

Ajuste computacional de dados experimentais

Filomena F.M. Freitas ^a
 Fernando M.S.S. Fernandes ^a
 Lídia M.P.C. Albuquerque ^a

INTRODUÇÃO

O objectivo deste trabalho é a aplicação do ajuste linear de dados a dois casos concretos com vista à determinação de grandezas físico-químicas, exemplificando como se pode complementar uma experiência laboratorial com microcomputadores. Em trabalho experimental é necessário, frequentemente, extrair informação de um conjunto de dados que consta de valores de uma ou mais variáveis independentes (x) e dos correspondentes valores de uma variável dependente (y). Em geral, os referidos dados estão afectados de erro e , além disso, a relação funcional $y=f(x)$ é complicada ou até mesmo, por vezes, desconhecida. Torna-se, assim, necessário aproximar $f(x)$ por uma outra função $g(x)$ pertencente a uma determinada família de funções que seja de tratamento matemático simples e cuja escolha obedece a determinadas condições [1].

Como consequência desta aproximação, surgem erros, designados por resíduos (r_i), que poderão satisfazer uma tolerância pré-estabelecida, desde que se faça uma escolha adequada da função a usar na aproximação.

A determinação dos parâmetros de ajuste é normalmente feita recorrendo ao método dos mínimos quadrados, não só porque o tratamento matemático nela envolvido é simples como também este método possui uma base estatística rigorosa para um estudo adequado de propagação de erros.

De acordo com o método anterior, os valores óptimos dos parâmetros da função aproximadora são os que minimizam a soma pesada dos quadrados dos resíduos:

$$\chi^2 = \sum_{i=1}^N \frac{r_i^2}{\sigma_i^2} = \sum_{i=1}^N \frac{1}{\sigma_i^2} (y_{i(\text{exp.})} - y_{i(\text{calc.})})^2 \quad (1)$$

em que

σ_i^2 — variância da amostra para y_i

$y_{i(\text{exp.})}$ — valor experimental de y_i

$y_{i(\text{calc.})}$ — valor calculado de y_i

N — número de pares de valores experimentais

EXEMPLO A — DETERMINAÇÃO COMPUTACIONAL DA TEMPERATURA CRÍTICA SUPERIOR DE SOLUÇÃO DE UMA MISTURA LÍQUIDA BINÁRIA

Fundamento teórico

Para um sistema líquido binário parcialmente miscível, a uma dada pressão e na vizinhança da temperatura crítica superior de solução (temperatura acima da qual os componentes são miscíveis em quaisquer proporções), a temperatura de miscibilidade completa (T) é uma função quadrática da composição total do sistema [2]:

$$T = a_0 + a_1 x + a_2 x^2 \quad (A.1)$$

em que x é a fracção molar de um dos componentes. Esta função apresenta um máximo que corresponde à temperatura crítica superior de solução. Então:

$$\left(\frac{dT}{dx}\right)_{\text{TCSS}} = 0 \Rightarrow x_{\text{TCSS}} = -\frac{a_1}{2a_2} \quad (A.2)$$

e

$$\text{TCSS} = a_0 + a_1 x_{\text{TCSS}} + a_2 x_{\text{TCSS}}^2 \quad (A.3)$$

Assim, ajustando um conjunto de N pares experimentais (T_i, x_i) à equação (A.1), determinam-se os valores óptimos dos parâmetros a_0, a_1 e a_2 com os quais se obtém a TCSS.

Tratamento estatístico

No caso presente a função de ajuste é do tipo:

$$y = a_0 + a_1 x + a_2 x^2 \quad (A.4)$$

em que $y=T$ e $x=x$.

Pelo método dos mínimos quadrados, os parâmetros de ajuste são os que minimizam a soma pesada dos quadrados dos resíduos:

$$\chi^2 = \sum_{i=1}^N \frac{1}{\sigma_i^2} (y_i - a_0 - a_1 x_i - a_2 x_i^2)^2 \quad (A.5)$$

Para facilitar o tratamento matemático admite-se que todos os valores experimentais estão afectados da mesma incerteza, pelo que se considera $\sigma_i = \sigma$.

