. Se, para algum $a \in A$, se tem

$$(a + P)^{n-1} = 1 + P$$

e se, para todo divisor próprio t de $n - 1$, se tem

$$(a + P)^t \neq 1 + P,$$

então P é um ideal primo de A .

Dem.: Na verdade, consideremos os seguintes elementos de A/P :

$$(1) \quad P, a + P, a^2 + P = (a + P)^2, \dots, a^{n-1} + P = (a + P)^{n-1}.$$

É fácil ver que, se $i, j \in \{1, 2, \dots, n-1\}$, então

$$(2) \quad i \neq j \text{ implica } a^i + P \neq a^j + P.$$

Com efeito, suponhamos que

$$(3) \quad a^i + P = a^j + P$$

tendo-se, por exemplo, $i > j$.

Ora, fazendo $n - 1 - i = k$, de (3) resultaria

$$a^{n-1} + P = a^{i+k} + P = a^{j+k} + P$$

e, como, por hipótese,

$$a^{n-1} + P = 1 + P,$$

ter-se-ia

$$a^{j+k} \in 1 + P$$

com $0 < j + k < n - 1$.

Designemos por d o menor inteiro racional positivo tal que

$$a^d \in 1 + P$$

e vejamos que, então, d é divisor de $n - 1$.

De facto, pelo algoritmo de divisão,

$$n - 1 = dm + r \quad \text{com} \quad 0 \leq r < d$$

e, por consequência,

$$\begin{aligned} 1 + P &= a^{n-1} + P = (a + P)^{n-1} = (a + P)^{d \cdot m + r} = \\ &= (a^d + P)^m \cdot (a^r + P) = (1 + P)^m \cdot \\ &\cdot (a^r + P) = (1 + P) \cdot (a^r + P) = a^r + P, \end{aligned}$$

isto é,

$$a^r \in 1 + P,$$

donde, pela definição de d , resulta que $r = 0$, quer dizer, d é divisor de $n - 1$.

Mas, como

$$d \leq j + k < n - 1,$$

não pode ter-se, por força da hipótese, $(a + P)^d = 1 + P$.

Esta contradição mostra a validade da implicação (2).

Assim, em (1) estão precisamente os n elementos que constituem o anel A/P . Então

o conjunto dos elementos de A/P que são diferentes de P constitui um grupo abeliano (cíclico), com respeito à multiplicação do anel A/P .

Isto significa que o anel A/P é um corpo, logo, um domínio de integridade e, portanto, P é um ideal primo, como queríamos mostrar.

BIBLIOGRAFIA

- [1] O. ORR, *Number Theory and its History*, Mc Graw-Hill Book Company, Inc., New York, 1948.
- [2] L. E. DICKSON, *History of the Theory of Numbers*, vol. 1, Chelsea Publ. Company, New York, 1952.
- [3] E. LUCAS, *Théorie des Nombres*, tome premier, nouveau tirage, Librairie Scientifique et Technique Albert Blanchard, Paris, 1961.
- [4] H. POLLARD, *The Theory of Algebraic Numbers*, The Carus Mathematical Monographs, No. 9, John Wiley and Sons, Inc., third impression, 1965.

Geometrização da Lógica de Proposições (a n dimensões)

por António José Mendes Silva
Universidade de Coimbra

I — Introdução e proposições básicas da teoria.

Traduzir a rigorosa linguagem da análise lógica, para a sugestiva linguagem dos traçados geométricos, foi a ideia central, que presidiu à elaboração deste método, com vista à criação de sistemas de cálculo mais rápidos e intuitivos.

As proposições básicas desta teoria, são, além de proposições de lógica pura, as seguintes condições definidoras do referencial lógico:

1 — «0» é a origem do referencial e a intersecção dos semi-eixos «fechados» — Postulado.

2 — Cada eixo é constituído por um *semi-eixo verdadeiro* e por um *semi-eixo falso* — Postulado.

3 — Os pontos dos eixos distintos da origem, representam proposições — Postulado.

A origem (ponto comum aos semi-eixos verdadeiros e falsos) não pode representar proposições-Teorema, pois se assim não fosse, uma proposição poderia ser *simultaneamente* verdadeira e falsa, o que peca contra o princípio da não contradição *aqui admitido*.

4 — Os pontos P dos quadrantes representam pares ordenados, cujos elementos são as suas coordenadas — Postulado.

Deste postulado e do teorema anterior deriva que os quadrantes lógicos são conjuntos disjuntos, pois não contêm os eixos-Corolário, porque se um ponto dos quadrantes, pertencesse a um eixo, pelo menos uma das suas coordenadas (que são proposições) estaria situada na origem, o que vai contra o teorema anterior, pelo qual a origem não pode representar proposições.

5 — O universo operacional duma operação entre proposições $x_1 \theta y_1$, é a reunião dos pontos P_1, P_2, P_3 e P_4 , que representam os pares ordenados $(x_1, y_1), (\sim x_1, y_1), (\sim x_1, \sim y_1)$ e $(x_1, \sim y_1)$, respectivamente — Postulado.

6 — O universo operacional duma operação entre expressões proposicionais $x \theta y$, é a reunião dos quatro quadrantes, q_1, q_2, q_3 e q_4 , ou seja: o conjunto dos pontos que representam respectivamente os pares ordenados $(x, y), (\sim x, y), (\sim x, \sim y)$ e $(x, \sim y)$ — Postulado.

Destas duas últimas condições se infere, que as deduções, silogismos e problemas, quer relativos a expressões proporcionais, quer relativos a proposições, se consideram referidos aos respectivos universos operacionais.

7 — *Domínio operacional*⁽¹⁾ é o conjunto de pontos do universo operacional em que

(1) Para simplificar, adiante diremos apenas *domínio*, em vez de domínio operacional.

a operação considerada é verdadeira — Postulado.

8 — Duas proposições são *contrárias*, isto é, uma é a negação da outra, quando estão situadas no mesmo eixo e ocupam posições simétricas relativamente à origem do referencial — Postulado.

A figura que a seguir apresentamos, ajuda a visualizar as condições do referencial lógico:

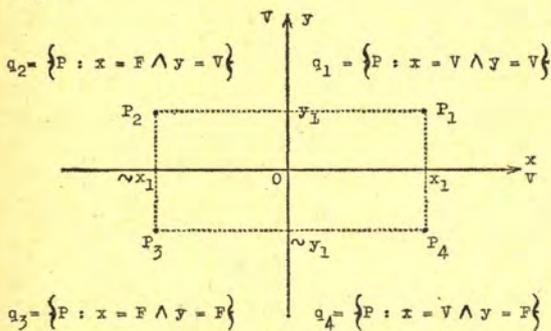


Fig. 1

II — Aplicação ao cálculo. Generalização.

1 — Com base na axiomática exposta, é fácil construir uma tabela com os *domínios operacionais* das principais operações lógicas. A tabela que a seguir apresentamos, foi feita para expressões proposicionais (para proposições a tabela é análoga). Na coluna dos «domínios operacionais» estão justificações entre parênteses.

OPERAÇÃO	DOMÍNIO OPERACIONAL
$x \wedge y$	q_1 (ambos verdadeiros)
$x \vee y$	$q_2 \cup q_4$ (um verdadeiro e outro falso)
$x \vee y$	$q_1 \cup q_2 \cup q_4$ (um ou dois verdadeiros)
$x \Rightarrow y$	$q_1 \cup q_2 \cup q_3$ ($x \Rightarrow y = F$ sse $x = V \wedge y = F$)
$x \Leftrightarrow y$	$q_1 \cup q_3$ (ambos verdadeiros ou ambos falsos)
$\sim(x \oplus y)$	domínio complementar do domínio de $x \oplus y$

É fácil provar, utilizando a *teoria das estruturas algébricas*, que:

$$(A, \vee) \simeq (D, \cup), \quad (A, \wedge) \simeq (D, \cap),$$

$$(A, \sim) \simeq (D, C), \quad (A, \Rightarrow) \simeq (D, \subset),$$

$$(A \Leftrightarrow) \simeq (D, =), \text{ etc.}$$

— *Sendo*: estes pares ordenados *grupoides*, A o conjunto das operações entre proposições, D o conjunto dos domínios operacionais das referidas operações e \simeq a relação de *isomorfia* definida pela aplicação que a uma operação entre proposições, faz corresponder o respectivo domínio operacional (1).

Ora o *princípio de isomorfia*, diz que se (E, θ) e (B, φ) são grupoides isomorfos, todas as propriedades lógicas de θ são verificadas por φ e vice versa; então, pela aplicação deste princípio ao caso presente, provamos que para efeitos de cálculo, os domínios operacionais podem substituir as correspondentes operações.

Obtém-se assim um novo sistema de cálculo, mais intuitivo e mais rápido que o cálculo clássico.

2 — Vejamos alguns exemplos:

a) Seja a demonstração de que:

$$x \Rightarrow y = \sim x \vee y.$$

De acordo com a axiomática (e as tabelas que dela derivam) e tendo em conta os isomorfismos apresentados, tem-se (ver fig. 2):

O domínio de $x \Rightarrow y$ é $q_1 \cup q_2 \cup q_3$.

O domínio de $\sim x$ ($x = F$) é $q_2 \cup q_3$.

O domínio de y ($y = V$) é $q_2 \cup q_1$.

Como o domínio de $\sim x \vee y$ (isomorfismo) é a *reunião* dos domínios de $\sim x$ e de y ,

(1) O caso das expressões proposicionais é análogo.

tem-se que este domínio é: $q_1 \cup q_2 \cup q_3$, que é também o domínio de $x \Rightarrow y$.

Logo (isomorfismo):

$$x \Rightarrow y = \sim x \vee y$$

c. q. d.

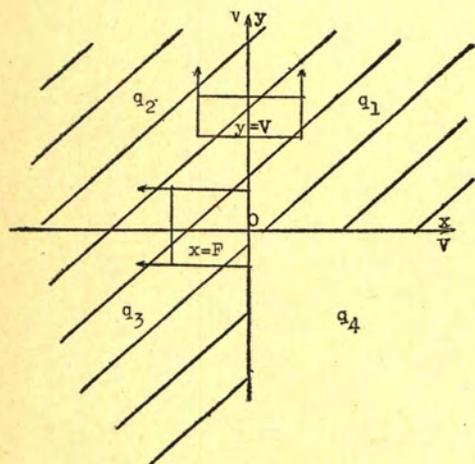


Fig. 2

b) Demonstração da 2.^a Lei de De Morgan

O domínio de $x_1 \vee y_1$ é $P_1 \cup P_2 \cup P_4$.

Logo:

O domínio de $\sim(x_1 \vee y_1)$ é P_3 .

Mas o domínio de $(x_1 = F \wedge y_1 = F) = \sim x_1 \wedge \sim y_1$ é P_3 .

Donde:

$$\sim(x_1 \vee y_1) = \sim x_1 \wedge \sim y_1.$$

c. q. d.

Anàlogamente se deduzia a 1.^a Lei de DE MORGAN.

c) Seja a demonstração de que

$$\sim(x_1 \dot{\vee} y_1) = x_1 \Leftrightarrow y_1.$$

O domínio de $\sim(x_1 \dot{\vee} y_1)$ é $P_1 \cup P_3$ visto que o de $x_1 \dot{\vee} y_1$ é $P_2 \cup P_4$. Mas como de $x_1 \Leftrightarrow y_1$ é $P_1 \cup P_3$:

$$\sim(x_1 \dot{\vee} y_1) = x_1 \Leftrightarrow y_1$$

c. q. d.

d) Seja o problema

– Simplificar a expressão:

$$(x \vee y) \wedge (x \Leftrightarrow y).$$

Traduzindo pelos isomorfismos apresentados:

$$(q_1 \cup q_2 \cup q_4) \cap (q_1 \cup q_3) = q_1,$$

atendendo a que segundo a axiomática, os quadrantes são conjuntos disjuntos; mas q_1 é o domínio de $x \wedge y$.

e – 1) Vamos demonstrar que o seguinte silogismo está correcto:

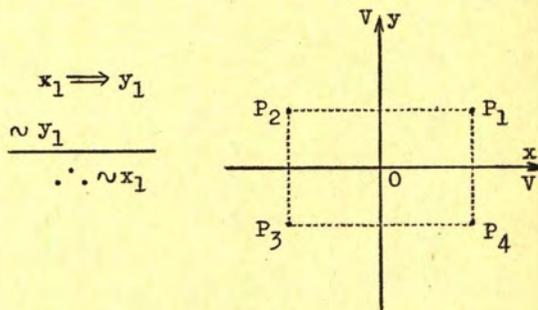


Fig. 3

O domínio de $x_1 \Rightarrow y_1$ é $P_1 \cup P_2 \cup P_3$. Como (2.^a premissa) $\sim y_1 = V$, tem-se $y_1 = F$. Ora, o único ponto de $P_1 \cup P_2 \cup P_3$ em que $y = F$ é P_3 e como em P_3 se tem $\sim x_1$, a conclusão é $\sim x_1$.

c. q. d.

e – 2) Pelo método clássico:

$$(x_1 \Rightarrow y_1) = (y_1 \vee \sim x_1)$$

e como:

$$(x_1 \vee y_1) = \sim x_1 \Rightarrow y_1$$

têm-se que:

$$(y_1 \vee \sim x_1) = (\sim y_1 \Rightarrow \sim x_1)$$

donde pela propriedade transitiva da igualdade:

$$(x_1 \Rightarrow y_1) = (\sim y_1 \Rightarrow \sim x_1).$$

Portanto a 1.^a premissa do silogismo é igual a:

$$\sim y_1 \Rightarrow \sim x_1.$$

Ora tendo-se $\sim y_1$ (segunda premissa), por definição de implicação, forçosamente se terá $\sim x_1$.

Da comparação entre os dois métodos se infere que este último é menos intuitivo.

3 — Trabalhámos em L^2 , sendo

$$L = \{V, F\}.$$

É fácil generalizar a teoria para L^n , ou seja para n dimensões. Sendo então os quadrantes dados por 2S_n (sequência de dois objectos, os valores lógicos V e F , agrupados n a n), pois neste caso geral temos, não os dois elementos dum par ordenado, mas os n elementos duma sequência de n elementos.

Em certos casos é possível (quando as operações são associativas) decompor um problema de n dimensões em vários problemas de 2, ou de 3, ou de m (com $m < n$) dimensões.

Para três dimensões, é conveniente a representação gráfica em perspectiva:

$${}^2S_3 = 2^3 = 8$$

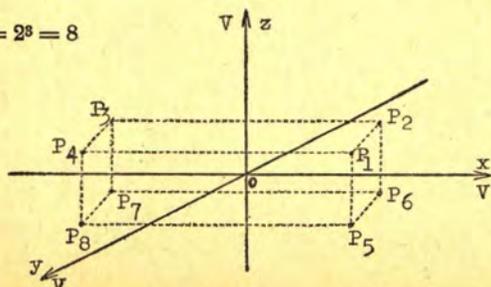


Fig 4

Mas claro que o estudo segundo esta teoria, utilizando Cálculo Combinatório, se pode fazer independentemente de qualquer representação gráfica, sendo então o número e a natureza dos quadrantes dados por 2S_n , no caso geral de n dimensões.

4 — Exemplo para três dimensões (fig. 4):

— Seja « $D(j)$ » o domínio operacional da operação « j » e U o seu universo operacional.

Temos:

$$D(z_1) = P_1 \cup P_2 \cup P_3 \cup P_4$$

e:

$$D(x_1 \vee y_1) = U - (P_5 \cup P_7) = C_U(P_5 \cup P_7)$$

donde (vide isomorfismo):

$$D(z_1 \wedge (x_1 \vee y_1)) = P_1 \cup P_2 \cup P_4$$

por outro lado:

$$D(z_1 \wedge x_1) = P_1 \cup P_2$$

e:

$$D(z_1 \wedge y_1) = P_1 \cup P_4$$

donde (vide isomorfismo):

$$D((z_1 \wedge x_1) \vee (z_1 \wedge y_1)) = P_1 \cup P_2 \cup P_4$$

logo:

$$z_1 \wedge (x_1 \vee y_1) = (z_1 \wedge x_1) \vee (z_1 \wedge y_1).$$

Deduzimos assim, dada a arbitrariedade de

$$x_1, y_1 \text{ e } z_1,$$

a propriedade distributiva da conjunção em relação à disjunção inclusiva.

Algumas propriedades da distribuição binomial negativa

por António Dorival Campos (*) e Euclides Custódio de Lima Filho

Departamento de Matemática Aplicada à Biologia
Faculdade de Medicina de Ribeirão Preto Universidade de São Paulo — Brasil

1. Consideremos uma população infinita π com uma proporção P desconhecida de elementos com uma característica específica A .

Amostremos π através de uma sequência de repetições independentes, com probabilidade constante, de sucessos e denotemos por X o número total de fracassos nesta sequência antes do r -ésimo sucesso, isto é, $X + r$ é o número de ensaios necessários para produzir na amostra $r > 1$ elementos com a característica A .

Temos:

$$(1.1) \quad g(x; r, P) = \binom{x+r-1}{x} P^r (1-P)^x \quad x = 0, 1, 2, \dots$$

Usando-se $\hat{P} = \frac{r-1}{X+r-1}$ como estimador de P , o qual foi proposto por HALDANE [2], demonstraremos que:

- i) X é uma estatística suficiente para P ;
- ii) $\{g(x; r, P), 0 < P < 1\}$ é completa;
- iii) $E[\hat{P}] = P$.

DEMONSTRAÇÃO:

i) Imediata, pois, $g(x; r, P)$ satisfaz o critério de fatoração.

(*) Bolsista do I. A. C. no Centro de Matemáticas Aplicadas, Lisboa.

ii) Seja $u(x)$ uma função contínua, independente de P , tal que:

$$(1.2) \quad E[u(X)] = 0$$

para todo P , $0 < P < 1$.

De (1.2), segue que:

$$(1.3) \quad E[u(X)] = \sum_{x=0}^{\infty} u(x) \cdot \binom{x+r-1}{x} P^r (1-P)^x = 0$$

para todo P , $0 < P < 1$.

Desenvolvendo-se (1.3) e tomando-se $1 - P = Q$, temos:

$$E[u(X)] = (1-Q)^r \left[u(0) + r u(1) Q + \frac{r(r+1)}{2} u(2) Q^2 + \dots \right] = 0.$$

Sendo $1 - Q \neq 0$ para todo $0 < Q < 1$, segue que a série em Q entre colchetes converge a 0 para todo $0 < Q < 1$, o que implica $u(x) = 0$ para todo $x = 0, 1, \dots$.

iii)

$$E[\hat{P}] = \sum_{x=0}^{\infty} \frac{r-1}{x+r-1} \binom{x+r-1}{x} P^r (1-P)^x = P \sum_{x=0}^{\infty} \binom{x+r-2}{x} P^{r-1} (1-P)^x.$$

Donde:

$$E[\hat{P}] = P.$$

Satisfeitas as três condições acima, concluímos que o estimador \hat{P} assim definido, tem variância mínima entre todos os estimadores não viciados de P .