^a Faculdade de Ciências — CECUL, Calçada Bento da Rocha Cabral, n.º 14 — 1200 Lisboa.

Os parâmetros, acima referidos, correspondem às incógnitas do seguinte sistema de equações lineares:

$$\begin{cases} \frac{2}{\sigma^2} \sum_{i=1}^N (y_i - a_0 - a_1 x_i - a_2 x_i^2) (-1) = 0 \\ \frac{2}{\sigma^2} \sum_{i=1}^N (y_i - a_0 - a_1 x_i - a_2 x_i^2) (-x_i) = 0 \\ \frac{2}{\sigma^2} \sum_{i=1}^N (y_i - a_0 - a_1 x_i - a_2 x_i^2) (-x_i^2) = 0 \end{cases} \quad (\text{A.6})$$

que se obtém derivando a equação (A.5) em ordem a cada um dos parâmetros e igualando a zero os resultados obtidos. A resolução deste sistema, por via computacional, requer a implementação da respectiva representação matricial:

$$\begin{bmatrix} N & \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 \\ \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 & \sum_{i=1}^N x_i^3 \\ \sum_{i=1}^N x_i^2 & \sum_{i=1}^N x_i^3 & \sum_{i=1}^N x_i^4 \end{bmatrix} \times \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N y_i x_i \\ \sum_{i=1}^N y_i x_i^2 \end{bmatrix} \quad (\text{A.7})$$

\uparrow MCS \uparrow MIS \uparrow MTIS

em que

MCS — matriz dos coeficientes do sistema
 MIS — matriz das incógnitas do sistema
 MTIS — matriz dos termos independentes do sistema

Utilizando o método de Gauss-Jordan [2,3] na resolução numérica deste sistema, obtém-se a matriz das incógnitas (ou parâmetros) e a inversa da matriz dos coeficientes do sistema (MCS^{-1}), muito útil do ponto de vista estatístico pois permite determinar as variâncias dos parâmetros de ajuste.

A variância de cada parâmetro [2] é dada por:

$$\sigma_{aj}^2 = \sigma^2 C'_{j+1, j+1} \quad (\text{A.8})$$

com $j=0, 1, \dots, n$, sendo

σ_{aj}^2 — variância do parâmetro de ajuste a_j
 σ^2 — variância do ajuste
 $C'_{j+1, j+1}$ — elemento da diagonal da inversa da matriz MCS
 n — grau do polinómio de ajuste

Neste caso:

$$\sigma^2 = \frac{1}{N-3} \sum_{i=1}^N [y_i - (a_0 + a_1 x_i + a_2 x_i^2)]^2 \quad (\text{A.9})$$

$y_{i(\text{calc.})}$

em que

$N-3$ — n.º de graus de liberdade, sendo 3 o número de parâmetros.

Os valores dos desvios padrão para x_{TCSS} e TCSS cal-

culam-se através das seguintes expressões:

$$\sigma_{x_{TCSS}} = x_{TCSS} \times \left(\frac{\sigma_{a_1}^2}{a_1^2} + \frac{\sigma_{a_2}^2}{a_2^2} - 2 \times \frac{\sigma_{a_1, a_2}^2}{a_1 \times a_2} \right)^{1/2} \quad (\text{A.10})$$

$$\sigma_{TCSS} = \left(\sigma_{a_0}^2 + \sigma_{a_1}^2 \times \frac{a_1^2}{4 a_2^2} + \sigma_{a_2}^2 \times \frac{a_1^4}{16 a_2^4} - \sigma_{a_0, a_1}^2 \times \frac{a_1}{a_2} + \sigma_{a_0, a_2}^2 \times \frac{a_1^2}{2 a_2^2} - \sigma_{a_1, a_2}^2 \times \frac{a_1^3}{4 a_2^3} \right)^{1/2} \quad (\text{A.11})$$

em que σ_{a_i, a_j}^2 , covariância entre os parâmetros a_i e a_j , é dada pelo produto $\sigma^2 \times C'_{i, j}$, sendo $C'_{i, j}$ um dos elementos com $i \neq j$ da inversa da matriz MCS.

Algoritmo

INÍCIO

```
{ Introdução de dados }
Ler PE { n.º de pares experimentais }
Para i variando de 1 a PE repetir
  Ler X(i), Y(i) { Pares de valores experimentais —
    fracção molar/temperatura }
GR ← 2
SOMY ← 0
{ Matriz das potências de x — MPX e
  somatório dos y — SOMY }
Para i variando de 1 a PE repetir
  PROD ← 1
  Para k variando de 1 a 2×GR repetir
    PROD ← PROD×X(i)
    MPX(i,k) ← PROD
    SOMY ← SOMY+Y(i)
{ Matriz dos somatórios das potências de x —
  MSPX }
Para k variando de 1 a 2×GR repetir
  SOMA ← 0
  Para i variando de 1 a PE repetir
    SOMA ← SOMA+MPX(i,k)
  MSPX(k) ← SOMA
{ Matriz dos coeficientes do sistema — MCS }
{ MI — Inversa da matriz MCS }
NE ← GR+1 { N.º de Equações }
Para i variando de 1 a NE repetir
  S ← i-1
  Para k variando de 1 a NE repetir
    MCS(i,k) ← MSPX(S)
    S ← S+1
  Se i=k então
    MI(i,k) ← 1
  Senão
    MI(i,k) ← 0
{ Matriz dos termos independentes do sistema —
  MTIS }
MTIS(1) ← SOMY
Para i variando de 1 a GR repetir
  S ← i+1
  Para k variando de 1 a PE repetir
    MTIS(S) ← MTIS(S)+Y(k)×MPX(k,i)
Chamar Subrotina GAUSSJORDAN
{ Matriz dos valores de y calculados — YCALC }
SQD ← 0 { Soma dos quadrados dos resíduos }
Para i variando de 1 a PE repetir
  YCALC(i) ← MIS(1)+MIS(2)×X(i)+MIS(3)×
    ×X(i)²
  SQD ← SQD+Y(i)-YCALC(i)²
VARAJ ← SQD/(PE-3) { Variância do ajuste }
{ Matriz das variâncias dos parâmetros — VARP }
```

```

Para i variando de 1 a NE repetir
| VARP(i) ← VARAJ × MI(i,i)
{ Covariância entre os parâmetros }
COV01 ← VARAJ × MI(1,2)
COV02 ← VARAJ × MI(1,3)
COV12 ← VARAJ × MI(2,3)
{ "Display" dos resultados }
Escrever 'Parâmetros de ajuste'
Para i variando de 1 a NP repetir
| Escrever MIS(i), '±', √VARP(i)
XTCSS ← -MIS(2)/(2 × MIS(3))
TCSS ← MIS(1) + MIS(2) × XTCSS + MIS(3) × XTCSS²
{ Desvios padrão de XTCSS e de TCSS }
DPX ← XTCSS × (VARP(2)/MIS(2)² + VARP(3)/MIS(3)² - 2 ×
    × COV23/(MIS(2) × MIS(3)))¹/²
DPT ← (VARP(1) + VARP(2) × MIS(2)²/4 MIS(3)² +
    + VARP(3) × MIS(2)⁴/16 MIS(3)⁴ - COV01 × MIS(2)/MIS(3) +
    + COV02 × MIS(2)²/2 MIS(3)² - COV12 × MIS(2)³/4 MIS(3)³)¹/²
Escrever 'TCSS=', TCSS, '±', DPT
Escrever 'xTCSS=', XTCSS, '±', DPX
Chamar Subrotina GRAFICO
FIM DO ALGORITMO