2. Determinação da variância de \hat{P} .

$$(2.1) \quad E[\hat{P}^2] = \\ = \sum_{x=0}^{\infty} \left(\frac{r-1}{x+r-1} \right)^2 \binom{x+r-1}{x} P^r (1-P)^x = \\ = (r-1)^2 P^r \sum_{x=0}^{\infty} \frac{1}{(x+r-1)^2} \binom{x+r-1}{x} (1-P)^x$$

Usando-se o facto que se a série

$$\psi(x) = \sum_{k=0}^{\infty} b_k x^k$$

é diferenciada m vezes dentro do seu intervalo de convergência, então,

$$\frac{\psi^{(m)}(x)}{m!} = \sum_{k=0}^{\infty} b_{k+m} \binom{k+m}{k} x^k,$$

de (2.1), temos:

$$(2.2) \quad E[\hat{P}^2] = (r-1)^2 P^r \frac{1}{(r-1)!} \cdot$$

$$\frac{d^{r-1}}{d Q^{r-1}} \left[\sum_{x=1}^{\infty} \frac{1}{x^2} Q^x \right].$$

Por outro lado, temos:

$$\frac{d}{d Q} \left[\sum_{x=1}^{\infty} \frac{1}{x^2} Q^x \right] = \sum_{x=1}^{\infty} \frac{1}{x} Q^{x-1} = \\ = \frac{1}{Q} [-\log(1-Q)].$$

Donde:

$$(2.3) \quad \frac{d^{r-1}}{d Q^{r-1}} \left[\sum_{x=1}^{\infty} \frac{1}{x^2} Q^x \right] = \\ = - \frac{d^{r-2}}{d Q^{r-2}} \left[\frac{\log(1-Q)}{Q} \right].$$

Substituindo-se (2.3) em (2.2), segue que:

$$(2.4) \quad E[\hat{P}^2] = - \frac{(r-1) P^r}{(r-2)!} \cdot \\ \cdot \frac{d^{r-2}}{d Q^{r-2}} \left[\frac{\log(1-Q)}{Q} \right].$$

Sabemos que:

$$\frac{d^{r-2}}{d Q^{r-2}} \left[\frac{\log(1-Q)}{Q} \right] =$$

$$\sum_{i=0}^{r-2} \binom{r-2}{i} \left[\frac{d^i}{d Q^i} \log(1-Q) \right] \left[\frac{d^{r-2-i}}{d Q^{r-2-i}} \left(\frac{1}{Q} \right) \right]$$

Donde:

$$(2.5) \quad \frac{d^{r-2}}{d Q^{r-2}} \left[\frac{\log(1-Q)}{Q} \right] = \\ = \frac{(-1)^{r-2} (r-2)! \log(1-Q)}{Q^{r-1}} + \\ + (r-2)! \sum_{i=1}^{r-2} \frac{(-1)^{r-1-i}}{i(1-Q)^i Q^{r-1-i}}.$$

Substituindo-se (2.5) em (2.4), obtemos:

$$E[\hat{P}^2] = \frac{(-1)^{r-1} (r-1) P^r \log(1-Q)}{Q^{r-1}} + \\ + (r-1) P^r \sum_{i=1}^{r-2} \frac{(-1)^{r-i}}{i(1-Q)^i Q^{r-1-i}} = \\ = \frac{(r-1) P^r}{Q^{r-1}} \left[(-1)^{r-1} \log(1-Q) + \right. \\ \left. + \sum_{i=1}^{r-2} \frac{(-1)^{r-i}}{i} \cdot \left(\frac{Q}{1-Q} \right)^i \right].$$

Portanto,

$$\text{Var}[\hat{P}] = \frac{P^r}{(1-P)^{r-1}} (r-1) \left[(-1)^{r-1} \log P + \sum_{j=5}^r \frac{(-1)^{j-1}}{r-j+1} \cdot \left(\frac{1-P}{P} \right)^{r-j+1} \right] - P^2.$$

Outras expressões obtidas através de séries infinitas da variância de \hat{P} são mencionadas no trabalho já citado de HALDANE [2] e também em KENDALL and STUART [3].

3. Determinação de um estimador não viciado e de variância mínima para a razão $R = \frac{1-P}{P}$.

A partir de (1.1) o estimador de máxima verossimilhança para P é expresso por:

$$(3.1) \quad \hat{P} = \frac{r}{r+X}.$$

De acordo com a propriedade de invariância dos estimadores de máxima verossimilhança, substituindo-se (3.1) na expressão de R , temos:

$$(3.2) \quad \hat{R} = \frac{X}{r}.$$

Dado que X distribui-se segundo uma binomial negativa, temos que as condições i), ii) de 1) estão satisfeitas.

Por outro lado, temos:

$$(3.3) \quad E[\hat{R}] = \frac{1}{r} E[X] = \frac{1}{r} \cdot \frac{r(1-P)}{P}.$$

Donde:

$$(3.4) \quad E[\hat{R}] = \frac{1-P}{P} = R.$$

Das expressões (3.2) e (3.4), concluímos que \hat{R} também satisfaz as condições i), ii) e iii), respectivamente.

Portanto, entre todos os estimadores não viciados de R , o definido por (3.2) é de variância mínima, cuja expressão é dada por:

$$\text{Var}[\hat{R}] = \frac{1}{r^2} \text{Var}[X] = \frac{1}{r^2} \cdot \frac{r(1-P)}{P^2}.$$

Donde:

$$(3.5) \quad \text{Var}[\hat{R}] = \frac{1}{r} \cdot \frac{(1-P)}{P^2} = \frac{(1+R)R}{r}.$$

Devemos acrescentar ainda que estes resultados podem ser obtidos trabalhando-se com o estimador de máxima verossimilhança de $\frac{1}{P}$ e decompondo-se $R = \frac{1}{P} - 1$.

Podemos observar também que esses resultados não são válidos quando consideramos amostras de tamanho fixo.

BIBLIOGRAFIA

- [1] W. FELLER (1957): *An Introduction to Probability Theory and Its Applications*. Vol. 1-2nd Edition John Wiley and Sons, New York.
- [2] J. B. S. HALDANE (1945): *On a method of estimating frequencies*. *Biometrika*, Vol. 33, 222-225.
- [3] M. G. KENDALL and A. STUART (1961): *The Advanced Theory of Statistics*. Vol. 2-3rd Edition, Charles Griffin and Company Limited, London.

MATEMÁTICAS SUPERIORES

PONTOS DE EXAME DE FREQUÊNCIA E FINAIS

MATEMÁTICAS GERAIS

I. S. C. E. F. — MATEMÁTICAS GERAIS — 1.ª cadeira —
Exame final — Ano lectivo 1969-70 — Ponto n.º 1
— 4-6-1970.

5773 — 1) Prove que

$$f^{-1}(A) \cup f^{-1}(B) = f^{-1}(A \cup B).$$

R.: Provemos em primeiro lugar que

$$f^{-1}(A) \cup f^{-1}(B) \subseteq f^{-1}(A \cup B).$$

$$x \in f^{-1}(A) \cup f^{-1}(B) \Rightarrow x \in f^{-1}(A) \vee x \in f^{-1}(B) \Rightarrow$$

$$\Rightarrow f(x) \in A \vee f(x) \in B \Rightarrow f(x) \in A \cup B \Rightarrow$$

$$\Rightarrow x \in f^{-1}(A \cup B).$$

Mostremos agora que

$$f^{-1}(A \cup B) \subseteq f^{-1}(A) \cup f^{-1}(B).$$

$$x \in f^{-1}(A \cup B) \Rightarrow f(x) \in A \cup B \Rightarrow f(x) \in A \vee f(x) \in B \Rightarrow$$

$$\Rightarrow x \in f^{-1}(A) \vee x \in f^{-1}(B) \Rightarrow x \in f^{-1}(A) \cup f^{-1}(B).$$

2) Sendo X subconjunto do espaço métrico A , prove que o seu fecho \bar{X} é a intersecção de todos os conjuntos fechados que contêm X .

Ache o supremo, o ínfimo e o fecho do conjunto linear

$$X =]1, 3] \cup \left\{ x \in \mathbb{R} : x = (-1)^n \frac{n}{3n+2} (n=1, 2, \dots) \right\}.$$

O conjunto X é fechado? É aberto? Porquê?

R.: Seja K a intersecção de todos os conjuntos fechados que contêm X . Então, como $K \supseteq X$, tem-se $\bar{K} \supseteq \bar{X}$. K é fechado e portanto $K = \bar{K}$. Assim $K \supseteq \bar{X}$. Por outro lado, \bar{X} é fechado, $\bar{X} \supseteq X$ e assim \bar{X} é um dos conjuntos cuja intersecção é K . Portanto $\bar{X} \supseteq K$. Assim $K = \bar{X}$.

Para o conjunto linear X apresentado no problema

$$\text{é } \sup X = 3, \text{ inf } X = -1/5, \bar{X} = X \cup \left\{ -\frac{1}{3}, \frac{1}{1} \right\}.$$

O conjunto X não é fechado nem aberto.

3) Considere os intervalos fechados $I_n = [l_n, L_n]$, suponha que $I_n \supseteq I_{n+1}$ ($n=1, 2, \dots$) e que $L_n - l_n \rightarrow 0$. Prove que existe um e um só ponto comum a todos os intervalos I_n .

$$\text{Calcule } \lim_{n \rightarrow \infty} \frac{\log \left(1 + \frac{\log n}{n^2} \right) - \frac{1}{n}}{\sqrt[n]{a} - 1}.$$

R.: Tem-se

$$l_1 \leq l_2 \leq \dots \leq l_n \leq \dots$$

$$L_1 \geq L_2 \geq \dots \geq L_n \geq \dots$$

e, com $l_1 \leq l_n \leq L_n \leq L_1$, as sucessões l_n e L_n têm limites finitos e, dado que $L_n - l_n \rightarrow 0$, tem-se $\lim l_n = \lim L_n = \xi$.

É claro que ξ é o único ponto comum a todos os intervalos I_n . Se houvesse outro η , teríamos $L_n - l_n \geq |\xi - \eta| \forall n \in \mathbb{N}$ e então não era verdade que $L_n - l_n \rightarrow 0$.

$$\lim \frac{\log \left(1 + \frac{\log n}{n^2} \right) - \frac{1}{n}}{\sqrt[n]{a} - 1} = \lim \frac{\frac{\log n}{n^2} - \frac{1}{n}}{e^{\frac{1}{n} \log a} - 1} =$$

$$= \lim \frac{\frac{\log n}{n^2} - \frac{1}{n}}{\frac{1}{\xi} \log a} = \lim \frac{\frac{\log n}{n} - 1}{\xi \log a} = -1 / \log a.$$

4) Considere a série $\sum a_n$ ($a_n \geq 0$). Mostre que, se o conjunto dos termos da sucessão $\sqrt[n]{a_n}$ possui um ponto de acumulação maior do que 1, a série é divergente.

$$\text{Estude a natureza da série } \sum \frac{1}{[4 + (-1)^n]^n}.$$

R.: Se $\{ \sqrt[n]{a_n} \}$ tem um ponto de acumulação maior do que 1, então $\lim \sqrt[n]{a_n} > 1$ e $\sum a_n$ diverge.

Como

$$\lim^n \sqrt[n]{\frac{1}{[4 + (-1)^n]^n}} = \lim \frac{1}{4 + (-1)^n} = \frac{1}{3} < 1,$$

a série dada converge.

5) Sendo f função contínua em $] -\infty, +\infty [$, (inclusivé nos pontos impróprios), mostre que

$$\forall x \in \mathbb{R}, \exists k \in \mathbb{R} : f(x) + k > 0.$$

Na mesma hipótese e supondo $f(-\infty) \cdot f(+\infty) < 0$, indique, justificando, o máximo da função definida

$$\text{por } g(x) = \frac{1}{a^2 + [f(x)]^2}.$$

R.: Como f é contínua em $] -\infty, +\infty [$, tem-se $\forall x \in \mathbb{R} \mid f(x) \mid < k$ (ou $\forall x \in \mathbb{R} -k < f(x) < k$) donde resulta $\forall x \in \mathbb{R}, \exists k \in \mathbb{R} f(x) + k > 0$.

Como f passa por todos os valores desde $f(-\infty)$ a $f(+\infty)$, toma o valor 0. Então, $\max g(x) = 1/a^2$.

6) Deduza a fórmula que dá a derivada da função definida por $f(x) = \log_a x$ com $a \neq e$.

$$R: y = \log_a x \Leftrightarrow x = a^y$$

$$\frac{dy}{dx} = 1 / \frac{dx}{dy} = \frac{1}{a^y \log a} = \frac{1}{x \log a}.$$

I. S. G. E. F. — MATEMÁTICAS GERAIS — 1.ª cadeira — Exame final — Ano lectivo 1969-70 — Ponto n.º 2 — 24-6-1970.

5774 — 1) Sendo a e b números reais, prove que

- i) $a \neq 0 \wedge b \neq 0 \Rightarrow (ab)^{-1} = a^{-1} b^{-1}$.
 ii) $|a| \geq a \quad |a| \geq -a$.

$$R: i) (ab)(a^{-1}b^{-1}) = ab(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = aa^{-1} = 1 \\ (a^{-1}b^{-1})(ab) = b^{-1}(a^{-1}a)b = b^{-1}b = 1. \\ \text{Logo, } (ab)^{-1} = a^{-1}b^{-1}.$$

$$ii) |a| = \begin{cases} a & (a \geq 0) \\ -a & (a < 0). \end{cases}$$

Sendo $a \geq 0$, vem $-a \leq 0$ e $|a| = a \geq 0 \geq -a$ o que dá $|a| = a$ e $|a| \geq -a$; supondo $a < 0$, vem $-a > 0$ e $|a| = -a > 0 > a$ o que dá $|a| > a$ e $|a| = -a$.

Portanto, em todos os casos, $|a| \geq a \wedge |a| \geq -a$.

2) Sendo X subconjunto do espaço métrico A , prove que $\text{int } X$ é a reunião de todos os conjuntos abertos contidos em X .

Ache $\text{int } X$, $\text{ext } X$, $\text{sup } X$ e $\text{inf } X$ para o conjunto linear

$$X = [1, 3] \cup \{5, 6\} \cup \left\{ x \in \mathbb{R} : x = \frac{2n + (-1)^n}{n+1} \right\}.$$

Justifique as respostas.

R: Seja G a reunião de todos os conjuntos abertos contidos em X . Então G é aberto. O $\text{int } X$ é conjunto aberto contido em X e portanto $\text{int } X \subseteq G$. Se $x \in G$, então x pertence a um conjunto aberto T ($T \subseteq X$). Como T é aberto, existe $\varepsilon > 0$ tal que $V_\varepsilon(x) \subseteq T$ e portanto $V_\varepsilon(x) \subseteq X$ e $x \in \text{int } X$. Logo $x \in G \Rightarrow x \in \text{int } X$ e portanto $G \subseteq \text{int } X$. Podemos pois afirmar que $G = \text{int } X$.

Para o exercício proposto é

$$\text{int } X =]1, 3[$$

$$\text{ext } X = \mathbb{R} - \left\{ [1, 3] \cup \{5, 6\} \cup \left\{ x \in \mathbb{R} : x = \frac{2k+1}{2k+2} \right\} \right\}$$

$$\text{sup } X = 6$$

$$\text{inf } X = 1/2$$

3) Supondo que $\lim u_n/v_n = 1$ e v_n é sucessão limitada, prove que $u_n - v_n \rightarrow 0$.

Dê um exemplo de sucessões u_n e v_n para as quais $u_n/v_n \rightarrow 1$ mas $\lim(u_n - v_n) \neq 0$.

$$\text{Calcule } \lim_{n \rightarrow \infty} \left(\sqrt[n]{a} - 1 \right)^{\frac{\log n}{n^2}}.$$

$$R: \frac{u_n}{v_n} \rightarrow 1 \Leftrightarrow \frac{u_n}{v_n} = 1 + \alpha_n \quad (\alpha_n \rightarrow 0) \Leftrightarrow u_n = v_n + \alpha_n v_n \Leftrightarrow u_n - v_n = \alpha_n v_n.$$

Or $\alpha_n v_n \rightarrow 0$ e a proposição está provada.

Tomando $u_n = n$ e $v_n = n+1$, vem $u_n/v_n \rightarrow 1$ mas $u_n - v_n = -1$ e portanto $u_n - v_n$ não tende para 0 porque v_n não é limitada.

$$\lim_{n \rightarrow \infty} \frac{\log n}{n^2} \log(\sqrt[n]{a} - 1) = \lim_{n \rightarrow \infty} \frac{\log n}{n^2} \log(e^{\frac{1}{n} \log a} - 1) = \\ = \lim_{n \rightarrow \infty} \frac{\log n}{n^2} \log\left(\xi \frac{1}{n} \log a\right)$$

$$= \lim_{n \rightarrow \infty} \frac{\log n}{n^2} (\log \xi + \log \log a - \log n) =$$

$$= \lim_{n \rightarrow \infty} \left[\frac{\log n}{n^2} \log \xi + \frac{\log n}{n^2} \log \log a - \left(\frac{\log n}{n}\right)^2 \right]$$

$$= 0.$$

Portanto, $\lim (n\sqrt[n]{a} - 1) \frac{\log n}{n^2} = 1$.

4) Seja p_n uma sucessão de termos positivos e

$$P_n = p_n \frac{a_n}{a_{n+1}} - p_{n+1} (a_n > 0).$$

Demonstre que: (i) $\lim P_n > 0 \Rightarrow \Sigma a_n$ converge; (ii) $\overline{\lim} P_n < 0 \wedge \Sigma (1/p_n)$ diverg, $\Rightarrow \Sigma a_n$ diverge.

Deduzo o corolário que se obtém fazendo $p_n = n$ e mostre que se trata de um corolário do critério de RAABE.

Estude a natureza das séries

$$\Sigma \frac{1 + (-1)^n}{n} \text{ e } \Sigma \frac{1 + (-1)^n}{n^2}.$$

R: Considerando $P > 0$, tome-se $\varepsilon > 0$ suficientemente pequeno por forma que $K = P - \varepsilon > 0$. Portanto, a partir de certa ordem, é

$$P_n \frac{a_n}{a_{n+1}} - p_{n+1} > K$$

ou

$$\frac{a_{n+1}}{a_n} < \frac{P_n}{P_{n+1} + K}$$

o que implica a convergência de Σa_n (1.ª parte do critério de KUMMER).

Com $\overline{P} < 0$, tome-se $\varepsilon > 0$ suficientemente pequeno por forma que $K = \overline{P} + \varepsilon < 0$ e então, a partir de certa ordem,

$$P_n \frac{a_n}{a_{n+1}} - p_{n+1} < 0$$

o que dá

$$\frac{a_{n+1}}{a_n} > \frac{P_n}{P_{n+1}},$$

condição que, juntamente com a divergência de $\Sigma (1/p_n)$, garante a divergência de Σa_n (2.ª parte do critério de KUMMER).

Fazendo $p_n = n$, vem

$$P_n = n \left(\frac{a_n}{a_{n+1}} - 1 \right) - 1.$$

$\lim P_n > 0 \Rightarrow \Sigma a_n C$, equivale a afirmar que

$$\lim n \left(\frac{a_n}{a_{n+1}} - 1 \right) > 1 \Rightarrow \Sigma a_n C.$$

$\overline{\lim} P_n < 0 \wedge \Sigma (1/n) D$. $\Rightarrow \Sigma a_n D$. equivale a

$$\overline{\lim} n \left(\frac{a_n}{a_{n+1}} - 1 \right) < 1 \Rightarrow \Sigma a_n D.$$

Para $\Sigma \frac{1 + (-1)^n}{n}$ basta notar que a soma dos 2n primeiros termos é a soma dos n primeiros termos da série divergente $\Sigma \frac{1}{n}$ para garantirmos que a série é divergente; para $\Sigma \frac{1 + (-1)^n}{n^2}$, a soma dos 2n primeiros é a soma dos n primeiros termos de $\Sigma \frac{1}{2n^2}$ (conv.) e portanto a série converge.

5) Demonstre que, sendo f função contínua incessantemente crescente em $[a, b]$, a equação $f(x) = k$ ($f(a) \leq k \leq f(b)$) admite uma solução única x_0 .