```

Aplicação

Utilizando o programa correspondente ao algoritmo anterior e os dados experimentais da tabela I [2], em que se apresentam vários valores de fracção molar (x) de hexano e as respectivas temperaturas (T) de miscibilidade completa para o sistema anilina-hexano, à pressão de 1 atm, obtiveram-se os seguintes resultados:

PARÂMETROS DE AJUSTE

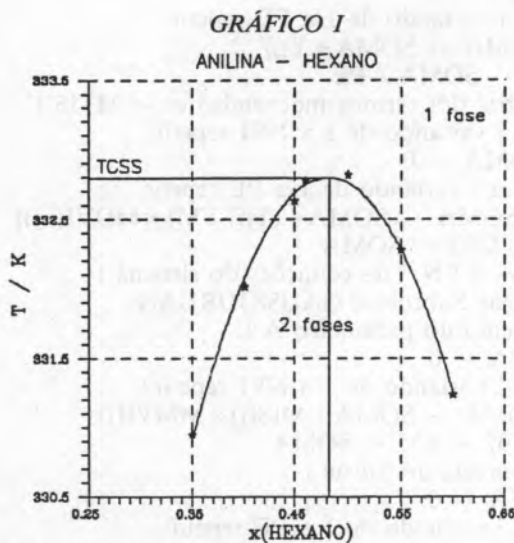
$$a_0 = 307 \pm 1$$

$$a_1 = 106 \pm 3$$

$$a_2 = 110 \pm 3$$

$$TCSS = 332,80 \text{ K} \pm 0,03 \text{ K}$$

$$x_{TCSS} = 0,481 \pm 0,001$$



Pressão de trabalho — 1 atm

TABELA I

x	0,350	0,400	0,450	0,460	0,500	0,550	0,600
T/K	330,95	332,02	332,63	332,78	332,82	332,28	331,24

EXEMPLO B — DETERMINAÇÃO COMPUTACIONAL DE FUNÇÕES TERMODINÂMICAS DE ACTIVAÇÃO A PARTIR DA EQUAÇÃO DE EVERETT E WYNNE-JONES

Fundamento teórico

A partir de um conjunto de medidas de velocidade específica (k) de uma reacção química, a diferentes temperaturas, é possível determinar as seguintes funções termodinâmicas de activação: variação de entalpia padrão de activação (Δ^*H°), variação de entropia padrão de activação (Δ^*S°) e variação da capacidade calorífica padrão de activação ($\Delta^*C_p^\circ$). Um dos processos de determinação das grandezas, acima referidas, consiste em ajustar os valores experimentais de $\ln k$ à equação de Everett e Wynne-Jones [4]:

$$\ln k = A + B/T + C \ln T \quad (B.1)$$

em que T representa a temperatura do sistema químico em estudo e A , B , C os parâmetros de ajuste. Por aplicação de uma análise de regressão múltipla a esta equação, estimam-se os valores dos referidos parâmetros com os quais se determinam Δ^*H° , Δ^*S° e $\Delta^*C_p^\circ$, de acordo com as seguintes expressões:

$$\Delta^*H^\circ = R [T(C-1) - B] \quad (B.2)$$

$$\Delta^*S^\circ = R [\ln(h/\kappa_B) + A + C - 1 + (C-1)\ln T] \quad (B.3)$$

$$\Delta^*C_p^\circ = R (C-1) \quad (B.4)$$

em que

h — constante de Planck

κ_B — constante de Boltzmann

Tratamento estatístico

Os parâmetros óptimos de ajuste são os que minimizam a soma dos quadrados dos resíduos:

$$\chi^2(A, B, C) = \sum_{i=1}^N \frac{1}{\sigma_i^2} (y_i - A - B X_{i1} - C X_{i2})^2 \quad (B.5)$$