R: Por ser contínua em $[a, b] \exists x_0 \in]a, b[$ $f(x_0) = k$ mas como é incessantemente crescente é $\forall x \neq x_0 f(x) \neq f(x_0) = k$ e portanto x_0 é a única solução.

6) Prove que sendo f função periódica, também f' é periódica.

Se f não-constante tem o período μ , é necessário que f' tenha o período μ ? Porquê?

R: $f(x + \mu) = f(x) \Rightarrow f'(x + \mu) = f'(x)$ e portanto f' também é periódica.

Supondo que a derivada tinha o período $\mu_1 < \mu$, seria $\mu = m \mu_1$ e viria $f'(x + \mu_1) = f'(x)$, o que implicaria $f(x + \mu_1) = f(x) + K$.

Então, obter-se-ia $f(x + m \mu_1) = f(x) + K m$ ou $f(x + \mu) = f(x) + K m$ donde resultaria $K m = 0$ ou $K = 0$. Mas então seria $f(x + \mu_1) = f(x)$ o que é contrário à hipótese de ser μ o período de f. Logo a derivada f' tem o mesmo período de f.

I. S. G. E. F. — MATEMÁTICAS GERAIS — 1.ª cadeira — Exame final — Ano lectivo 1969-70 — Ponto n.º 3.

5775 — 1) Sendo n inteiro positivo, prove que

$$\left(\frac{-1 + i\sqrt{3}}{2} \right)^{5n} + \left(\frac{-1 - i\sqrt{3}}{2} \right)^{5n} = 2.$$

R.:

$$\begin{aligned} \left(\frac{-1 + i\sqrt{3}}{2} \right)^{5n} &= \left(-\frac{1}{2} + i\frac{\sqrt{3}}{2} \right)^{5n} = \\ &= \left(\text{cis } \frac{2\pi}{3} \right)^{5n} = \text{cis } 2n\pi = 1 \end{aligned}$$

$$\left(\frac{-1 - i\sqrt{3}}{2}\right)^{5n} = \left(-\frac{1}{2} - i\frac{\sqrt{3}}{2}\right)^{5n} =$$

$$= \left(\operatorname{cis} \frac{4\pi}{3}\right)^{5n} = \operatorname{cis} 4n\pi = 1$$

Logo

$$\left(\frac{-1 + i\sqrt{3}}{2}\right)^{5n} + \left(\frac{-1 - i\sqrt{3}}{2}\right)^{5n} = 2.$$

2) A é o espaço métrico com a distância definida do modo seguinte:

$$d(x, y) = \begin{cases} 0 & (x = y) \\ 1 & (x \neq y) \end{cases}.$$

Seja $X \subseteq A$. X é aberto? É fechado? É limitado? Justifique as respostas.

Tomando o conjunto linear

$$X = \{1, 2, 3\} \cup \left\{x \in \mathbb{R} : x = \frac{e^n}{n} \ (n = 1, 2, \dots)\right\},$$

indique $\operatorname{int} X$, $\operatorname{ext} X$, $\operatorname{front} X$, $\sup X$ e $\inf X$. O conjunto X é fechado? É aberto? Justifique as respostas.

R.: Atendendo a que $V_\varepsilon(a) = \{a\}$ ($\varepsilon \leq 1$) e $V_\varepsilon(a) = \mathbb{R}$ ($\varepsilon > 1$), é fácil concluir que X é aberto e portanto não é fechado. É claro que o conjunto X é limitado pois $V_\varepsilon(a) = X$ ($\varepsilon > 1$).

Para o conjunto linear apresentado tem-se: $\operatorname{int} X = \emptyset$, $\operatorname{ext} X = \mathbb{R} - \{1, 2, 3\} \cup \left\{x : x = \frac{e^n}{n}\right\}$, $\operatorname{front} X = X$, $\sup X = +\infty$, $\inf X = 0$ e X é fechado.

3) Supondo que $u_n - v_n \rightarrow 0$ e que $1/v_n$ é sucessão limitada, prove que $\lim u_n/v_n = 1$. Dê um exemplo de sucessões u_n e v_n para as quais $u_n - v_n \rightarrow 0$ mas $\lim u_n/v_n \neq 1$.

$$\text{Calcule } \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n^n}\right)^{e^n}.$$

R.:

$$\lim (u_n - v_n) = 0 \iff u_n - v_n = \alpha_n \ (z_n \rightarrow 0) \iff$$

$$\iff u_n = v_n + \alpha_n \iff \frac{u_n}{v_n} = 1 + \frac{\alpha_n}{v_n}$$

donde se conclui facilmente, em consequência da hipótese sobre $1/v_n$, que $\frac{u_n}{v_n} \rightarrow 1$.

$$\text{Com } u_n = 2/n \text{ e } v_n = 1/n, \text{ vem } u_n - v_n = \frac{1}{n} \rightarrow 0$$

mas, no entanto, $\lim \frac{u_n}{v_n} = 2$ porque $1/v_n$ não é limitada.

Atendendo a que

$$\lim e^n \log \left(1 + \frac{1}{n^n}\right) = \lim e^n n \frac{1}{n^n} = \lim \left(\frac{e}{n}\right)^n = 0,$$

$$\left(1 + \frac{1}{n^n}\right)^{e^n} \rightarrow 1.$$

4) Seja $\sum a_n$ ($a_n \geq 0$) uma série convergente e b_n ($b_n \geq 0$) uma sucessão limitada. Mostre que a série $\sum a_n b_n$ é convergente.

Dada a série $\sum_1^\infty \frac{(2x+1)^n}{n(n+2)}$, determine o intervalo de convergência e a natureza nos extremos desse intervalo.

R.: Como $\frac{a_n b_n}{a_n} = b_n < K$, então a convergência de $\sum a_n$ implica a da série $\sum a_n b_n$.

Como $\left|\frac{u_n + 1}{u_n}\right| \rightarrow (2x + 1)$, a série dada converge absolutamente para $|2x + 1| < 1$ ($-1 < x < 0$) e diverge para $|2x + 1| > 1$ ($x > 0 \vee x < -1$).

Para $x = 0$, obtém-se a série $\sum_1^\infty \frac{1}{n(n+2)}$ absolutamente convergente e, para $x = -1$, vem

$$\sum_1^\infty (-1)^n \frac{1}{n(n+2)}$$

que é também absolutamente convergente.

5) Sendo f função contínua e biunívoca em $[a, b]$, prove que f é incessantemente crescente ou decrescente em $[a, b]$.

R.: Sendo f biunívoca em $[a, b]$, vem $f(a) \neq f(b)$. Suponhamos que $f(a) < f(b)$. Provaremos que f é crescente em $[a, b]$.

Sejam $x_1, x_2 \in [a, b]$ tais que $x_1 < x_2$. Deveremos ter $f(x_1) < f(x_2)$. Com efeito, se fosse $f(x_1) = f(x_2)$, a função não era biunívoca. Se fosse $f(x_1) > f(x_2)$, duas hipóteses teríamos de considerar: $f(x_2) > f(a)$ e $f(x_2) > f(b)$. No primeiro caso, vinha $f(x_2) < f(a) < f(b)$ e portanto existiria $\bar{x} \in]x_2, a[$ tal que $f(\bar{x}) = f(a)$, o que é absurdo em face da biunivocidade; no segundo caso, ter-se-ia $f(a) < f(x_2) < f(x_1)$ e, análogamente, existiria $\bar{x} \in]a, x_1[$ tal que $f(\bar{x}) = f(x_2)$, o que é absurdo pela mesma razão.

Logo, podemos efectivamente concluir que

$$\forall x_1, x_2 \in [a, b] \quad x_1 < x_2 \implies f(x_1) < f(x_2).$$

Analogamente se provaria a proposição na hipótese $f(a) > f(b)$.

6) Sendo k parâmetro real e sabendo que a equação da tangente à curva representativa de $y = \frac{x+k}{x}$ é $x - 4y - 4 = 0$, ache o valor do parâmetro k e as coordenadas do ponto de tangência.

R.: $y' = -\frac{k}{x^2}$ e portanto terá de ser

$$\begin{cases} -\frac{k}{x^2} = \frac{1}{4} \\ y = 1 + \frac{k}{x} \\ x - 4y - 4 = 0 \end{cases}$$

sistema cuja solução é

$$\begin{cases} k = -4 \\ y = 0 \\ x = 4 \end{cases}$$

I. S. C. E. F. — MATEMÁTICAS GERAIS — Exame final —
 Ano lectivo de 1969/70 — Época de Outubro — Ponto
 n.º 4 — 1-10-1970.

5776 — 1) Demonstre que $(A - C) \cup (B - D) \subseteq (A \cup B) - (C \cap D)$. Dê um exemplo em que se tenha $(A - C) \cup (B - D) = (A \cup B) - (C \cap D)$.

R:

$$\begin{aligned} (A - C) \cup (B - D) &= (A \cap \bar{C}) \cup (B \cap \bar{D}) = (A \cup B) \cap \\ &\cap (A \cup \bar{D}) \cap (\bar{C} \cup B) \cap (\bar{C} \cup \bar{D}) \subseteq (A \cup B) \cap (\bar{C} \cup \bar{D}) = \\ &= (A \cup B) \cap [\bar{(C \cap D)}] = (A \cup B) - (C \cap D). \end{aligned}$$

Tomando

$$\begin{aligned} U &= \{1, 2, 3, 4, 5\}, A = \{1, 2\}, B = \{2, 3, 4\}, \\ C &= \{2, 3\}, D = \{1, 2, 3\} \end{aligned}$$

vem

$$(A - C) \cup (B - D) = (A \cup B) - (C \cap D) = \{1, 4\}.$$

2) Sendo X subconjunto do espaço métrico A , prove que $\text{int } X = \text{est } \bar{X}$ e $\text{front } X = \text{front } \bar{X}$.

Determine a e b por forma que o conjunto linear $X = \{3, a\} \cup \{x \in R : x = \frac{bn}{n+2} (n=1, 2, \dots)\}$ tenha

o derivado $X' = \{3\}$ e $\text{inf } X = 1/2$. Qual é o $\text{sup } X$? X é fechado? É aberto? Justifique as respostas.

R: a p. int. $X \Rightarrow V_\varepsilon(a) \subseteq X \Rightarrow V_\varepsilon(a) \cap \bar{X} = \emptyset \Rightarrow$
 \Rightarrow a p. ext. X ; a p. ext. $\bar{X} \Rightarrow V_\varepsilon(a) \cap \bar{X} = \emptyset \Rightarrow$
 $\Rightarrow V_\varepsilon(a) \subseteq X \Rightarrow$ a p. int. X .

Se a é ponto fronteiro de X (de \bar{X}) em qualquer sua vizinhança, por mais pequeno que seja o valor de ε , há sempre pontos de X e de \bar{X} . Logo, a também é ponto fronteiro de \bar{X} (de X).

Como $\lim_{n \rightarrow \infty} \frac{bn}{n+2} = b$, terá de ser $b = 3$. $\text{inf } X = 1/2 \Rightarrow a = 1/2$; $\text{sup } X = 3$ e X é fechado.

3) Dada a sucessão u_n , suponha que u_{2n}, u_{2n+1} e u_{3n} são convergentes. Prove que u_n converge.

Diga qual é o conjunto dos sublimites da sucessão

$$1, \frac{1}{2}, 1, \frac{1}{2}, \frac{1}{3}, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \dots$$

Indique, justificando, os seus limites máximo e mínimo.

R: Supondo que $u_{2n} \rightarrow u, u_{2n+1} \rightarrow v$ e $u_{3n} \rightarrow w$, notemos que $3n$ é número par (quando n for par) ou ímpar (quando n for ímpar) e portanto as subsucessões de u_{3n} que se obtêm com n par e n ímpar são, respectivamente, subsucessões de u_{2n} e u_{2n+1} . Podemos então garantir que $u = v = w$ e portanto u_n converge para u .

O conjunto dos sublimites da sucessão dada é

$$\left\{1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, \dots, 0\right\}$$

Os limites máximo e mínimo são, respectivamente, 1 e 0.

4) Prove que, sendo $u_n > -1 (n \geq 0)$, a convergência de $\sum_0^\infty u_n$ e $\sum_0^\infty u_n^2$ implica a convergência de $\sum_0^\infty \log(1 + u_n)$.

Discuta a natureza da série

$$\sum_{n=1}^\infty \frac{1 \cdot 3 \dots (2n+1)}{2 \cdot 4 \dots 2n} (1 - x^2)^n.$$

R: Notando que $u_n \rightarrow 0, \log(1 + u_n) = u_n - \lambda u_n^2$ onde $\lambda \rightarrow \frac{1}{2}$ quando $u_n \rightarrow 0$.

Tem-se então $\sum_0^{\infty} \log(1 + u_n) = \sum_0^{\infty} u_n - \sum_0^{\infty} \lambda u_n^2$.

Ora como $\sum_0^{\infty} u_n$ converge e $\sum_0^{\infty} \lambda u_n^2$ também converge, porque $\frac{\lambda u_n^2}{u_n} \rightarrow \frac{1}{2} \neq 0, \infty$, a série $\sum_0^{\infty} \log(1 + u_n)$ (soma de séries convergentes) também converge.

Para a série $\sum_1^{\infty} \frac{1 \cdot 3 \cdots (2n+1)}{2 \cdot 4 \cdots 2n} (1-x^2)^n$ vem

$$\left| \frac{u_{n+1}}{u_n} \right| = \frac{2n+3}{2n+2} |1-x^2| \rightarrow |1-x^2|$$

Como

$$|1-x^2| < 1 \Leftrightarrow -1 < 1-x^2 < 1 \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} 1-x^2 < 1 \Leftrightarrow x \neq 0 \\ 1-x^2 > -1 \Leftrightarrow -\sqrt{2} < x < \sqrt{2} \end{cases}$$

a série converge para $-\sqrt{2} < x < 0$ e $0 < x < \sqrt{2}$.

Por outro lado,

$$|1-x^2| > 1 \Leftrightarrow \begin{cases} 1-x^2 > 1 \text{ (impossível)} \\ 1-x^2 < -1 \Leftrightarrow x < -\sqrt{2} \vee x > \sqrt{2} \end{cases}$$

e portanto a série diverge para $x < -\sqrt{2}$ e $x > \sqrt{2}$.

Para $x=0$, vem a série $\sum_1^{\infty} \frac{1 \cdot 3 \cdots (2n+1)}{2 \cdot 4 \cdots 2n}$ que

é divergente pois $\frac{u_{n+1}}{u_n} = \frac{2n+3}{2n+1} \rightarrow 1+0$.

Para $x = \pm \sqrt{2}$ vem $\sum_1^{\infty} (-1)^n \frac{1 \cdot 3 \cdots (2n+1)}{2 \cdot 4 \cdots 2n}$

e, como $\left| \frac{u_{n+1}}{u_n} \right| = \frac{2n+3}{2n+2} > 1$, o seu termo geral não tende para zero e portanto a série diverge.

5) Seja f contínua em $[a, b]$ e admita que $f([a, b]) \subseteq [a, b]$. Prove que, sob estas condições, a equação $f(x) = x$ tem sempre uma raiz em $[a, b]$. Exemplifique com $f(x) = \sqrt{x}$ em $[0, 4]$.

R.: Nos casos em que $f(a) = a$ ou $f(b) = b$ a proposição é óbvia. Se for $f(a) > a$ e $f(b) < b$ a função $g(x) = f(x) - x$ toma sinais contrários em a e b e portanto $g(x)$ anula-se num ponto c interior (teorema de BOLZANO-CAUCHY): $g(c) = 0$ ou $f(c) = c$.

Com $f(x) = \sqrt{x}$ em $[0, 4]$ a equação $\sqrt{x} = x$ tem as raízes $x=0$ e $x=1$.

6) A função F satisfaz à condição $F(c-0) < F(c) < F(c+0)$ no ponto interior c do seu domínio. Estude a existência de $F'(c)$.

$$R.: F'_-(c) = \lim_{x \rightarrow c-0} \frac{F(x) - F(c)}{x - c} = +\infty$$

$$F'_+(c) = \lim_{x \rightarrow c+0} \frac{F(x) - F(c)}{x - c} = +\infty$$

Portanto, no ponto $x=c$ vem $F'(c) = +\infty$.

I. S. C. E. F. — MATEMÁTICAS GERAIS — Exame final — Ano lectivo de 1969/70 — Época de Outubro — Ponto n.º 5 — 1-10-1970.

5777 — 1) Considere as aplicações de R em R definidas por $f(x) = x^5$, $g(x) = x - 1$ e $h(x) = x$.

Determine $f \circ g$, $g \circ f$, $(f \circ g) \circ h$, $f \circ (g \circ h)$, $g^{-1} \circ f^{-1}$ e $(f \circ g)^{-1}$.

$$R.: f \circ g = (x-1)^5, g \circ f = x^5 - 1, (f \circ g) \circ h = (x-1)^5, f \circ (g \circ h) = (x-1)^5, g^{-1} \circ f^{-1} = 1 + \sqrt[5]{y}, (f \circ g)^{-1} = 1 + \sqrt[5]{y}.$$

2) Prove as proposições seguintes:

- Seja a um número real qualquer e $X = \{x: x \text{ é racional } \wedge x < a\}$, então $a = \sup X$.
- Seja A conjunto linear não-vazio limitado inferiormente e B o conjunto dos números reais tais que $x \in A$, então B é superiormente limitado e $\inf A = -\sup B$.

Indique o conjunto derivado, o supremo e o ínfimo do conjunto linear

$$X = \left\{ x: x = (-1)^n + \frac{1}{m} \text{ (} m, n = 1, 2, \dots \text{)} \right\}.$$

R.:

- O número a é majorante de X e, como $\forall \delta > 0, \exists x \in X$ a $-\delta < x < a$, é evidente que $a = \sup X$.
- Seja $l = \inf A$, tem-se $\forall x \in A, x \geq l$ e $\forall \delta > 0, \exists x' \in A: x' < l + \delta$. Portanto, $\forall -x \in B, -x \leq -l$ e $\forall \delta > 0, \exists -x' \in B, -x' > -l - \delta$. Logo $-l$ é o $\sup B$.

$$\text{Para } X = \left\{ x: x = (-1)^n + \frac{1}{m} \text{ (} m, n = 1, 2, \dots \text{)} \right\}$$

tem-se $X' = \{-1, 1\} \cup X$, $\inf X = -1$, $\sup X = 2$.

3) Mostre que, sendo $u_n > 0$ e $u_{n+1}/u_n \leq k < 1$ (k constante) $\forall n \in N$, então $\lim u_n = 0$. O que

se pode afirmar quando se sabe apenas que $u_{n+1}/u_n < 1$? Justifique.

Sendo $u_n = n!/n^n$, indique o $\lim u_n$ a partir de $\lim u_{n+1}/u_n$.

R.: $u_{n+1}/u_n \leq k < 1 \Rightarrow \sum u_n$ converge $\Rightarrow u_n \rightarrow 0$.
Quando $u_{n+1}/u_n < 1$ apenas se pode afirmar que $u_{n+1} < u_n$ e portanto $u_n \rightarrow u \geq 0$.

Para $u_n = n!/n^n$ vem $\lim \frac{u_{n+1}}{u_n} = \frac{1}{e} < 1$ e portanto $\lim u_n = 0$.

4) Sendo $a_n > 0$ ($n = 1, 2, \dots$) e $b_n = (a_1 + a_2 + \dots + a_n)/n$, prove que $b_1 + b_2 + \dots + b_n > a_1 \left(1 + \frac{1}{2} + \dots + \frac{1}{n}\right)$.

Tomando a série $\sum a_n$, aproveite o resultado anterior para justificar que $\sum b_n$ é sempre divergente.

Estude a natureza da série

$$\sum \frac{(p+1)(p+2)\dots(p+n)}{(q+1)(q+2)\dots(q+n)}$$

R: Como

$$b_1 = a_1$$

$$b_2 = \frac{a_1 + a_2}{2}$$

$$b_3 = \frac{a_1 + a_2 + a_3}{3}$$

.....

$$b_n = \frac{a_1 + a_2 + \dots + a_n}{n}$$

resulta imediatamente

$$b_1 + b_2 + \dots + b_n > a_1 \left(1 + \frac{1}{2} + \dots + \frac{1}{n}\right)$$

Esta desigualdade serve para provar que

$$b_1 + b_2 + \dots + b_n \rightarrow +\infty$$

pois

$$1 + \frac{1}{2} + \dots + \frac{1}{n} \rightarrow +\infty$$

Para estudar a série dada, notemos que

$$\frac{a_{n+1}}{a_n} = \frac{p+n+1}{q+n+1} \rightarrow 1$$

e o critério da razão não explica a natureza da série. Como

$$n \left(\frac{a_n}{a_{n+1}} - 1 \right) = \frac{(q-p)n}{n+p+1} \rightarrow q-p,$$

o critério de RAABE, permite concluir que a série converge com $q > p+1$ e diverge com $q \leq p+1$.

5) Seja f contínua no intervalo $[0, 2\pi]$ tal que $f(0) = f(2\pi)$. Prove que existe um ponto c nesse intervalo tal que $f(c) = f(c+\pi)$. Sugestão: considere a função auxiliar $g(x) = f(x) - f(x+\pi)$.

R: Observemos que $g(0) = f(0) - f(\pi)$ e $g(\pi) = f(\pi) - f(2\pi) = f(\pi) - f(0)$. Se $f(0) = f(\pi)$, a proposição é óbvia. Com $f(0) \neq f(\pi)$, a função contínua $g(x)$ tem sinais contrários em 0 e π e o teorema de BOLZANO-CAUCHY garante que existe um ponto c entre 0 e π tal que $g(c) = f(c) - f(c+\pi) = 0$.

6) Seja

$$f(x) = \begin{cases} \frac{x-1}{\sqrt{x}-1} & (x \neq 1) \\ 2 & (x = 1) \end{cases}$$

Prove que f é diferenciável para $x = 1$.

R: Como

$$\begin{aligned} f'(1) &= \lim_{x \rightarrow 1} \frac{\frac{x-1}{\sqrt{x}-1} - 2}{x-1} = \lim_{x \rightarrow 1} \frac{x-1-2(\sqrt{x}-1)}{(x-1)(\sqrt{x}-1)} = \\ &= \lim_{x \rightarrow 1} \frac{(\sqrt{x}-1)^2}{(x-1)(\sqrt{x}-1)} = \lim_{x \rightarrow 1} \frac{1}{\sqrt{x}+1} = \frac{1}{2}, \end{aligned}$$

a função é diferenciável.

I. S. C. E. F. — MATEMÁTICAS GERAIS — Exame final — Ano lectivo de 1969-70 — Época de Outubro — Ponto n.º 6 — 6-10-1970.

5778 — 1) Prove as seguintes propriedades no conjunto R dos números reais:

i) $-(-a) = a$

ii) $a \leq b \iff -a \geq -b$.

R: i) Dado o número real $a = [A_1, A_2]$, tem-se $-a = [-A_2, -A_1]$ e é claro que $-(-a) = [A_1, A_2]$.

ii) Sendo $a = [A_1, A_2]$ e $b = [B_1, B_2]$, tem-se $a = b \Leftrightarrow A_1 = B_1 \wedge A_2 = B_2$ mas então também $-A_1 = -B_1 \wedge -A_2 = -B_2$ o que mostra que $-a = -b$.

Supondo, por exemplo, $a > b$ tem-se $A_1 \cap B_2 \neq \emptyset$ o que indica que

$$-b = [-B_2, -B_1] > [-A_2, -A_1] = -a.$$

2) Prove por indução que todo o conjunto linear finito tem mínimo e máximo. Mostre que

$$\max \{a, b\} = \frac{1}{2} [a + b + |a - b|]$$

$$\min \{a, b\} = \frac{1}{2} [a + b - |a - b|]$$

e prove também que

$$\max \{a + c, b + d\} \leq \max \{a, b\} + \max \{c, d\}.$$

Ache o supremo e o ínfimo de

$$\{x \in R : (x - a)(x - b)(x - c)(x - d) < 0\}$$

com

$$a < b < c < d.$$

R.: A proposição é óbvia para um conjunto de dois elementos. Supondo que a propriedade é verdadeira para um conjunto de m elementos, se tomarmos um conjunto com $m + 1$ elementos e considerarmos um seu subconjunto de m elementos, o mínimo existe. De facto, é o menor de dois elementos: o mínimo do subconjunto e o elemento que lhe não pertence. Mutatis mutandis, o raciocínio aplica-se ao máximo.

Supondo

$$a \leq b, \max \{a, b\} = b = \frac{1}{2} (a + b + b - a).$$

Com $a \leq b$, vem

$$\max \{a, b\} = b = \frac{1}{2} (a + b + b - a).$$

Análogamente se prova que

$$\min \{a, b\} = \frac{1}{2} [a + b - |a - b|].$$

$$\max \{a + c, b + d\} =$$

$$= \frac{1}{2} [(a + c) + (b + d) + |(a + c) - (b + d)|] \leq$$

$$\leq \frac{1}{2} [(a + c) + (b + d) + |a - b| + |c - d|] =$$

$$= \frac{1}{2} [a + b + |a - b|] + \frac{1}{2} [c + d + |c - d|] =$$

$$= \max \{a, b\} + \max \{c, d\}.$$

Para o conjunto

$$X = \{x \in R : (x - a)(x - b)(x - c)(x - d) < 0\},$$

vem $\sup X = d$ e $\inf X = a$.

3) Dada a sucessão $a, a + b, 1 + 1/2, a, a + b, 1 + 1/3, a, a + b, 1 + 1/4, \dots$ determine os números a e b por forma que a sucessão seja convergente. Justifique.

Admitindo que $u_n \rightarrow u$ e $v_n \rightarrow v$, mostre que, exceptuando os casos $(u = 1, v = \infty)$, $(u = \infty, v = 0)$ e $(u = 0, v = 0)$, a sucessão $w_n = u_n^{v_n}$ ($u_n > 0$) possui o limite u^v .

$$\text{Calcule } \lim \left(1 + 2 \operatorname{tg} \frac{1}{n}\right)^n.$$

R.: $a = 1$ e $b = 0$.

Notando que $\log w_n = v_n \log u_n \rightarrow v \log u$, facilmente se conclui que $w_n \rightarrow u^v$.

$$n \log \left(1 + 2 \operatorname{tg} \frac{1}{n}\right) = 2n \operatorname{tg} \frac{1}{n} =$$

$$= 2n \frac{\operatorname{sen} \frac{1}{n}}{\frac{1}{n}} \cdot \frac{1}{\cos \frac{1}{n}} \rightarrow 2$$

e portanto

$$\left(1 + 2 \operatorname{tg} \frac{1}{n}\right)^n \rightarrow e^2.$$

4) Dada a série $\sum u_n$, demonstre que

i) $|u_{n+1}/u_n| \leq h < 1 \Rightarrow \sum u_n$ é absolutamente convergente.

ii) $|u_{n+1}/u_n| \geq 1 \Rightarrow \sum u_n$ é divergente.

Estude a natureza da série de termo geral

$$u_n = \frac{a^n}{(1 + a)(1 + a^2) \dots (1 + a^n)} \quad (a > 0).$$

R.:

$|u_{n+1}/u_n| \leq h < 1 \Rightarrow \sum |u_n| \text{ conv.} \Rightarrow \sum u_n \text{ conv. absol.}$

$|u_{n+1}/u_n| \geq 1 \Rightarrow \sum |u_n| \text{ div} \Rightarrow u_n \text{ não é evanescente} \Rightarrow \sum u_n \text{ div.}$

$$\frac{u_{n+1}}{u_n} = \frac{a}{1+a^{n+1}} \rightarrow \begin{cases} a & (a < 1) \\ 1/2 & (a = 1) \\ 0 & (a > 1) \end{cases}$$

e o critério da razão permite concluir que a série converge para todos os valores de a .

5) Sendo f monótona em $[a, b]$, justifique que f só pode possuir descontinuidades de 1.ª espécie nesse intervalo. Estude a continuidade da função

$$f(x) = \begin{cases} 0 & (x \text{ rac.}) \\ x & (x \text{ irrac.}) \end{cases}$$

R.: Uma função monótona tem sempre limites laterais nos pontos de $[a, b]$ e portanto só pode possuir descontinuidades de 1.ª espécie.

A função

$$f(x) = \begin{cases} 0 & (x \text{ rac.}) \\ x & (x \text{ irrac.}) \end{cases}$$

é contínua para $x = 0$ pois $|f(x) - f(0)| = |x|$. Em todos os outros pontos apresenta descontinuidades de 2.ª espécie.

6) Considere a função

$$g(x) = \begin{cases} \sqrt{-x} & (x < 0) \\ \sqrt{x} & (x \geq 0) \end{cases}$$

e calcule $g'(0)$. Esboce a imagem da função na vizinhança de $x = 0$.

R.: Notando que

$$\frac{g(x) - g(0)}{x} = \begin{cases} -\frac{1}{\sqrt{-x}} & (x < 0) \\ \frac{1}{\sqrt{x}} & (x > 0) \end{cases}$$

resulta $g'(0) = \mp \infty$.

I. S. C. E. F — MATEMÁTICAS GERAIS — Exame final — Ano lectivo de 1969-70 — Época de Outubro — Ponto n.º 7 — 6-10-1970.

5779 — 1) Prove a seguinte propriedade: se $a, b \in R$ e $a > b$, então $a + c > b + c \forall c \in R$.

R.: Com efeito, seja $a = [A_1, A_2]$, $b = [B_1, B_2]$, $c = [C_1, C_2]$, $a + c = [D_1, D_2]$ e $b + c = [E_1, E_2]$. Como $a > b$, existe $a'_1 \in A_1$ e $b'_2 \in B_2$ tais que $a'_1 > b'_2$ ou $a'_1 - b'_2 > 0$. Então, poderemos determinar $c'_1 \in C_1$ e $c'_2 \in C_2$ tais que $c'_2 - c'_1 < a'_1 - b'_2$ ou $b'_2 + c'_2 < a'_1 + c'_1$. Mas $b'_2 + c'_2 \in E_2$ e $a'_1 + c'_1 \in D_1$ e portanto $a + c > b + c$.

2) Prove que λ é ponto de acumulação de A se e só se em qualquer vizinhança de λ existe pelo menos um elemento de A distinto de λ .

Mostre que $\sup A = \sup \bar{A}$, onde \bar{A} designa o fecho de A . Indique dois conjuntos A e B tais que $A \cap B = \emptyset$ e $\bar{A} \cap \bar{B} \neq \emptyset$.

Indique o ínfimo, o supremo e o fecho do conjunto $X = \{x \in R : x = n^{(-1)^m} \text{ (} m, n = 1, 2, \dots)\}$.

R.: Se λ é ponto de acumulação de A , numa $V_\varepsilon(\lambda)$, por menor que seja ε , existe uma infinidade de elementos de A e portanto existe sempre um elemento de A distinto de λ . Reciprocamente, se em qualquer $V_\varepsilon(\lambda)$ existe um elemento de A distinto de λ , há uma infinidade de elementos de A em qualquer $V_\varepsilon(\lambda)$.

Notando que $\bar{A} = A \cup A'$, basta examinar as hipóteses $\sup \bar{A} > \sup A$ e $\sup A > \sup \bar{A}$. No primeiro caso haveria $\bar{a} > \sup A$, o que é absurdo; no segundo caso, haveria $a > \sup \bar{A}$ que é uma contradição. Logo $\sup A = \sup \bar{A}$.

Sendo $A =]a, b[$ e $B =]b, c[$ vem $A \cap B = \emptyset$ e $\bar{A} \cap \bar{B} = \{b\}$. $\inf X = 0$, $\sup X = +\infty$, $\bar{X} = X \cup \{0\}$.

3) Indique, justificando, quais das proposições seguintes são verdadeiras:

- Se $u_{2n} \rightarrow u$ e $u_{2n+1} \rightarrow v$, u e v são os únicos sublimites de u_n .
- Se o conjunto dos termos de uma sucessão não tem máximo nem mínimo, a sucessão é divergente.
- Se $\forall n \in N \ u_n > 0 \wedge u_n \rightarrow 0$, então u_n é decrescente.

$$\text{Calcule } \lim n \left(\sqrt[3]{2 + \frac{1}{n}} - \sqrt[3]{2} \right).$$

R.: A proposição i) é obviamente verdadeira. Para ii) basta notar que o ínfimo e o supremo de (u_n) são pontos de acumulação, e portanto sublimites de u_n , para provar que ii) é verdadeira. Para iii), observemos que $u_n = \frac{2 + (-1)^n}{n} > 0$ é evanescente e no entanto não é decrescente. Logo iii) é falsa.

$$n \left(\sqrt[3]{2 + \frac{1}{n}} - \sqrt[3]{2} \right) = n \sqrt[3]{2} \left(\sqrt[3]{1 + \frac{1}{2n}} - 1 \right) = \\ = n \sqrt[3]{2} \frac{1}{3} \xi \frac{1}{2n} \rightarrow \frac{\sqrt[3]{2}}{6}.$$

4) Se a série $\sum a_n$ ($a_n \geq 0$) é divergente e S_n designa a soma dos seus n primeiros termos, prove que $\sum (\sqrt{S_{n+1}} - \sqrt{S_n})$ também diverge.

Estude a natureza das séries

$$\sum_0^{\infty} (-1)^{n+1} e^{-n \operatorname{sen} x} \text{ e } \sum_1^{\infty} \frac{1}{n! x^n}.$$

R.: Se $\sum a_n$ diverge, então $S_n \rightarrow +\infty$ e a série de MENGOLO $\sum (\sqrt{S_{n+1}} - \sqrt{S_n})$ também diverge.

Notando que $\sum_0^{\infty} (-1)^{n+1} e^{-n \operatorname{sen} x}$ é série geométrica de razão $-e^{-\operatorname{sen} x}$, ela será convergente se e só se $\frac{1}{e^{\operatorname{sen} x}} < 1$ ou $\operatorname{sen} x > 0$, isto é, para $2k\pi < x <$

$< (2k+1)\pi$. A série de potências $\sum_1^{\infty} \frac{1}{n! x^n}$ converge absolutamente para $x \neq 0$.

5)

- i) Dê exemplo de uma função contínua num conjunto fechado limitado que não tome todos os valores desde o mínimo ao máximo.
- ii) Represente geometricamente uma função contínua em $[a, b]$ excepto no ponto interior c onde possui uma descontinuidade de primeira espécie com $f(c-0) < f(c) < f(c+0)$. Calcule $f'(c)$ para esta função.

$$\text{R.: i) } f(x) = \begin{cases} 1 & (0 \leq x \leq 1) \\ 2 & (x = 2) \end{cases}$$

ii)

$$f'_+(c) = +\infty \text{ e } f'_-(c) = +\infty, \text{ isto é, } f'(c) = +\infty.$$

6) Prove que, sendo g diferenciável para $x = a$, então

$$\lim_{h \rightarrow 0} \frac{g(a+h) - g(a-h)}{2h} = g'(a)$$

R.: Por definição de diferenciabilidade podemos escrever

$$g(a+h) - g(a) = h[g'(a) + \alpha] \quad (\lim_{h \rightarrow 0} \alpha = 0)$$

$$g(a-h) - g(a) = -h[g'(a) + \beta] \quad (\lim_{h \rightarrow 0} \beta = 0)$$

Subtraindo membro a membro, vem

$$g(a+h) - g(a-h) = 2h g'(a) + h(\alpha + \beta)$$

ou

$$\frac{g(a+h) - g(a-h)}{2h} = g'(a) + \frac{\alpha + \beta}{2},$$

o que prova a proposição.

Enunciados e soluções dos n.ºs 5773 a 5779 de Fernando de Jesus

I. S. T. — MATEMÁTICAS GERAIS — Ponto N.º 4—7-1-70.

5780 — 1) Justificando cuidadosamente as respostas, indique se são:

- a) reflexivas,
- b) simétricas,
- c) transitivas

as relações F , G e H , definidas em R pela forma seguinte:

$$x F y \Leftrightarrow |x - y| > 0$$

$$x G y \Leftrightarrow x^2 - y^2 \in Q \quad (Q \text{ é o conjunto dos racionais})$$

$$x H y \Leftrightarrow \exists_{u \in R} x = u y.$$

Nos casos em que se trate de relações de equivalência, indique a classe de equivalência do elemento 0 (zero) e uma outra classe de equivalência, distinta desta.

2) Seja f a aplicação de R em si mesmo definida por

$$f(x) = \begin{cases} -x & \text{se } x < 0 \\ \frac{1}{x} & \text{se } x > 0 \end{cases}.$$

a) Indique, justificando, se f é injectiva ou sobrejectiva.

b) Dê exemplos concretos de um conjunto limitado (em R) cujo transformado por f seja ilimitado e de um conjunto ilimitado que f transforme num conjunto limitado.

c) Mostre que

$$(f \circ f)(x) = f(-x), \quad \forall x \in R$$

(considere separadamente os casos $x > 0$, $x = 0$, $x < 0$).

Pondo $f_1 = f$ e, para todo o $n \in N$, $f_{n+1} = f_n \circ f$, aproveite o resultado anterior para determinar explicitamente a aplicação f_n .

3) Sendo u_n e v_n os termos gerais de duas sucessões de termos reais e $w_n = u_n + v_n$, indique, justificando, quais das proposições seguintes são necessariamente verdadeiras e mostre, por meio de exemplos concretos que as restantes podem não o ser:

- a) Se u_n e v_n são divergentes, w_n também o é.
- b) Se u_n e v_n são limitadas, w_n também o é.
- c) Se u_n e v_n são ilimitadas, w_n também o é.
- d) Se u_n e v_n são crescentes, w_n também o é.
- e) Se u_n e v_n são monótonas, w_n também o é.

4) a) Prove que, para que $c \in R$ seja ponto de acumulação do conjunto

$$A_1 \cup A_2 \cup \dots \cup A_p$$

($p \in N$) é necessário e suficiente que c seja ponto de acumulação de um, pelo menos, dos conjuntos A_1, A_2, \dots, A_p .

b) Se em vez de uma reunião de conjuntos, em número finito, se tratasse da reunião de uma infinidade de conjuntos, a condição seria ainda suficiente? E necessária? Justifique.

I. S. T. — MATEMÁTICAS GERAIS — Ponto N.º 2 — 7-1-70.

5781 — 1) Justificando cuidadosamente as respostas, indique quais dos seguintes subconjuntos de R são

- a) abertos,
- b) fechados,
- c) limitados (em R):

$$A = \{x: \exists_{y \in R} x^2 + y^2 < 0\}$$

$$B = \left\{x: \frac{1}{x} \in Z\right\} \quad (Z \text{ é o conjunto dos inteiros})$$

$$C = \{x: |x - 1| + |x + 1| > 4\}.$$

2) Seja f a aplicação de R em si mesmo definida pela forma seguinte:

$$f(x) = \begin{cases} -x & \text{se } x \in Q \quad (Q \text{ é o conjunto dos racionais}) \\ \frac{1}{x} & \text{se } x \in R \setminus Q. \end{cases}$$

- a) Prove que f é bijectiva.
- b) Indique, justificando, quais são os valores de x que verificam a condição:

$$f(x) = x.$$

e) Designando por A o transformado por f do intervalo $[1, 2]$, indique (quando possível) o supremo, o ínfimo, o máximo e o mínimo do conjunto A .

d) Pondo $f_1 = f \circ e$, para todo o $n \in N$, $f_{n+1} = f_n \circ f$, determine explicitamente a aplicação f_n .