em que $y_i = \ln k_i$, $X_{i1} = 1/T_i$, $X_{i2} = \ln T_i$ e N o número de pares experimentais (k_i , T_i). Assim, admitindo que todos os valores experimentais, y_i , estão afectados da mesma incerteza, derivando este somatório em ordem a cada um dos parâmetros e, igualando a zero as respectivas derivadas, obtém-se o seguinte sistema linear de 3 equações:

$$\begin{cases} A \sum_{i=1}^N 1 + B \sum_{i=1}^N X_{i1} + C \sum_{i=1}^N X_{i2} = \sum_{i=1}^N y_i \\ A \sum_{i=1}^N X_{i1} + B \sum_{i=1}^N X_{i1}^2 + C \sum_{i=1}^N X_{i1} X_{i2} = \sum_{i=1}^N y_i X_{i1} \\ A \sum_{i=1}^N X_{i2} + B \sum_{i=1}^N X_{i1} X_{i2} + C \sum_{i=1}^N X_{i2}^2 = \sum_{i=1}^N y_i X_{i2} \end{cases} \quad (B.6)$$


```

VARAJ ← SOMA / (PE-NVI-1)
{ Variâncias dos parâmetros B e C }
VARB ← VARAJ × MI(1,1)
VARC ← VARAJ × MI(2,2)
{ Covariância entre B e C }
COVBC ← VARAJ × MI(1,2)
{ Variância do parâmetro A }
VARA ← VARB × MMVI(1)2 + VARC × MMVI(2)2 +
      + 2 × COVBC × MMVI(1) × MMVI(2)
{ Covariância entre A e C }
COVAC ← -MMVI(2) × VARC - MMVI(1) × COVBC
{ "DISPLAY" dos parâmetros e respectivos desvios
  padrão }
Escrever 'A=' , MIS(0), '±', √VARA
Escrever 'B=' , MIS(0), '±', √VARB
Escrever 'C=' , MIS(0), '±', √VARC
{ Cálculo das funções termodinâmicas de activação e
  respectivos desvios padrão, "DISPLAY" dos resul-
  tados }
CG ← 1.9868 { Constante dos gases }
CPB ← -ln (6.6255 × 10-27 / (1.3805 × 10-16))
      { ln(h/kB) }
Para i variando de 1 a PE repetir
  HO ← CG × (TD × (MIS(2)-1) - MIS(1)) { Δ*H° }
  SO ← CG × (CPB + MIS(2)-1 + (MIS(2)-1) × lnTD)
      { Δ*S° }
  CO ← CG × MIS(2)-1 { Δ*Cp° }
  DPH ← CG × (TEMPR(i)2 × VARC + VARB-2 ×
    × TEMPR(i) × COVBC)1/2 { Desv. Padrão
    de Δ*H2° }
  DPS ← CG × (VARA + (1 + ln(TEMPR(i)))2 ×
    × VARC + 2 × (1 + ln(TEMPR(i))) ×
    × COVAC)1/2 { Desv. padrão de Δ*S° }
  DPC ← CG × √VARC { Desv. Padrão de Δ*Cp° }
  Escrever TEMPR(i), HO, DPH, SO, DPS, CO, DPC
Chamar Subrotina GRAFICO
FIM DO ALGORITMO

```

Aplicação

Com os dados experimentais da Tabela II [6], em que se apresentam vários valores de velocidades específicas (*k*) e as respectivas temperaturas (*T*), à pressão de 1 atm, e, utilizando o programa relativo ao algoritmo anterior, obtiveram-se os seguintes resultados:

PARÂMETROS DE AJUSTE

$$\begin{aligned}
 A &= -23,91 \pm 129,21 \\
 B &= -9530 \pm 5820 \\
 C &= 7,95 \pm 19,25
 \end{aligned}$$

VARIAÇÕES PADRÃO DE ACTIVAÇÃO DE:

TEMPERATURA /K	ENTALPIA /cal mol ⁻¹	ENTROPIA /cal mol ⁻¹ K ⁻¹	CAP. CALORÍF. /cal mol ⁻¹ K ⁻¹
283,15	22844 ± 763	-2,946 ± 2,636	14 ± 38
288,15	22913 ± 582	-2,704 ± 2,000	14 ± 38
293,15	22982 ± 411	-2,466 ± 1,407	14 ± 38
303,15	23120 ± 219	-2,003 ± 0,723	14 ± 38
308,15	23189 ± 312	-1,777 ± 1,003	14 ± 38
313,15	23258 ± 468	-1,555 ± 1,499	14 ± 38
318,15	23327 ± 644	-1,336 ± 2,051	14 ± 38
323,15	23396 ± 827	-1,121 ± 2,618	14 ± 38

GRÁFICO II

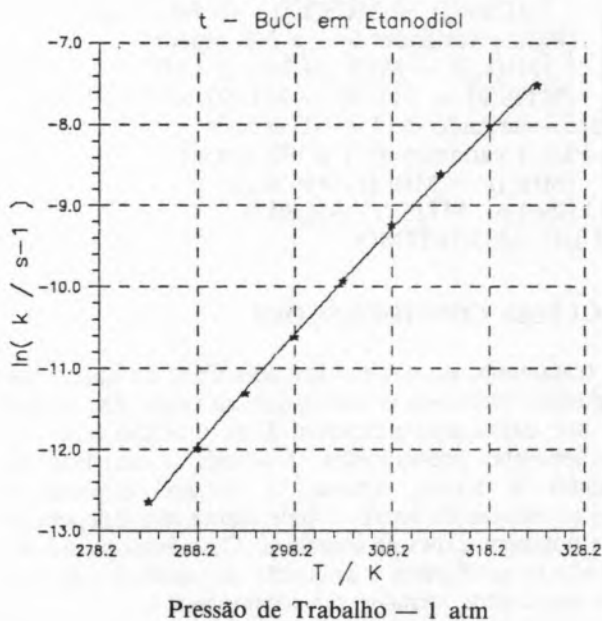


TABELA II

T / °C	* <i>k_m</i> × 10 ⁶ / s ⁻¹
10	3,245
15	6,250
20	12,36
25	24,61
30	48,71
35	97,61
40	182,4
45	323,3
50	543,8

* Velocidade específica média.

Algoritmo da Subrotina GaussJordan

INICIO

```

Para i variando de 1 a NE-1 repetir
  MAX ← ABS(MCS(i,i))
  L ← i
  Para k variando de i+1 a NE repetir
    MD ← ABS(MCS(k,i))
    Se MD > MAX então
      MAX ← MD
      L ← k
  Se MAX = 0 então
    Escrever 'Não se pode resolver'
    Parar
  Se i <> L então
    Para k variando de i a NE repetir
      MCS(i,k) ↔ MCS(L,k)
    Para k variando de 1 a NE repetir
      MI(i,k) ↔ MI(L,k)
      MTIS(i) ↔ MTIS(L)
    Para k variando de i+1 a NE repetir
      MU ← -MCS(k,i) / MCS(i,i)
      Para j variando de i a NE repetir
        MCS(k,j) ← MCS(k,j) + MCS(i,j) × MU
      Para j variando de 1 a NE repetir
        MI(k,j) ← MI(k,j) + MI(i,j) × MU
        MTIS(k) ← MTIS(k) + MTIS(i) × MU
  Para i variando de 2 a NE repetir
    Para k variando de i-1 a 1 passo -1 repetir
      MU ← -MCS(k,i) / MCS(i,i)