3) Sendo u_n o termo geral de uma sucessão de termos reais e $v_n = u_n^2$, indique, justificando, quais das proposições seguintes são necessariamente verdadeiras e mostre, por meio de exemplos concretos, que as restantes podem não o ser:

- a) Se v_n é convergente, u_n também o é.
- b) Se v_n é divergente, u_n também o é.
- c) Se v_n é majorada, u_n é limitada.
- d) Se v_n é crescente, u_n é crescente.
- e) Se v_n é crescente, u_n é monótona.

4) Convencionemos dizer que um conjunto $A \subset R$ é simétrico em relação à origem sse

$$\forall_{x \in R} (x \in A \implies -x \in A)$$

e que um conjunto $B \subset R$ é fechado a respeito da adição sse

$$\forall_{x, y \in R} (x \in B \wedge y \in B \implies x + y \in B).$$

Nestas condições, sendo C um subconjunto de R , considere a relação G , em R , definida por

$$x G y \iff x - y \in C$$

e prove que:

- a) G é reflexiva sse $0 \in C$.
- b) G é simétrica sse C é simétrico em relação à origem.
- c) G é transitiva sse C é fechado a respeito da adição.

Enunciados de J. Campos Ferreira

Université Libre de Bruxelles — Faculté des Sciences Appliquées — ALGÈBRE-ANALYSE.

1ère candidature.

5782 — 1. En quels points les fonctions suivantes sont-elles discontinues

- a) $f(x) = [x]$ où $[x]$ représente la partie entière de x

$$b) f(x) = \frac{1}{1+2\sin x}$$

$$c) f(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ e^{-\frac{1}{x}} & \text{si } x > 0. \end{cases}$$

2. Montrer que la fonction $y = \frac{1}{x}$ est continue sur l'intervalle $]0, 1[$ mais n'est pas uniformément continue.

3a) Montrer à l'aide de la définition de la dérivée que

$$(e^{ax})' = a e^{ax} \quad a \in \mathbb{R}.$$

Application: Calculer les dérivées des fonctions suivantes à partir de a)

1. $Sh x$
2. $Ch x$
3. $Th x$
4. $ln x$.

4. Etudier la fonction $y = |x|$ est-elle continue et dérivable en tout point x ?

5. En quels points la fonction $y = |\cos x|$ est-elle non dérivable?

6. Étudier la variation des fonctions suivantes

$$a) \begin{cases} 0 & \text{si } x = 0 \\ \cos \frac{1}{x} & \text{si } x \neq 0 \end{cases} \quad b) \begin{cases} 0 & \text{si } x = 0 \\ x \cos \frac{1}{x} & \text{si } x \neq 0 \end{cases}$$

$$c) \begin{cases} 0 & \text{si } x = 0 \\ x^2 \cos \frac{1}{x} & \text{si } x \neq 0. \end{cases}$$

En quels points ces fonctions sont-elles continues? Dérivables? Continument dérivables? Représentez les sur un graphe.

1ère candidature.

5783 — 1. Calculer les dérivées partielles de la fonction.

$$f = (xy)^z$$

2. Calculer les dérivées partielles de la fonction $f = e^{\sin y/x}$.

3. Calculer les dérivées partielles d'ordre 1 de y par rapport aux variables indépendantes strictement positives x_1, x_2, x_3 , si

$$y = \sqrt{u^2 + v^2} \text{ et } u = \log x_1 \cdot x_2 \cdot x_3 \quad v = e^{\sin(x_1 + x_2 + x_3)}.$$

4. Calculer à 1/10 près $\frac{\partial f}{\partial x}$ (2.1) et $\frac{\partial f}{\partial y}$ (2.1)

$$\text{si } f(x, y) = \sqrt{x \sin y + \frac{x}{y}}.$$

5. Soit $X(t)$ une matrice dont les éléments $x_{ij}(t)$ sont des fonctions dérivables de t . Calculer

$$a) \frac{d}{dt} X^{-1} \quad b) \frac{d}{dt} X^2.$$

6. Calculer les dérivées partielles d'ordre 1 de la fonction

$$f(x, y, z) = x^{(y^z)}.$$

7. Montrer que l'équation $y^3 + 5xy^2 + 10y - 10x^2 = 0$ définit, dans le voisinage de $x = 0$ une et une seule fonction réelle $y(x)$.

Développer cette fonction $y(x)$ en série de TAYLOR jusqu'au terme en x^5 inclus.

$$8. \text{ Si } \begin{cases} x = \varphi + \sin \varphi \\ y = 1 - \cos \varphi \end{cases}$$

sont les équations paramétriques d'une courbe plane d'équation cartésienne $y = F(x)$, calculer $F'(x)$ et étudier la variation de $F'(x)$ pour $0 < x < 2\pi$.

9. Considérons la surface d'équation $xy + yz - xz = 2$.

Déterminer pour quelles valeurs de x et y il est possible d'appliquer le théorème des fonctions implicites. Si (x_0, y_0) est l'un de ces points, calculer (de deux manières: en explicitant ou en n'explicitant pas z en fonction de x et y) les dérivées partielles premières $\frac{\partial z}{\partial x}(x_0, y_0)$, $\frac{\partial z}{\partial y}(x_0, y_0)$ et la différentielle totale $dz(x_0, y_0)$.

10. Calculer la matrice jacobienne de u, v par rapport à x, y si u et v sont les fonctions définies implicitement par les relations:

$$\begin{cases} u^2 - v^2 + 2x + 3y = 0 \\ uv + x - y = 0. \end{cases}$$

1^{ère} candidature.

5784 — 1. Démontrez que :

Si A et B sont des matrices carrées symétriques, alors AB est une matrice symétrique si et seulement si A et B commutent ($A \cdot B = B \cdot A$).

2. Sans effectuer le produit, trouvez le rang de la matrice $C = A \cdot B$ où

$$A = \begin{pmatrix} 2 & -1 \\ 3 & 0 \end{pmatrix} \text{ et } B = \begin{pmatrix} 1 & 2 & 3 \\ -4 & 0 & 5 \end{pmatrix}.$$

3. On considère la transformation linéaire de l'espace vectoriel R^2 définie dans la base canonique par la matrice $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Montrer qu'il n'existe aucune base de R^2 par rapport à laquelle la matrice de cette transformation linéaire soit diagonale.

4. Trouver les valeurs propres et les vecteurs correspondants de la matrice $\begin{pmatrix} 1 & 0 & 2 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix}$.

5. Soit $\int \sin^m x \cos^n x dx = I_{m,n}$.

Etablir la formule de réduction

$$(m+n)I_{m,n} = -\sin^{m-1} x \cos^{n+1} x + (m-1)I_{m-2,n} \\ \forall m, n \text{ sauf si } m = -n = 1.$$

Application: Calculer $I_{3,2} = \int \sin^3 x \cos^2 x dx$.

6. Démontrer la forme de récurrence

$$n I_n = -\sin^{n-1} x \cos x + (n-1) I_{n-2}$$

avec

$$I_n = \int \sin^n x dx \quad (n \neq -1, +1).$$

Application: $\int \sin^7 x dx$.

7. Démontrer

$$\int_0^{\pi/2} \sin^n x dx = \begin{cases} \frac{(n-1)!!}{n!!} & \text{si } n \text{ est un entier positif impair} \\ \frac{(n-1)!!}{n!!} \frac{\pi}{2} & \text{si } n \text{ est un entier positif pair} \end{cases}$$

$$n!! = n(n-2)(n-4) \dots$$

Enunciados de F. et J. Teixeira

BOLETIM BIBLIOGRÁFICO

Nesta secção, além de extractos de críticas aparecidas em revistas estrangeiras, serão publicadas críticas de livros e outras publicações de Matemática de que os Autores ou Editores enviarem dois exemplares à Redacção.

185 — NICOLAE POPESCU — *Categorii Abeliene* — Editura Academiei Republicii Socialiste România — Bucuresti, 1971.

A teoria das categorias abelianas tem-se desenvolvido nos últimos vinte anos e pelos seus métodos e resultados desempenha um papel de relevo na Matemática abstracta. Tem origem na Algebra, particularmente na teoria dos módulos sobre um anel e tem sido desenvolvida pela necessidade de se obter o quadro natural para muitas noções algébricas. Do mesmo modo, a teoria das categorias abelianas tem aplicações importantes na teoria dos anéis, na geometria algébrica, na geometria analítica, na topologia algébrica, etc.

O presente trabalho é uma monografia que aborda problemas importantes da teoria das categorias abelianas.

O primeiro capítulo expõe (sem demonstrações) noções e resultados fundamentais da teoria das categorias.

Ao segundo capítulo descrevem-se as noções fundamentais sobre as categorias pré-aditivas, aditivas pré-abelianas e abelianas, demonstrando os resultados clássicos sobre as categorias abelianas, os produtos e somas fibradas teoremas de isomorfismo, somas directas. No parágrafo oitavo, o mais importante do capítulo, o Autor ocupa-se dos limites indutivos e projectivos nas categorias abelianas, demonstra a equivalência das quatro formas da condição Ab5 de GROTHENDIECK e apresenta consequências.

O terceiro capítulo, «Functores aditivos» contém o estudo dos funtores aditivos, categorias dos funtores aditivos, categorias de módulos, caracterização de categorias de funtores aditivos e categorias de módulos o teorema do produto tensorial, a exactidão dos

functores, objectos projectivos e injectivos, envolvente injectiva, a existência de envolventes injectivos em algumas categorias.

O quarto capítulo intitulado «A localização» é o mais importante de todos, desenvolve a teoria geral da localização que desempenha grande papel na Álgebra na Geometria algébrica, etc. Inicia-se pela teoria das categorias das fracções aditivas que é uma generalização bastante abstracta das fracções ordinárias, destacando-se o teorema 2.4 que fornece a técnica das construções na teoria da localização: as aplicações aqui expressas são a construção da categoria espectral duma categoria abeliana com complementar e a construção da categoria cociente duma categoria abeliana por uma subcategoria espessa. No parágrafo sexto estuda-se uma forma particular da categoria cociente para a qual o functor canónico admite um adjunto à direita, iniciando assim no livro o estudo das aplicações da teoria da localização. Segue-se o estudo das sub-categorias localizantes e da importância dos objectos injectivos na teoria da localização. Dá-se grande relevo à localização nas categorias dos módulos a saber a construção das sub-categorias localizantes utilizando a noção de sistema localizante (GABRIEL), estuda-se a construção do localisado de um módulo em relação a uma subcategoria localizante e as suas consequências importantes, parágrafos 10 a 12. A aplicação da teoria da localização faz-se ainda na caracterização das categorias abelianas com geradores e limites indutivos exactos, localização nas categorias de functores aditivos e aplicação na construção de functores exactos à esquerda e a aplicação no teorema de imersão MITCHELL. Continua com o estudo de algumas localizações de uma categoria de módulos caracterizando os epimorfismos achatados (à esquerda) dos anéis e com aplicação ao estudo (à esquerda) de um anel. O capítulo termina pelo estudo de algumas localizações especiais e aplicações da localização na teoria dos feixes abelianos sobre um espaço topológico.

O quinto capítulo trata do teorema de KRULL-REMACK-SCHMIDT para as categorias abelianas e começa pelas formas clássicas deste teorema (existência e unicidade — AZUMAYA, АТЯН). Continua com o estudo das categorias espectrais e as l. c. categorias para as quais o teorema de KRULL-REMACK-SCHMIDT sob a forma GABRIEL é válido. Estuda as l. c. categorias particulares: categorias com a dimensão de KRULL-GABRIEL definida, as categorias semi-artinianas, as categorias localmente noetherianas, localmente artínianas e localmente finitas. Dá a caracterização das categorias localmente noetherianas e aplicações ao estudo dos anéis noetherianos (à esquerda). Termina por algumas considerações sobre uma forma geral do teorema de

KRULL-REMACK-SCHMIDT sobre uma categoria de GROTHENDIECK qualquer.

O sexto capítulo trata da «Teoria da decomposição» e dá a definição e as propriedades gerais duma teoria axiomática da decomposição que tem casos particulares da teoria da decomposição primária e terciária: Estuda a ligação entre estas teorias de decomposição e o anterior teorema de KRULL-REMACK-SCHMIDT.

O sétimo capítulo intitulado «A dualidade» tem por objectivos a construção do dual duma categoria de GROTHENDIECK que é uma categoria dos módulos topológicos sobre um anel topológico dum tipo particular.

Os temas abordados neste livro e a profundidade com que os mesmos são tratados reflectem bem o alto nível científico que a Academia das Ciências da República Popular da Roménia tem conseguido imprimir ao País no campo da matemática, promovendo não só o desenvolvimento desta ciência nas suas formas mais abstratas como nas aplicações mais concretas, como reforço de todas as actividades produtivas nacionais: agricultura, indústria, econometria, etc.

J. G. T.

186 — N. P. BHATIA e G. P. SZEGÖ — *Stability Theory of Dynamical Systems* — Die Grundlehren der mathematischen Wissenschaften — Band 161, Springer Verlag. Berlin-Heidelberg-New York.

A teoria da estabilidade dos sistemas dinâmicos pode dizer-se ter sido iniciada em termos de teoria de equações diferenciais ordinárias no século XIX com LIAPOUNOV e POINCARÉ.

Posteriormente e nas últimas décadas deve o seu maior incremento à escola soviética; LEFSCHETZ alertando os matemáticos dos USA do atrazo, chama a atenção sobre o interesse que a teoria apresenta na resolução de problemas que se colocam na produção de energia nuclear (década 50 e começo de 60).

As oito centenas de trabalhos citados nas referências bibliográficas do presente livro exemplificam e confirmam os aspectos modernos e as origens da teoria.

LIAPOUNOV define de uma maneira precisa os conceitos de estabilidade, estabilidade assintótica e instabilidade e fornece um «método» de análise das propriedades de estabilidade de uma determinada solução de uma equação diferencial ordinária. Tanto a definição como o método caracterizam em termos estritamente locais as propriedades de estabilidade de uma solução de equação diferencial. Pelo contrário a teoria de POINCARÉ difere da anterior na medida em

que as propriedades globais das equações diferenciais no plano desempenham o papel relevante.

Um dos aspectos básicos da teoria de POINCARÉ está na introdução do conceito de trajectória, isto é, uma curva no plano x, \dot{x} , paramétrica em relação ao tempo t , que pode obter-se pela eliminação da variável entre equações dadas, reduzindo assim a equação diferencial à ordem primeira, a respeito de x e \dot{x} . Nesta via POINCARÉ adopta um referencial no qual estuda o comportamento qualitativo das equações diferenciais. POINCARÉ não está interessado na integração de tipos particulares de equações, mas em classificar todos os comportamentos possíveis de todas as equações diferenciais de segunda ordem. Pela introdução do conceito de trajectória POINCARÉ está apto a formular e a resolver, como problemas topológicos, os problemas da teoria das equações diferenciais. Desta forma POINCARÉ abre o caminho para a noção abstracta de sistema dinâmico que pode ser atribuída essencialmente a A. A. MARKOV e H. WHITNEY. Estes dois autores observaram separadamente que é possível estudar a teoria qualitativa de famílias de curvas (trajectórias) num espaço X conveniente, contanto que estas famílias sejam de certa forma condicionadas no seu comportamento, isto é, se forem definidas como tendo sido geradas por um grupo de transformações topológicas a um parâmetro, actuando sobre X .

G. D. BIRKHOFF deu grande desenvolvimento à teoria dos sistemas dinâmicos podendo ser verdadeiramente considerado como o fundador dessa teoria. A sua célebre monografia de 1927 sobre os Sistemas Dinâmicos é a base de muitos trabalhos de investigação das décadas de 30 e 40, que hoje em dia não estão desactualizados.

BIRKHOFF traçou as duas vias fundamentais na teoria dos sistemas dinâmicos, a saber a teoria topológica e a teoria ergódica.

Em 1947 V. V. NEMYTSKII e V. V. STEPANOV publicaram a «Teoria Qualitativa das Equações Diferenciais» que até hoje tem servido de referência de base aos mais importantes trabalhos da teoria dos sistemas dinâmicos. Em 1949 NEMYTSKII escreveu um trabalho que faz o ponto sobre os problemas topológicos na teoria dos sistemas dinâmicos e reúne todos os resultados da teoria topológica até o fim da década de 40.

Durante os anos de 50 faz-se um grande esforço na generalização do conceito de um sistema dinâmico em relação aos grupos transformações topológicas e muitos autores aparecem com obras notáveis.

Mais recentemente a teoria foi ampliada com a introdução de novos problemas considerados à LIAPOUNOV, ausente nos primeiros trabalhos sobre sistemas dinâmicos e grupos de transformação topológicos.

Neste sentido, a obra de TARO URA, em particular a sua teoria dos prolongamentos e as suas relações com a estabilidade mostrou claramente que a parte significativa da teoria da estabilidade é de natureza topológica e pertence à concepção básica da teoria dos sistemas dinâmicos. Os trabalhos de V. I. ZUBOV contribuem bastante para esta extensão da teoria com a introdução dos espaços métricos.

O presente volume foi concebido para apresentar de uma forma simples os resultados recentes da teoria dos sistemas dinâmicos nos espaços métricos e especialmente a teoria da estabilidade nas suas aplicações concretas à teoria das equações diferenciais. Não se refere a vários capítulos recentemente criados como a teoria da estabilidade estrutural, a teoria ergódica e a teoria geral dos grupos de transformações topológicas. Mantém-se um nível de acessibilidade aos estudantes universitários dos médios e e últimos anos (com conhecimentos de espaços métricos e de equações diferenciais) para o que não se faz referência aos sistemas dinâmicos locais.

Os capítulos I a VII contêm a teoria básica dos sistemas dinâmicos nos espaços métricos e os Capítulos VIII e IX contêm aplicações e extensões da teoria da estabilidade (Cap. V) aos sistemas dinâmicos definidos por equações diferenciais ordinárias.

Mais pormenorizadamente, o Capítulo I contém a definição de um sistema dinâmico e alguns exemplos que indicam vários campos de aplicação.

O Capítulo II contém noções elementares invariantes sob certas transformações topológicas dos sistemas dinâmicos. O Capítulo III trata principalmente dos conjuntos mínimos e sua estrutura. O Capítulo IV dedica-se ao estudo dos sistemas dinâmicos dispersivos e paralelizáveis e termina a teoria básica exposta. O Capítulo V desenvolve o tema principal do livro, a saber a teoria de estabilidade e da atracção (a teoria da atracção aqui apresentada difere essencialmente da de ZUBOV que assenta no conceito de atracção fraca). O Capítulo VI dedica-se a um problema mais específico: a classificação dos escoamentos nas proximidades de um conjunto compacto invariante. O Capítulo VII contém a teoria dos prolongamentos de T. URA com aplicação à estabilidade absoluta e recorrência generalizada. O Capítulo VIII trata da teoria geométrica da estabilidade para as equações diferenciais autónomas ordinárias incluindo várias extensões do método directo de LIAPOUNOV. O Capítulo IX dedica-se ainda a um problema mais específico dos conceitos de atracção e estabilidade via funções de LIAPOUNOV não contínuas.

187 — JEAN PIERRE KAHANE — *Séries de Fourier absolument convergentes — Ergebnisse der Mathematik und ihrer Grenzgebiete* — Springer-Verlag Berlin, 1970.

«Este livro tem por objectivo o estudo das funções da classe A , isto é, as funções contínuas sobre o círculo cuja série de FOURIER é absolutamente convergente. O pertencer à classe A é uma propriedade local. A teoria descritiva consiste em comparar esta propriedade local a outras propriedades (interessando o módulo de continuidade, por exemplo). É esta a mais antiga orientação, definida neste campo por S. BERNSTEIN. Os resultados recentes dizem respeito sobretudo as restrições das funções da classe A a conjuntos fechados E do círculo; estas restrições definem uma classe de funções $A(E)$. Para outros conjuntos fechados, E , do círculo, toda a função sobre E pertence a $A(E)$ e diz-se que E é um conjunto de HELSON.

Por outro lado, A é uma álgebra de BANACH (a norma de uma função é a soma dos módulos dos coeficientes de FOURIER). Esta estrutura descoberta por WIENER, sugere grande número de problemas, entre os quais o de considerar a classe $A(E)$ como álgebras cocientes de A

$$A(E) = A/I_E$$

em que I_E é o ideal fechado de A constituído pelas funções de A que se anulam sobre E .

Nestes termos levantam-se as seguintes questões:

- 1 — Quais são os endomorfismos de A ?
- 2 — Quais as funções que operam em A , isto é, as funções F (definidas, por exemplo, sobre um intervalo real I) tais que, para toda f de A com valor em I se tenha $F(f) \in A$?
- 3 — Quais são os ideais fechados de A ?
- 4 — Quais as sub-álgebras fechadas de A ?
- 5 — Sendo dada uma função f de A , quais são as funções F , definidas sobre o conjunto dos valores de f , tais que $F(f) \in A$?

As questões 1 e 2 foram levantadas em 1934 por P. LEVY e foram resolvidas completamente em 1953 (BEURLING e HELSON) e 1958 (KATZNELSON). Um caso particular da questão 3 é o seguinte problema que, por razões históricas, é designado «problema de síntese espectral»: Todo o ideal fechado de A é um I_E ? Deve-se a resposta negativa a MALLIAVIN (1959). As questões 4 e 5 estão de certa forma relacionadas e esta última foi apresentada sob a forma expressa por MALLIAVIN também em 1959 sob o nome de «cálculo simbólico individual»: Em 1939 porém MARCINKIEWICZ tinha já trabalhado na mesma direcção. Apesar das ligações evidentes entre as questões 5 e 2 foi apenas recentemente (1966-67) que o cálculo simbó-

lico individual permitiu reencontrar o teorema de KATZNELSON.

Todos estes problemas postos em relação a A , transpõem-se para $A(E)$. Melhor, verifica-se que, excepção do primeiro, a melhor via de acesso para o estudo em A consiste no seu estudo em certos $A(E)$. O poderoso método das álgebras tensoriais introduzido por VAROPOULOS (1965-1967) consiste em identificar certos $A(E)$ como álgebras $V(D)$ de funções definidas sobre o quadrado D^2 do grupo compacto $D = \{ |z| = 1 \}^N$ (onde a adição é definida como a multiplicação das coordenadas). Por seu turno identifica-se uma sub-álgebra fechada da $V(D)$ a $A(D)$, álgebra das séries de FOURIER absolutamente convergentes sobre o grupo D (e não mais sobre o círculo!) $A(D)$ pode ainda interpretar-se como a álgebra das séries de FOURIER-WALSH absolutamente convergentes.

Muitas questões que se põem sobre os $A(E)$ ficam em aberto. Em particular, mal se conhece a significação do isomorfismo de duas álgebras $A(E)$ e $A(E')$.

JEAN PIERRE KAHANE, é um dos matemáticos da actualidade mais qualificados para expôr as orientações e as vias que se apresentam no domínio da matemática que, na forma anterior, é tratado na introdução do livro de que é Autor. A sua competência permite-lhe não só ser colaborador duma das mais categorizadas colecções de obras de matemática — *Ergebnisse der Mathematik und ihrer Grenzgebiete* — mas ainda nela fazer ressaltar os aspectos mais salientes duma teoria com mais de dois séculos de existência iniciada na célebre Teoria Analítica do Calor. A longa bibliografia exposta, de 172 referências, varre completamente todo o espectro das tendências manifestadas e dos resultados obtidos. É uma obra de alto nível científico. J. G. T.

Notícias críticas próximas sobre as seguintes obras:

- A. F. MONNA — *Analyse non-Archimédienne* — Band 56.
 LEOPOLDO NACHBIN — *Topology on spaces of Holomorphic Mappings* — Band 47.
 CARLO MIRANDA — *Partial differential Equations of Elliptic Type* — Band 2.
 H. BUSEMANN — *Recent synthetic Differential Geometry* — Band 54.
 H. S. WILF — *Finite Sections of Some Classical Inequalities* — Band 52.
 S. LOPES DE MEDRANO — *Involutions on Manifolds* — Band 59.
 Volumes da col. «*Ergebnisse der Mathematik*» und «*Ihrer Grenzgebiete*».
 J. STOER-C. WITZGALL — *Convexity and optimization in Finite Dimensions I* — Band 163.
 A. GROTHENDIECK-J. A. DIEUDONNÉ — *Elements de Géométrie Algébrique I* — Band 166.
Grundlehren der mathematischen Wissenschaften in Enizelardarstellungen.

Panorama da Geodesia Contemporânea (*)

por F. Teixeira de Queiroz

1. O problema geodésico.

Como acontece com toda a ciência em rápido desenvolvimento, é extremamente difícil a apresentação dum panorâmica da geodesia contemporânea porque, não só alguns dos factos expostos podem à altura da publicação estar já desactualizados, mas também, e principalmente, porque a exposição dum ciência em tais condições está muito mais marcada do que qualquer outra por influências subjectivas, que sempre existem, daquele que a expõe. A qualquer desses perigos nos expomos ao apresentar esta panorâmica dum ciência que, como tantas outras, adquiriu depois da segunda guerra mundial um desenvolvimento que se pode classificar de explosivo.

Deve-se um tal desenvolvimento a quatro factores fundamentais:

1) Uma melhor colaboração entre as nações a respeito dos problemas geodésicos.

2) O aparecimento de máquinas de calcular electrónicas que permitem a resolução de sistemas de equações lineares a um muito grande número de incógnitas.

3) O aparecimento de instrumentos electrónicos medidores de distâncias.

4) A colocação em órbita de satélites artificiais.

Parece-nos útil, porém, antes de abordarmos a influência que cada um destes pontos tem em geodesia, fazer um apanhado dos objectivos e métodos da geodesia clássica pois, só assim, será possível inserir e fazer ressaltar o papel que cada um dos factores acima assinalados desempenha no desenvolvimento e renovação desta ciência.

*

A geodesia é uma ciência que tem como objectivo a determinação das dimensões e da forma da Terra. É uma das ciências mais antigas pois, já na Grécia clássica, AERATÓSTENES realizou uma determinação do raio da Terra.

Contudo, é por alturas da revolução francesa que a geodesia principia a adquirir o aspecto que tem hoje. Nessa altura, a observação dum cadeia de triângulos centrada em Paris e estendendo-se para Norte e Sul da Europa, levou alguns geodetas a formularem a hipótese de ser a Terra sensivelmente um elipsóide alongado de revolução (elipsóide de CASSINIS). Tal hipótese deu origem a uma grande controvérsia pois estava em contra-dição com a teoria de NEWTON e HUYGENS a qual conduzia a admitir ser um elipsóide achatado de revolução o que melhor se adaptava à forma da Terra. A fim de decidir

(*) O presente artigo consta de dois extractos, a Introdução e o Capítulo V, do trabalho apresentado pelo Autor ao Concurso organizado pela «Gazeta de Matemática» com o apoio da Fundação Calouste Gulbenkian e que mereceu o 2.º prémio da Secção B.

qual o elipsóide que melhor se adaptava à Terra, a Academia das Ciências de Paris mandou, em 1735, duas expedições para regiões de latitude elevada com o objectivo de aí medirem arcos que se podessem comparar com o da Europa Central. Uma das expedições, enviada para a Patagónia, depois de inúmeras peripécias, ficou impossibilitada de chegar a resultados conclusivos. A outra, enviada para a planície de Lapland, pôde provar ter a Terra uma forma mais próxima dum] elipsóide achatado do que dum alongado.

Com a controvérsia acima referida ficou posto um dos problemas fundamentais da geodesia: a determinação do elipsóide que melhor se ajusta à superfície da Terra.

Este problema é duma importância capital pois é sobre uma tal superfície que, na geodesia clássica, se efectuam os cálculos das observações geodésicas. Portanto, estes conduzem a resultados diferentes conforme o elipsóide escolhido, afastando-se tanto mais dos valores observáveis quanto mais diferir a superfície terrestre desse elipsóide. Evidentemente não se pode pensar que toda a massa de observações, feita ou a fazer, possa ser referida automaticamente a um elipsóide satisfatório. Escolhido um, ele é conservado durante o maior tempo possível, já que a mudança de milhares de dados observados, dum elipsóide para outro é uma operação muito dispendiosa. Além disso, não sendo a Terra uma superfície regular, pode haver um elipsóide que se ajuste melhor a uma dada região e que não seja satisfatório para outra.

Fixado um elipsóide de referência, todo o ponto do espaço pode ser determinado pela normal ao elipsóide que passa por ele e pela distância do ponto (afectada dum sinal) a esse elipsóide. Em particular, podem ser referidos a essa superfície todos os pontos da superfície terrestre. Na prática, é porém impossível, duma maneira simples, determinar

num ponto da superfície terrestre, a normal a um elipsóide dado. Contudo, sempre que o elipsóide escolhido seja uma boa aproximação da superfície terrestre, a vertical desse lugar é uma linha vizinha da normal referida. É por isso que todas as observações geodésicas são baseadas nessa vertical.

Assim, nas observações astronómicas determinam-se os parâmetros da vertical que passa pelo lugar de observação.

Na triangulação geodésica medem-se rectilíneos de diedros cujas arestas são tangentes a verticais.

No nivelamento geométrico determina-se a equidistância dos planos horizontais (Planos perpendiculares à referida vertical) que passam em lugares vizinhos.

Finalmente nas medições da gravidade determina-se qual a superfície equipotencial (superfície normal às verticais) que passa por um lugar dado.

Desta forma, nas observações geodésicas que referimos, ressalta a importância que tem o campo da gravidade terrestre, já que a vertical dum lugar não é senão a tangente, nesse lugar, à linha de corrente desse campo que passa por ele.

Aparece-nos assim, como segundo objectivo da geodesia, o estudo do campo da gravidade terrestre.

Da determinação das diferentes superfícies equipotenciais e das linhas de corrente desse campo deduz-se uma superfície (a superfície equipotencial de nível zero) que é uma melhor aproximação da superfície terrestre do que o elipsóide. Ela será a superfície que contém a superfície de equilíbrio dos oceanos. Em geodesia dá-se o nome de geoide a uma tal superfície.

Ficam assim definidas em geodesia três superfícies: a superfície terrestre, o geoide e o elipsóide de referência. A geodesia clássica desenvolve-se usando-as alternadamente: as observações são feitas sobre a primeira em relação à segunda e calculadas

sobre a terceira. Das vantagens e inconvenientes dum tal método teremos ocasião de falar no decorrer deste trabalho.

2. A Associação Internacional de Geodesia.

O estudo da figura da Terra não se pode compreender sem observações que abranjam, senão a totalidade, pelo menos uma grande extensão da superfície terrestre. Para a realização de tais observações é fundamental um perfeito acordo entre as nações, não só para que aquelas tenham sensivelmente a mesma precisão mas, principalmente, para que se possam interligar.

O trabalho de coordenação e de sistematização das observações, bem como a compilação dos estudos acerca da forma da Terra, tem sido realizado pela Associação Internacional de Geodesia (AIG).

A AIG que tem hoje, praticamente, representações de todas as nações completou há poucos anos o seu primeiro centenário. Criada com o nome de Associação Geodésica Internacional, funcionou desde 1862 até 1914, altura em que a guerra interrompeu a sua actividade. Tornada a criar em 1922, já com o nome de AIG, a sua actividade tem-se mantido até à actualidade.

Apraz-nos salientar que o bom entendimento entre as secções das diferentes nações permitiu colocar ao serviço da AIG assuntos tais como as coordenadas geográficas de pontos bem definidos, os quais até há bem pouco tempo eram classificados como segredo militar.

A AIG foi primitivamente criada para interligar os países do centro da Europa mas tornou-se rapidamente uma organização à escala planetária, podendo hoje reenviçar com orgulho o título da mais antiga organização científica mundial.

A sua actividade foi estruturada numa sequência de congressos:

Em Berlim nos anos de 1864 e de 1867, em Viena (1871), em Dresde (1870), em Stuttgart (1877), em Munique (1880), em Roma (1883), novamente em Berlim (1886), em Paris (1889), em Copenhague (1903), em Budapeste (1906), em Cambridge (1909). Com esta reunião terminou a primeira fase da vida desta organização.

Se quiséssemos apontar os nomes de alguns geodetas que se distinguiram durante este período três aparecem imediatamente destacados: GAUSS, BESSEL e HELMERT. Porém, numa organização deste tipo, é o trabalho contínuo de observação, sistematização e cálculo que mais convém destacar.

Durante o primeiro período de actividade da organização pode-se apresentar como obra realizada:

A Convenção Internacional do Metro que permitiu uniformizar as unidades de comprimento.

O Serviço Internacional de Latitudes ao qual está entregue a conservação da hora.

A repetição, com precisão muito maior, da observação dos arcos em regiões de latitude elevada, a que já fizemos referência no início deste trabalho.

A medição de diversos arcos de paralelos e de meridianos.

Determinações de desvios da vertical, pesquisas sobre isostasia e sobre marés terrestres, etc.

A seguir à guerra de 1914-1918 voltou a pôr-se a necessidade da Instituição. A mesma foi integrada numa União Geodésica e Geofísica Internacional da qual passou a ser uma das Associações.

Realizaram-se as Assembleias Gerais de Roma (1922), Madrid (1924), Praga (1927), Estocolmo (1930), Lisboa (1933), Edinburgo (1936), Washington (1939), e a seguir à segunda guerra mundial, Oslo (1948), Bruxelas (1951), Roma (1954), Toronto (1957), Helsinquia (1960) e Berkeley (1964).

A estrutura da AIG tem evoluído sendo

actualmente formada por cinco secções: determinação geométrica de posições, nivelamento e movimentos dos solos, astronomia geodésica e satélites artificiais, gravimetria e geodesia física.

Para dar uma primeira ideia da problemática da geodesia contemporânea, damos a seguir um apanhado dos principais problemas que são o objecto de cada uma destas secções:

I Secção: Triangulações e trilaterações, poligonais de alta precisão, bases, medições electromagnéticas e ópticas de distâncias, nivelamento trigonométrico, refacção atmosférica, cálculo.

II Secção: Altitudes, definições, nivelamento de precisão, métodos e instrumentos, resultados, precisões e cálculos, movimentos horizontais e verticais do solo.

III Secção. Métodos de observação em astronomia geodésica, variação das latitudes e das longitudes, problemas da hora, utilização geométrica dos satélites artificiais.

IV Secção: Medições absolutas e relativas da gravidade, estabelecimento duma rede gravimétrica mundial homogénia, bases de calibração, estabelecimento de cartas de anomalias da gravidade, utilização prática dos métodos de redução.

V Secção: Desvios da vertical, reduções da gravidade, interpretação física das anomalias da gravidade, determinação do potencial exterior e do campo gravítico terrestre, utilização dinâmica dos satélites artificiais, marés terrestres.

A estruturação da AIG em cinco secções não é suficiente para abordar os diferentes problemas da geodesia moderna. Por isso, sempre que um problema o justifique, é criada uma Comissão Permanente ou, pelo tempo que for julgado necessário, um Grupo Especial de Estudo (que pode eventualmente ser desdobrado em subgrupos) a que ficam ades-

trictos os membros da AIG a que ele interesse. O problema é estudado por esse grupo duma forma permanente e, quando se justifique, em sessões de trabalho.

A seguir ao congresso de Berkeley as comissões permanentes ficaram a ser:

1.^a Comissão: Nova compensação de conjunto das triangulações europeias. 2.^a Comissão: Estabelecimento duma rede europeia de nivelamento unificada (REUN). 3.^a Comissão: comissão gravimétrica internacional. 4.^a Comissão: marés terrestres. 5.^a Comissão: bibliografia geodésica. 6.^a Comissão: movimentos recentes da crosta terrestre. 7.^a Comissão: satélites artificiais.

Quanto aos grupos de estudo foram criados os seguintes (indicamos entre parêntesis a secção a que o grupo diz respeito):

GEE 1 (I) Problemas teóricos respeitantes ao cálculo e à compensação das grandes triangulações. GEE 2 (I) Cálculo das redes de SHORAN (e de processos análogos) e das grandes geodésicas. GEE 3 (II) Estudo dos erros sistemáticos dos nivelamentos de precisão por métodos de análise estatística. GEE 4 (III) Estudo crítico dos métodos de astronomia geodésica. GEE 5₁ (IV) Determinações absolutas da gravidade. Ligações entre estações absolutas. Fórmula da gravidade normal. GEE 5₂ Técnica das observações gravimétricas. GEE 6 (IV) Estabelecimento duma base europeia de calibração de gravímetros. Bases de calibração GEE 7 (IV) Sobre a oportunidade da adopção dum sistema único de decomposição em zonas para o cálculo das reduções topográficas (e isostáticas) e para o cálculo dos desvios absolutos da vertical. GEE 8 (V) Reduções das observações gravimétricas com vista à sua utilização na fórmula de STOKES. GEE 9 (III) Emprego em geodesia dos resultados das observações lunares e de satélites artificiais

(eclipses, ocultações, fotografias sobre um campo estelar, etc.). GEE 10 (V) Determinação do geóide europeu pelos desvios relativos da vertical. GEE 11 (V) Interpretação geofísica das anomalias da gravidade. GEE 12 (I) Utilização nos diferentes domínios da geodesia das modernas máquinas de cálculo (máquinas de cartões perfurados, de fita magnética, etc.). GEE 13 (I) Determinação dos movimentos da crosta terrestre (exceptuando o das marés) por todas as medidas susceptíveis de serem realizadas à superfície da Terra. GEE 14 (I) Especificações das precisões requeridas para as triangulações e as trilaterações. Pontos de LAPLACE. Emprego combinado de trilaterações e triangulações. GEE 15 (V) Estudo das secções do geóide que atravessam largas regiões do mar (Creta, Egipto, Canadá, etc.). GEE 16 (V) Estudo prático sobre a aplicação à fórmula de STOKES das anomalias definidas pela teoria de J. DE GRAAF HUNTER. GEE 17 (IV) Resultados obtidos por observação de satélites artificiais acerca do campo da gravidade terrestre. GEE 18 (IV) Métodos para a determinação absoluta da gravidade. GEE 19 (I) Medições electrónicas de distâncias. GEE 20 (IV) Medições da gravidade no mar (em submarinos e em navios). GEE 21 Cálculos numéricos com interesse para as grandes triangulações. GEE 22 (II) Estudo do nível médio dos mares. GEE 23 Estudo da refacção atmosférica. GEE 24 Notações e nomenclatura a usar em geodesia. GEE 25 (III) Observações ópticas de satélites artificiais. GEE 26 (III) Ligações geodésicas por meio de satélites artificiais. GEE 27 (III) Navegação por meio de satélites artificiais. GEE 28 (V) Dinâmica dos satélites artificiais.

Demos uma tão extensa lista de agrupamentos da AIG por nos parecer que assim destacávamos melhor um certo número de problemas da geodesia moderna. Teremos

o ocasião de nos referir a alguns particularmente.

A Associação Internacional de Geodesia publica uma revista, o «Bulletin Géodésique», ao qual fomos buscar os elementos relativos a esta Associação.