```

```

Para j variando de 1 a NE repetir
  MCS(k,j) ← MCS(k,j) + MCS(i,j) × MU
Para j variando de 1 a NE repetir
  MI(k,j) ← MI(k,j) + MI(i,j) × MU
  MTIS(k) ← MTIS(k) + MTIS(i) × MU
Para i variando de 1 a NE repetir
  Para j variando de 1 a NE repetir
    MI(i,j) ← MI(i,j) / MCS(i,i)
    MIS(i) ← MTIS(i) / MCS(i,i)
FIM DO ALGORITMO

```

ALGUMAS CONSIDERAÇÕES

No tratamento estatístico dos dois tipos de ajuste considerados, omitiu-se o factor que dá conta das incertezas nos dados experimentais. Essa omissão baseou-se nos seguintes pressupostos: *i* — toda a incerteza das medidas é devida, apenas, à variável dependente; *ii* — os valores da variável dependente são independentes e possuem a mesma incerteza. Caso estes pressupostos não se verifiquem a aplicação do método dos mínimos quadrados torna-se mais complexa.

Neste tratamento está, ainda, implícito que: *i* — a distribuição das incertezas é do tipo gaussiano, condição que está na base do método dos mínimos quadrados; *ii* — não existem erros sistemáticos, o que a acontecer originaria correlações entre as incertezas, vindo os resíduos alterados em relação à sua distribuição original. A aplicação do método dos mínimos quadrados no ajuste de dados experimentais a uma função tem vantagens em relação a outros métodos, tais como: *i* — conduzir a sistemas de equações lineares que são de fácil resolução; *ii* — permitir fazer tratamento de erros, pois o citado método possui uma base estatística rigorosa.

A utilização do método de Gauss-Jordan, na resolução do sistema de equações, justifica-se quando se pretende fazer um tratamento estatístico dos resultados experimentais, pois com este método obtém-se a inversa da

matriz dos coeficientes do sistema que, como se viu, é muito útil na determinação dos erros dos parâmetros de ajuste. No entanto, este método utiliza mais aritmética do que, por exemplo, o método de eliminação de Gauss, cuja utilização é preferível caso não se esteja interessado na inversa da matriz dos coeficientes do sistema.

Entre as várias fontes de erro que contribuem para que as soluções obtidas por via numérica difiram das reais, é de salientar os erros de arredondamento. Estes assumem, por vezes, um papel importante no ajuste de dados experimentais, quando se obtêm sistemas de equações mal condicionadas (sistemas em que pequenas variações nos valores dos coeficientes originam soluções bastante diferentes). Neste caso, é necessário uma análise numérica do problema. Assim, é importante que se tenha em atenção o risco que se corre quando se utilizam programas cujo conteúdo se desconhece.

NOTA: Os resultados apresentados foram obtidos com programas testados e corridos num microcomputador PC compatível.

A subrotina GRÁFICO cria um ficheiro de pontos experimentais e calculados com o qual se obtêm os gráficos recorrendo a um programa de traçado de gráficos, nomeadamente ao programa GRAPHER (Golden Software, Inc., 1986).

Referências

- [1] F.F.M. Freitas, "Ajuste de Dados Experimentais a uma Função", relatório apresentado no âmbito das Provas de Aptidão Pedagógica e Capacidade Científica, Faculdade de Ciências, Lisboa, 1988.
- [2] A.C. Norris, "Computational Chemistry — An Introduction to Numerical Methods", 1.ª ed., John Wiley & Sons, Ltd, Chichester, 1981.
- [3] A. Constantinides, "Applied Numerical Methods with Personal Computers", 1.ª ed., McGraw-Hill Book Company, New York, 1987.
- [4] D.H. Everett e W.F.K. Wynne-Jones, Trans. Faraday Soc., 1939, 35, 1390.
- [5] C.A.N. Viana, L.M.P.C. Albuquerque e R.M.C. Gonçalves, Rev. Port. Quím., 1980, 22, 117.
- [6] L.M.P.C. Albuquerque, Tese de Doutoramento, Lisboa, 1979.

O Boletim tem uma nova rubrica

OPINIÃO