*

Redigidas estas notas em 1965, são publicadas em 1970. Como não notar desactualizações? As palavras com que abrimos estas notas confirmaram-se plenamente. Contudo, quere-nos parecer que o contexto que serviu de base à redacção desta panorâmica da geodesia ainda se mantém o que justifica a sua publicação. Ao fazê-lo, importa assinalar os principais factos que introduzem desactualizações e os principais pontos onde uma influência subjectiva marcada, pode ter conduzido a um erro de apreciação.

Assim, importa assinalar os seguintes factos ocorridos durante este período :

a) A conclusão dos projectos de observação baseados nos satélites ECOS, GEOS e PA-GEOS.

b) A introdução dum novo elipsóide internacional de referência: o IAU, 1963.

c) O início dum novo capítulo da geodesia: o estudo da forma da Lua. A este novo capítulo deu-se o nome de selenodesia (*).

d) O início e possível conclusão do projecto SECOR.

Como influência subjectiva, assinalamos a importância excessiva que demos às observa-

(*) Quere-nos parecer que o nome de geodesia lunar teria sido mais apropriado. O que realmente se está a passar neste momento é o facto do objecto da geodesia estar a sofrer uma mutação: de ciência que estuda a forma da Terra tende a transformar-se na ciência que estuda a forma dum Planeta qualquer. Onde se justifica a conservação do nome da ciência mas não o seu conteúdo. Tal como aconteceu à geometria.

ções ópticas de satélites em detrimento das observações baseadas no efeito DOPPLER-FIZEAU. O segundo tipo de observações revela-se de momento mais preciso do que o primeiro.

Como resultado de tal precisão, a geodesia por satélites encontra-se de momento diferenciada em dois sectores. Um, o da «geodesia rica» utiliza satélites activos do tipo GEOS e SECOR. É a geodesia dos Estados Unidos e da União Soviética. O outro, o da «geodesia pobre», utiliza satélites passivos do tipo ECO, PAGEOS ou Diamante. É a geodesia da Europa.

ELEMENTOS DE GEODESIA DINÂMICA

1. Introdução.

Mostramos nos dois últimos capítulos a importância que tinha em geodesia a determinação do potencial do campo da gravidade. Essa determinação fazia-se decompondo o campo da gravidade num campo de gravidade normal e num campo de perturbação. Ao primeiro correspondia um potencial de gravidade normal e ao segundo um potencial de perturbação, T .

A geodesia dinâmica tem por objectivo estudar o campo de gravitação terrestre. Para isso, decompõe este num campo de gravitação normal e num campo de perturbação. Para campo de gravitação normal toma-se um campo de simetria esférica. Em tais condições, o potencial de perturbação é o mesmo usado em geodesia física quando se toma como superfície de referência uma esfera.

Se desenvolvermos o potencial de perturbação em série de LEGENDREANOS, tal como o fizemos nos capítulos anteriores, o problema central da geodesia dinâmica será o da deter-

minação dos coeficientes J_n , J_{nm} , e K_{nm} : Essa determinação faz-se pela análise das perturbações causadas no movimento dos satélites artificiais devido à existência do potencial de perturbação.

Torna-se então necessário separar as perturbações causadas pelo campo de perturbação daquelas devidas a diversos efeitos parasitários. Destes, os mais importantes são o atrito da atmosfera, a atracção do Sol e da Lua e a pressão da radiação solar. Destas perturbações parasitárias as de mais difícil controle são a pressão de radiação e o atrito atmosférico pois dependem grandemente da actividade solar. A fim de minimizar os seus efeitos, são escolhidos satélites cujas órbitas tenham um perigeu não muito pequeno e um apogeu não muito grande, pois a acção da atmosfera faz-se sentir quando o satélite está mais próximo da Terra e a perturbação causada pela pressão de radiação é tanto maior quanto mais afastado está o satélite da Terra. Além disso escolhem-se para estudo satélites pequenos e pesados.

A órbita é escolhida em conformidade com a perturbação a analisar. Para isso procura-se uma órbita que torne máximas as perturbações de interesse geodésico e mínimas as perturbações parasitárias. Em geral, uma dada perturbação não é estudada só relativamente a uma dada órbita. Procuram-se observar satélites em diversas órbitas, com diferentes inclinações (afastadas em geral da inclinação crítica), para se obter uma determinação suficientemente precisa. Um método a usar é o da escolha de satélites onde os harmónicos em estudo criem fenómenos de ressonância.

2. As equações do movimento dum satélite.

As equações do movimento dum satélite artificial sob a acção dum campo criado por um potencial da forma

$$(1) \quad W = \frac{fM}{r} + T = \frac{fM}{r} \cdot \left(1 + \sum \frac{J_n P_n + \Sigma (J_{nm} \cos m\lambda + K_{nm} \sin m\lambda) P_{nm}}{r^n} \right)$$

são mais facilmente deduzidas se utilizarmos um sistema de coordenadas esféricas. Com efeito, consideremos um sistema de coordenadas esféricas, com origem no centro de massas do corpo atraente e introduzidas por meio de

$$\begin{aligned} x &= r \cos \theta \cos \eta \\ y &= r \cos \theta \sin \eta \\ z &= r \sin \theta \end{aligned}$$

Será então

$$\begin{aligned} ds^2 &= dx^2 + dy^2 + dz^2 = \\ &= dr^2 + r^2 d\theta^2 + r^2 \cos^2 \theta d\eta^2 \end{aligned}$$

e a força viva do satélite, que supomos de massa m , será

$$(2) \quad 2T = m(\dot{r}^2 + r^2 \dot{\theta}^2 + r^2 \cos^2 \theta \dot{\eta}^2).$$

Com auxílio das equações de LAGRANGE obtemos para componentes da aceleração da partícula.

$$\begin{aligned} \gamma_r &= \frac{1}{m} \left(\frac{d}{dt} \frac{\partial T}{\partial \dot{r}} - \frac{\partial T}{\partial r} \right) = \frac{1}{m} Q_r \\ \gamma_\theta &= \frac{1}{m} \left(\frac{d}{dt} \frac{\partial T}{\partial \dot{\theta}} - \frac{\partial T}{\partial \theta} \right) = \frac{1}{m r} Q_\theta \\ \gamma_\eta &= \frac{1}{m} \left(\frac{d}{dt} \frac{\partial T}{\partial \dot{\eta}} - \frac{\partial T}{\partial \eta} \right) = \frac{1}{m r \cos \theta} Q_\eta \end{aligned}$$

onde as segundas igualdades são uma consequência de ser o trabalho realizado pelo campo, quando o satélite sofre um deslocamento dP , igual a

$$\begin{aligned} dW &= Q_i dP^i = Q_r dr + Q_\theta d\theta + Q_\eta d\eta = \\ &= m \gamma^i dP_i = m(\gamma^r dr + \gamma^\theta r d\theta + \gamma^\eta r \cos \theta d\eta). \end{aligned}$$

Com a força viva dada por (2) e o potencial da intensidade de campo dado por (1), as equações do movimento podem ser escritas

$$\begin{aligned} (3) \quad \gamma^r &= r'' - r \theta'^2 - r \cos^2 \theta \eta'^2 = \frac{\partial W}{\partial r} \\ \gamma^\theta &= \frac{1}{r} (r^2 \theta')' - r \cos \theta \sin \theta \eta'^2 = \frac{1}{r} \frac{\partial W}{\partial \theta} \\ \gamma^\eta &= \frac{1}{r \cos \theta} (r^2 \cos^2 \theta \eta')' = \frac{1}{r \cos \theta} \frac{\partial W}{\partial \eta}. \end{aligned}$$

A decomposição do potencial, no potencial de gravitação normal e no potencial de perturbação T , permite-nos escrever, em vez de (3), as equações com a forma

$$\begin{aligned} r'' - r \theta'^2 - r \cos^2 \theta \eta'^2 + \frac{fM}{r^2} &= \frac{\partial T}{\partial r} \\ \frac{1}{r} (r^2 \theta')' + r \cos \theta \sin \theta \eta'^2 &= \frac{1}{r} \frac{\partial T}{\partial \theta} \quad (I) \\ \frac{1}{r \cos \theta} (r^2 \cos^2 \theta \eta')' &= \frac{1}{r \cos \theta} \frac{\partial T}{\partial \eta}. \end{aligned}$$

3. O problema dos dois corpos.

Se anularmos os segundos membros do sistema de equações I, obtemos as equações do movimento dum satélite sob a acção do campo de gravitação normal escolhido. É o conhecido problema dos dois corpos, visto o anulamento do segundo membro equivaler a admitir que o corpo atraente tem simetria esférica. As equações I tomam a forma

$$\begin{aligned} r'' - r \theta'^2 - r \cos^2 \theta \eta'^2 + \frac{fM}{r} &= 0 \\ \frac{1}{r} (r^2 \theta')' + r \cos \theta \sin \theta \eta'^2 &= 0 \quad (II) \\ \frac{1}{r \cos \theta} (r^2 \cos^2 \theta \eta')' &= 0. \end{aligned}$$

Adoptamos, ao estabelecer o sistema de equações, um sistema de coordenadas esférico cujo plano de referência era arbitrário. Vamos agora admitir que esse plano contém no instante t_0 , não só o ponto material, como ainda é paralelo à velocidade dele no mesmo instante. Por outras palavras, vamos admitir que no instante t_0 é $\eta = \eta' = 0$. Em tais condições, a última das equações II admite o integral primeiro $r^2 \cos^2 \theta \eta' = k$ que se reduz a $\eta' = 0$. Com uma segunda integração resultará $\eta = k_1 = 0$. Vemos que a trajectória do ponto material é uma curva plana que está agora contida no plano de referência. Em tais condições o sistema II toma a forma mais simples

$$(2) \quad r'' - r \theta'^2 + \frac{fM}{r^2} = 0$$

$$(r^2 \theta')' = 0.$$

A segunda equação dá origem ao integral primeiro

$$(3) \quad \frac{1}{2} r^2 \theta' = k$$

que exprime que a velocidade areolar é constante. A eliminação do tempo entre a primeira das equações (2) e (3) dá-nos uma nova equação que determina a trajectória. Seja $1/r = u = f(\theta)$ a sua solução. Será então

$$\frac{dr}{dt} = \frac{dr}{du} \cdot \frac{du}{d\theta} \theta' = -\frac{1}{u^2} \cdot \frac{2k}{r^2} \cdot \frac{du}{d\theta} =$$

$$= -2k \frac{du}{d\theta}$$

$$\frac{d^2 r}{dt^2} = -2k \frac{d^2 u}{d\theta^2} \cdot \theta' = -4k^2 u^2 \frac{d^2 u}{d\theta^2}$$

e, portanto, essa equação escrever-se-á

$$(4) \quad \frac{d^2 u}{d\theta^2} + u - m = 0$$

em que $m = \frac{fM}{4k^2}$. Esta equação tem uma solução da forma $u - m = k \cos(\theta - \theta_0)$ o que mostra que a trajectória, visto não ter pontos impróprios, é uma elipse. Escrevemos a solução de (4) com a forma

$$(5) \quad \frac{1}{r} = \frac{1 + e \cos(\theta - \theta_0)}{a(1 - e^2)} = \frac{1 + e \cos \theta}{a(1 - e^2)}.$$

A integração do sistema II leva ao aparecimento de seis constantes. A forma como eliminamos η , mostra-nos que duas das constantes deverão determinar o plano da trajectória. Duas outras determinam a forma da elipse (5) descrita pelo satélite. Finalmente, as duas restantes são necessárias para determinar o movimento do satélite na sua órbita. Tomaremos, para definir a órbita descrita pelo ponto, o semi-eixo maior, a , e a excentricidade, e . O plano da órbita ficará definido pelo conhecimento da sua inclinação, i sobre o plano de referência XOY e ainda pelo conhecimento dum seu ponto nesse plano. Chamamos nodo ascendente o ponto em que o satélite passa do hemisfério Sul para o hemisfério Norte. A outra constante a ser introduzida é a ascensão recta (ou a longitude) do nodo ascendente, Ω . Por fim, para determinar o movimento do astro, introduzimos o instante, τ , em que ele passa mais próximo do corpo atraente (Instante da passagem no periastro) e o argumento do periastro, ω , que é o arco da órbita compreendido entre o nodo ascendente e o periastro. A estas constantes a , i , ω , Ω , τ , dá-se o nome de constantes elípticas.

Para acabar de resolver o problema dos dois corpos, torna-se necessário fazer ainda a integração de (3). Para a fazermos, vamos introduzir a variável auxiliar E dada por

$$(6) \quad r = a(1 - e \cos E).$$

Será

$$1 - e^2 = (1 + e \cos v) \cdot (1 - e \cos E)$$

e portanto

$$(7) \quad \cos v = \frac{\cos E - e}{1 - e \cos E},$$

$$\sin v = \frac{\sqrt{1 - e^2} \sin E}{1 - e \cos E}$$

ou, o que é equivalente,

$$(8) \quad \operatorname{tg}^2 \frac{v}{2} = \frac{1 - \cos v}{1 + \cos v} = \frac{1 + e}{1 - e} \operatorname{tg}^2 \frac{E}{2}.$$

Finalmente, derivando (7), teremos

$$\sin v \frac{dv}{dE} = \frac{(1 - e^2) \sin E}{(1 - e \cos E)^2}$$

ou seja

$$(9) \quad \frac{dv}{dE} = \frac{\sqrt{1 - e^2}}{1 - e \cos E}.$$

Utilizando esta relação e (6) para fazer uma mudança de variável em (3), somos conduzidos a

$$(10) \quad r^2 \frac{dv}{dt} = 2k = a^2 \sqrt{1 - e^2} (1 - e \cos E) \frac{dE}{dt}$$

Introduzindo a constante auxiliar

$$n = \frac{2k}{a^2 \sqrt{1 - e^2}}$$

e fazendo a integração temos

$$(11) \quad E - e \sin E = n(t - \tau).$$

A posição do corpo atraído será então determinada por meio de

$$M = n(t - \tau)$$

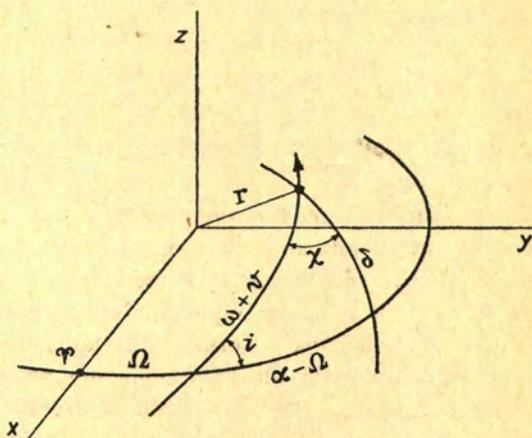
$$E = e \sin E = M$$

$$\operatorname{tg} \frac{v}{2} = \sqrt{\frac{1 + e}{1 - e}} \operatorname{tg} \frac{E}{2} \quad (III)$$

$$r = a(1 - e \cos E)$$

$$\operatorname{tg}(\alpha - \Omega) = \cos i \operatorname{tg}(\omega + v)$$

$$\sin \delta = \sin i \sin(\omega + v).$$



Usando (6), (9) e (10) vemos que

$$\left(\frac{dr}{dt}\right)^2 + r^2 \left(\frac{dv}{dt}\right)^2 = \frac{n^2 a^2}{r^2} (a^2 - a^2 e^2 \cos^2 E).$$

Usando novamente (6) podemos formar a constante das forças vivas

$$h = \frac{1}{2} \left[\left(\frac{dr}{dt}\right)^2 + r^2 \left(\frac{dv}{dt}\right)^2 \right] - \frac{fM}{r} = -\frac{n^2 a^2}{2} + \frac{n^2 a^5 - fM}{r}$$

o que mostra ser

$$(12) \quad 2h = -n^2 a^2 \quad n^2 a^5 = fM.$$

Finalmente, a partir de (10), vemos que o momento cinético da unidade de massa, constantemente normal ao plano da órbita, tem

a grandeza

$$(13) \quad r^2 \frac{dv}{dt} = n a^2 \sqrt{1-e^2} = \sqrt{f M a (1-e^2)}.$$

As suas coordenadas no referencial celeste que introduzimos anteriormente, serão

$$\sqrt{f M a (1-e^2)} \operatorname{sen} i \operatorname{sen} \Omega, \\ \sqrt{f M a (1-e^2)} \operatorname{sen} i \cos \Omega, \sqrt{f M a (1-e^2)} \cos i.$$

4. A análise das perturbações.

Mostrámos já como obter as equações de LAGRANGE quando se toma como coordenadas esféricas r, α, δ . Isto é tomámos como plano de referência o plano equatorial. Porém, o método que usámos é geral: poderíamos ter tomado como plano de referência um plano qualquer. As equações das perturbações simplificam-se grandemente quando se toma para plano de referência o plano da órbita no instante t . Com efeito, nesse caso, o momento de inércia é normal ao plano de referência. Em vez de α usamos v . Podemos decompor a força perturbadora numa componente radial, R , numa componente transversa, T (situada no plano da órbita) e numa componente bitransversa, B (Perpendicular ao plano da órbita). Em tais condições as equações de LAGRANGE são:

$$\frac{da}{dt} = \frac{2}{n\sqrt{1-e^2}} [e \operatorname{sen} v R + (1 + e \cos v) T]$$

$$\frac{de}{dt} = \frac{\sqrt{1-e^2}}{na} [\operatorname{sen} v R + (\cos v + \cos E) T]$$

$$\frac{di}{dt} = \frac{\sqrt{1-e^2}}{na} \frac{\cos(\omega + v)}{1 + e \cos v} B$$

$$\frac{d\Omega}{dt} = \frac{\sqrt{1-e^2}}{na \operatorname{sen} i} \frac{\operatorname{sen}(\omega + v)}{1 + e \cos v} B \quad (IV)$$

$$\frac{d\omega}{dt} + \cos i \frac{d\Omega}{dt} = \frac{\sqrt{1-e^2}}{nae} \left[-\cos v R + \right. \\ \left. + \operatorname{sen} v \left(1 + \frac{1}{1 + e \cos v} \right) T \right]$$

$$\frac{d\tau}{dt} - \frac{1}{n} \left(\frac{d\omega}{dt} + \cos i \frac{d\Omega}{dt} \right) = \frac{1}{n^2 a} \\ \left[\frac{2(1-e^2) \operatorname{sen} E - 3ne \operatorname{sen}^2 v (t-\tau)}{(1 + e \cos v) \operatorname{sen} E} R - \right. \\ \left. - \frac{3 \operatorname{sen} v}{\operatorname{sen} E} n(t-\tau) T \right].$$

Como em I a força perturbadora é dada pelas suas componentes radial $\left(R = \frac{\partial T}{\partial r} \right)$, meridiana $\left(M = \frac{1}{r} \frac{\partial T}{\partial \delta} \right)$ e perpendicular $\left(P = \frac{1}{r \operatorname{sen} \delta} \frac{\partial T}{\partial \alpha} \right)$ será

$$(1) \quad \begin{aligned} R &= R \\ T &= M \cos \chi + P \operatorname{sen} \chi \\ B &= M \operatorname{sen} \chi - P \cos \chi \end{aligned}$$

em que χ é o ângulo que o plano meridiano que contem o satélite faz com o plano da órbita. Pela figura da página 95 vemos que

$$(2) \quad \begin{aligned} \cos i &= \cos \delta \operatorname{sen} \chi \\ \operatorname{sen} i \cos(\omega + v) &= \cos \delta \cos \chi \\ \operatorname{sen} i \operatorname{sen}(\omega + v) &= \operatorname{sen} \delta. \end{aligned}$$

KING-HELE toma em primeira aproximação $P=0$. Corresponde isso a considerar o corpo atraente como sendo de revolução. Uma tal aproximação, porém, só pode ser válida quando a inclinação da órbita não for pequena. Para pequenos valores de i , o factor que multiplica $\frac{\partial T}{\partial \alpha}$ é grande e portanto P toma valores consideráveis. Isso mostra-nos que se devem usar satélites com órbitas de pequena inclinação quando se desejam determinar os harmónicos em xadrez.

Uma variante das equações de LAGRANGE quando, como no caso presente, as forças de perturbação derivam dum potencial T , é

$$\begin{aligned} \frac{da}{dt} &= -\frac{2}{n^2 a} \frac{\partial T}{\partial \tau} \\ \frac{de}{dt} &= -\frac{\sqrt{1-e^2}}{n a^2 e} \frac{\partial T}{\partial \omega} - \frac{1-e^2}{n^2 a^2 e} \frac{\partial T}{\partial \tau} \\ \frac{di}{dt} &= -\frac{1}{n a^2 \sqrt{1-e^2} \sin i} \frac{\partial T}{\partial \Omega} \\ &\quad - \frac{\operatorname{tg} \frac{i}{2}}{n a^2 \sqrt{1-e^2}} \frac{\partial T}{\partial \omega} \end{aligned} \tag{V}$$

$$\begin{aligned} \frac{d\Omega}{dt} &= \frac{1}{n a^2 \sqrt{1-e^2} \sin i} \frac{\partial T}{\partial i} \\ \frac{d\omega}{dt} &= \frac{\sqrt{1-e^2}}{n a^2 e} \frac{\partial T}{\partial e} + \frac{\operatorname{tg} \frac{i}{2}}{n a^2 \sqrt{1-e^2}} \frac{\partial T}{\partial i} \\ \frac{d\tau}{dt} &= \frac{2}{n^2 a} \frac{\partial T}{\partial a} + \frac{1-e^2}{n^2 a^2 e} \frac{\partial T}{\partial e} \end{aligned}$$

Vários outros sistemas equivalentes têm sido usados. Entre eles destaca-se o obtido por DELAUNAY para o estudo do movimento lunar. Em todos esses sistemas, bem como nos que já referimos, torna-se necessário exprimir o segundo membro em função dos elementos elípticos e do tempo.

As perturbações a estudar são decompostas como habitualmente, em perturbações periódicas e seculares. No estudo das perturbações seculares não é necessário entrar em consideração com os coeficientes J_{nm} e K_{nm} pois, a existência do movimento de rotação da Terra, é suficiente para fazer com que as perturbações correspondentes sejam periódicas.

O potencial de perturbação causa perturbações mais sensíveis no movimento do periéu e do nodo ascendente. Daí, utilizarem-se

quase exclusivamente, as equações em Ω e ω para o estudo do potencial terrestre. A introdução dos elementos da órbita nas equações das perturbações permite dar às equações das perturbações seculares a forma

$$\begin{aligned} \frac{d\Omega}{dt} &= \sum c_n J_n + \sum c_{nm} J_n J_m \\ \frac{d\omega}{dt} &= \sum d_n J_n + \sum d_{nm} J_n J_m \end{aligned} \tag{3}$$

em que os coeficientes c_n , c_{nm} , d_n e d_{nm} são funções dos elementos elípticos. Assim, por exemplo, KING-HELE (considerando só quantidades até à ordem de J_2^2) obtém

$$\begin{aligned} \frac{d\Omega}{dt} &= n \left(\frac{1}{1-e^2} \right)^2 \cos i \left[\frac{3}{2} J_2 + \right. \\ &\quad + \frac{9}{4} J_2^2 \left(\frac{1}{1-e^2} \right) \left(\frac{19}{12} \sin^2 i - 1 \right) + \\ &\quad \left. + \frac{15}{16} J_4 \left(\frac{1}{1+e^2} \right)^2 (y \sin^2 i - 4) \right] \\ \frac{d\omega}{dt} &= \frac{3}{4} n \left(\frac{1}{1-e^2} \right)^2 J_2 (5 \cos^2 i - 1). \end{aligned} \tag{4}$$

A segunda destas equações mostra-nos a importância que têm as órbitas com a inclinação $i = 63,4^\circ$ para as quais $5 \cos^2 i = 1$. É de notar que um grande número de satélites soviéticos têm uma inclinação próxima. Isso permite diminuir as perturbações causadas por J_2 no movimento do periéu.

A utilização de (3), calculado para diferentes satélites em órbitas distintas, e com inclinações variadas, permite, depois de determinado

$\frac{d\Omega}{dt}$ por observações determinar os coeficientes J_n . KING-HELE, por exemplo, faz uma determinação simultânea de J_2 , J_4 e J_6 usando o Sputnik 2 ($i = 65^\circ$), o Vanguard 1 ($i = 34^\circ$) e o Explorer 4 ($i = 50^\circ$).

Em 1963 melhora o resultado obtido pelo estudo do movimento do nodo de sete satélites com inclinações variando de $28,8^{\circ}$ até $97,4^{\circ}$. KOZAI em 1962 utiliza igualmente sete satélites mas com um número menor de órbitas.

Breve análise dos resultados obtidos.

Foram já realizadas várias determinações dos coeficientes J_n . Assim, dos valores comunicados ao congresso de Berkeley, destacamos os de KING-HELE e KOZAI a que já fizemos referência.

Ambas as determinações foram feitas pelo estudo do movimento secular do nodo ascendente. Os valores obtidos foram:

(Entre parêntesis indicam-se as precisões obtidas)

$$J_2 \cdot 10^6 \quad J_4 \cdot 10^6 \quad J_6 \cdot 10^6 \quad J_8 \cdot 10^6 \quad J_{10} \cdot 10^6 \quad J_{12} \cdot 10^6$$

KOZAI

$$\begin{array}{cccccc} +1082.48 & -1.84 & +0.39 & +0.002 & & \\ (0.04) & (0.09) & (0.09) & (0.07) & & \end{array}$$

KING-HELE

$$\begin{array}{cccccc} +1082.86 & -1.03 & +0.72 & +0.34 & -0.50 & +0.44 \\ (0.1) & (0.2) & (0.2) & (0.2) & (0.2) & (0.2) \end{array}$$

Os coeficientes de ordem ímpar foram estudados por meio das variações periódicas do perigeu, da excentricidade, da inclinação e do nodo ascendente. KOZAI dá-nos os valores

$$\begin{array}{cccc} J_3 \cdot 10^6 & J_5 \cdot 10^6 & J_7 \cdot 10^6 & J_9 \cdot 10^6 \\ -2.562 & -0.064 & -0.470 & -0.117 \\ (0.007) & (0.007) & (0.010) & (0.011) \end{array}$$

Estes resultados melhoram muito os que tinham sido obtidos por métodos gravimétricos. Com efeito JEFFREYS tinha chegado aos valores:

$$J_2 = (+1093 \pm 5) \cdot 10^{-6}, \quad J_3 = (-0.4 \pm 2.8) \cdot 10^{-6} \\ J_4 = (-2.2 \pm 2.1) \cdot 10^{-6}$$

valores estes que são de muito menor precisão.

Poderá supor-se, pelos valores dados, que os métodos da geodesia dinâmica vêm substituir, com vantagem, os métodos da gravimetria. Isso só em parte é verdade. Com efeito, a observação de satélites permite determinar, com uma muito grande precisão, os harmônicos de ordem inferior. A precisão com que se fazem estas determinações vai diminuindo à medida que a ordem dos harmônicos vai aumentando. Pelo contrário, a precisão com que se fazem as determinações pelos métodos gravimétricos é independente da ordem dos harmônicos. Além disso, certamente essa precisão melhorará à medida que forem aumentando o número de observações gravimétricas realizadas. Desta forma, é de esperar que os métodos gravimétricos e os métodos da geodesia dinâmica se completem harmoniosamente para, conjuntamente, determinarem o potencial terrestre.

5. Conclusão.

Ao apresentar os diferentes métodos usados para determinar a forma da Terra, procuramos mostrar os problemas ligados a cada um. Para isso tivemos que os estudar separadamente desarticulando em parte a ciência geodésica. Na verdade esta, para fazer esse estudo, não recorre alternadamente a um e outro método mas, bem ao contrário, procura utilizá-los simultaneamente.

Já demos alguns exemplos da colaboração existente entre os diferentes métodos utilizados. Assim, vimos como o conceito de cota geopotencial era obtido a partir do nivelamento e da gravimetria. Vimos igualmente, no fim do parágrafo anterior, que o potencial terrestre era determinado pela colaboração da gravimetria e da geodesia dinâmica. Outros exemplos de trabalho conjunto dos diferentes ramos da geodesia podem ser apresentados.

No Capítulo I mostra-se como por meio de triangulação podemos definir referenciais ligados à Terra. A cada datum escolhido corresponde uma origem e um sistema de eixos bem determinado. A extensão e a conexão das diferentes redes permite reduzir os diferentes referenciais a utilizar a um número mínimo. O recurso a métodos de triangulação com grandes lados (triangulação por meio de fochos e satélites, medição de lados por meio de shoran, etc.) permitirá mesmo realizar um referencial único ligado à Terra. Este é impossível de realizar pelos métodos clássicos dada a impossibilidade de triangular as grandes extensões cobertas pelos oceanos.

Os métodos da geodesia geométrica são, porém, impotentes para fazer com que a origem de coordenadas do referencial terrestre adoptado seja o centro de massas terrestre. Será então necessário recorrer a métodos gravimétricos e de geodesia dinâmica para impor a condição que, nesse referencial, seja $J_1 = 0$.

Os diferentes pontos da Terra não são coordenados nesse referencial. São-no mais simplesmente pela introdução dum elipsoide auxiliar de forma muito vizinha da da Terra. Este pode ser de revolução ou triaxial. No primeiro caso, o seu achatamento será determinado por métodos gravimétricos e de geodesia dinâmica. Para isso impõe-se a condição do elipsoide ser uma superfície de nível dum campo de gravidade normal tal que uma função bem determinada de J_2 e J_4 tenha o mesmo valor nesse campo e no campo da gravidade terrestre.

Se abandonarmos a condição do elipsoide de referência ser de revolução, podemos então introduzir a condição do campo de gravidade normal ter os mesmos coeficientes J_{21} e K_{21} que o campo da gravidade terrestre. Isso determina a elipticidade do equador e a longitude do semi-eixo maior.

Vemos que a forma do elipsoide de referência é determinada com maior precisão

por métodos gravimétricos e de geodesia dinâmica. Estes são, porém, impotentes para nos darem as dimensões desse elipsoide. Temos então que recorrer a métodos da geodesia geométrica para determinar o valor do seu raio equatorial. Esta determinação está ligada ao ajuste de redes geodésicas.

Por meio desta última operação faz-se a colocação do elipsoide escolhido e, portanto, faz-se a escolha de referencial terrestre. Como dissemos o referencial obtido não tem o seu centro no centro de massas terrestre. Além disso, a cada rede geodésica corresponde um referencial distinto. A unificação dos diferentes referenciais é uma tarefa da geodesia dinâmica.

Com efeito na geodesia dinâmica, estuda-se o movimento dum satélite em relação a um referencial celeste que tem a sua origem no centro de massas terrestre. A passagem desse referencial celeste para qualquer referencial terrestre é obtida pelo estudo do movimento do pólo e das variações da velocidade de rotação da Terra. Este dar-nos-á elementos para estabelecer, em cada instante, uma rotação de eixos. A mudança de referencial ficará determinada a menos de uma translacção.

Conhecidos os coeficientes J_i , J_{ik} e K_{ik} do potencial de atracção terrestre, podemos determinar o movimento dum satélite artificial. Observado este de diferentes pontos duma dada rede geodésica, podemos obter as coordenadas dele no referencial ligado a essa rede e em qualquer instante. A evolução do satélite permite-nos, além disso, determinar a translacção que nos faltava na mudança de referenciais a que nos referimos acima.

Finalmente, conhecidas, em qualquer instante, as coordenadas dum satélite relativamente a uma dada rede podemos, por meio de observações, ligar qualquer ponto de coordenadas desconhecidas a essa rede. O satélite comporta-se neste caso como um mensageiro que leva as coordenadas duma dada rede a qualquer ponto da Terra.

Neste caso, ao contrário do que acontecia no método de triangulação por meio de satélites, as observações não têm de ser simultâneas visto serem conhecidas as coordenadas do satélite observado. A precisão obtida não é contudo tão grande como no referido método, visto o satélite estar sujeito a perturbações que nem sempre são controláveis. Para que a precisão seja o maior possível, importa que as observações sejam realizadas em ocasiões em que as coordenadas do satélite sejam bem conhecidas. Para isso observa-se, do ponto desconhecido, o satélite pouco depois dele ter sido coordenado na rede adoptada. Neste caso o movimento do satélite foi pouco perturbado pelos valores mal conhecidos do potencial. Pode ainda observar-se o satélite depois dele ter dado uma revolução completa pois neste caso só há a considerar, apenas, as perturbações seculares.

O problema da determinação das coordenadas dum ponto terrestre pela observação dum satélite reduz-se então a :

Determinar as coordenadas dum ponto P quando dele se observaram satélites nas posições $S_1, S_2, S_3, \dots, S_n$, conhecidas.

A observação do satélite sobre um campo de estrelas permite determinar os correspon-

dentes vectores direccionais e_1, e_2, e_3, \dots . O ponto P será então determinado pelas equações vectoriais

$$\overrightarrow{S_i - P} = |S_i - P| \vec{e}_i.$$

A introdução de coordenadas aproximadas do ponto P permite formar as equações de observação.

O método que acabamos de indicar foi utilizado pelo Instituto Smithsoniano para interligar as principais redes geodésicas.

BIBLIOGRAFIA

- W. A. HEISKANEN and F. A. VENING-MEINESZ, *The Earth and its gravity field*. Bonford Geodesy.
- M. S. MOLODENSKI, V. E. EREMEEV M. I. YURKINA, *Methods for study of the external gravitational field and figure of the Earth*.
- I. I. MUELLER, *Introduction to Satellite Geodesy*.
- G. VEIS, *Geodetic uses of artificial satellites*.
- Comunicações aos congressos da Associação Internacional de Geodesia.
- The use of Artificial Satellites for Geodesy. Contemporary Geodesy.
- Geodesy in Space Age.
- Foram ainda consultados diversos artigos de «Bulletin Géodésique».
- Survey Review.

LITERATURA MATEMÁTICA RECIENTE

Editor — MASSON ET C.^{ie}, Paris

- J. BASS — *Cours de Mathématiques — Troisième Édition Revue et Corrigée.*
— *Exercices de Mathématiques.*
- M. BOUIX — *Les Fonctions Généralisées ou Distributions.*
- M. A. TONNELET — *Les Vérifications Experimentales de la Relativité Générale.*
- A. HOCQUENGHEM & P. JAFFARD — *Mathématiques. Tome I. Éléments de calcul différentiel et intégral.*

Editor — LIBRAIRIE SCIENTIFIQUE ALBERT BLANCHARD — Paris

- MARCEL DOLIGÉZ — *Gravitation — Contribution à la théorie corpusculaire de la gravitation.*
- H. LAURENT — *Théorie des Jeux de Hasard.*

Editor — DUNOD, Paris

Monographies Universitaires de Mathematiques

22. A. G. KUROSCHE — *Algèbre générale.*
23. I. M. GUELFAND et N. Y. VILENKIN — *Les distributions. Tome 4: Applications de l'analyse harmonique.*
24. C. FOURGEAUD et A. FUCHS — *Statistique.*
25. J. GARSOUX — *Analyse mathématique.*
26. A. GUICHARDET — *Analyse harmonique commutative.*
27. G. HOCHSCHILD — *La structure des groupes de Lie.*
28. MME Y. CHOQUET-BRUHAT — *Geometrie différentielle et systèmes extérieures.*
29. PHAM MAU QUAN — *Introduction à la géométrie des variétés différentiables,*
30. R. ISAACS — *Jeux différentiels. Théorie des jeux appliqués aux domaines de la guerre, des poursuites, du contrôle et de l'optimisation.*
31. O. A. LADYZENSKAJA et N. N. ORAL'OEVA — *Equations aux dérivées partielles de type elliptique.*
32. J. LÉVY-BRUHL — *Introduction aux structures algébriques.*

**BENTO
DE JESUS
CARAÇA**

CONFERÊNCIAS E OUTROS ESCRITOS

Livraria Sá da Costa — Rua Garrett, 100-102

**ENSAIOS
de Geografia Humana e Regional**

FOR

ORLANDO RIBEIRO
(1 Volume)

Livraria Sá da Costa — Rua Garrett, 100-102

E. S. CABRERA y H. J. MEDICI

**LA ENSEÑANZA DE LAS MATEMATICAS
EN LA REPUBLICA ARGENTINA**

Matemática 1 a 4

Trigonometria, nociones de limite continuidad y derivada

Ejercicios 1 a 4

Astronomia Elemental

Libreria del Colegio — BUENOS AIRES

GAZETA DE MATEMÁTICA

Número avulso: 30 escudos

Assinatura relativa a 1971 (4 números) 100 escudos

Assinatura para o estrangeiro, 200 escudos

PONTOS DE EXAME

Uma das secções permanentes da *Gazeta de Matemática* é constituída pelos pontos de Matemática do exame do 3.º ciclo do ensino liceal e de exames de aptidão às Universidades e pontos de exames de frequência e finais das cadeiras de matemática das escolas superiores.

2.ª EDIÇÃO DO VOL. II (N.º 5 e 8)

Continua aberta a inscrição para a nova edição do ano II da *Gazeta de Matemática* (n.º 5 a 8) ao preço de escudos 30. Esta nova edição oferece aos leitores da *Gazeta de Matemática* a possibilidade de completarem as suas colecções no formato e características actuais e com os textos cuidadosamente revistos. Logo que as inscrições atinjam o número de 300, proceder-se-á, à composição, impressão e distribuição da nova edição do ano II. Depois de publicada, a segunda edição do volume II será vendida ao preço de escudos 40.

CONDIÇÕES DE ASSINATURA

A administração da *Gazeta de Matemática* aceita, quando pedidas directamente, assinaturas de quatro números, ao preço de escudos 100, para o que basta indicar o nome, a morada e o local de cobrança. As assinaturas são renovadas automaticamente no

seu termo, salvo aviso prévio em contrário. Todas as assinaturas têm início com o primeiro número publicado em cada ano.

ASSINATURAS GRATUITAS

Todo o assinante que indique à administração da *Gazeta de Matemática* dez novos assinantes beneficiará de uma assinatura gratuita durante o ano seguinte ao da sua assinatura.

NÚMEROS ATRASADOS

Estão completamente esgotados os números 5 a 15, da *Gazeta de Matemática*. Os restantes números são vendidos aos preços seguintes:

N.º 1-4 (2.ª edição do ano I, no formato actual e com o texto cuidadosamente revisto)	Portu- gal	40\$00	Estran- geiro	75\$00
N.º 16 a 49, cada		12\$50		25\$00
N.º 50, 76-77		60\$00		100\$00
51 a 75 { número simples		17\$50		25\$00
N.º 78 a 99 { " duplo		35\$00		50\$00
101 a 108 { " duplo		35\$00		50\$00
N.º 100		100\$00		150\$00
N.º 109-112, 113-116 e 117-120		120\$00		200\$00

A administração da *Gazeta de Matemática* executa qualquer encomenda à cobrança pelo correio.

ANGARIE ASSINANTES PARA A «GAZETA DE MATEMÁTICA».

Concorrerá, assim, para o melhoramento
de uma revista sem objectivos comerciais

PREÇO ESC. 120\$00

ADMINISTRAÇÃO DA «GAZETA DE MATEMÁTICA»
Rua Diário de Notícias, 134-1.º-Esq.º — LISBOA - 2 — Telefone 369449