

PROGRAMA FITGL PARA DECOMPOSIÇÃO DE UM ESPECTRO COMPLEXO NAS RESPECTIVAS BANDAS COMPONENTES ⁽¹⁾

F. A. P. BASTO ⁽²⁾

J. A. N. F. GOMES

A. A. S. C. MACHADO*

Laboratório de Química
Faculdade de Ciências
Porto — Portugal

1 — INTRODUÇÃO

Em continuação de trabalho realizado previamente neste laboratório [1], e aproveitando alguma experiência adquirida previamente por um dos autores [2], escreveu-se um programa, FITGL, para a decomposição de um espectro nas respectivas bandas componentes. Procurou-se dar ao programa máxima versatilidade, pelo que se permite ao utilizador a fixação de parte dos parâmetros das funções componentes e, ainda, o ajuste ou fixação prévia de uma linha-base acima da qual é feita a interferência das várias funções componentes. Com esse programa fez-se um estudo sistemático da decomposição de um espectro constituído por duas bandas gaussianas parcialmente sobrepostas, iguais em intensidade e largura mas variavelmente afastadas uma da outra, com o fim de determinar as limitações do programa e de definir o melhor modo de o usar na decomposição de espectros electrónicos. Nesta nota apresenta-se uma versão do referido programa e as conclusões a que se chegou quanto às suas possibilidades e maneira de utilização.

2 — DISCUSSÃO DO PROGRAMA

Tal como o programa anterior [1], este programa faz, por iterações sucessivas, uma minimização dos erros pelo critério dos mínimos quadrados, após linearização da dependência entre o modelo adoptado e os parâmetros; não se apresentam aqui pormenores do método por ele ser bem conhecido [1, 3, 4]. O método tem, porém, uma séria limitação: por vezes as iterações divergem e, noutros casos, as iterações convergem muito lentamente (por vezes para uma solução que não é a pretendida). Isto resulta de, na prática, poderem estar grosseiramente erradas as correcções calculadas para os parâmetros, as quais só seriam exactas caso a função de erro a minimizar dependesse linearmente destes parâmetros. Por outro lado, em problemas de optimização simultânea de um elevado número de parâmetros, um programa deste tipo pode produzir grandes variações simultâneas de vários parâmetros sem que a melhoria do ajuste seja

⁽¹⁾ Adaptado, em parte, do relatório do trabalho de Seminário (ano lectivo de 1972/73) de F. A. P. Basto.

⁽²⁾ Endereço actual: Escola Industrial Infante D. Henrique, Porto, Portugal.

apreciável, e é extremamente difícil estabelecer um método de rejeição ou atenuação das correcções calculadas em cada iteração. No presente programa introduziram-se, porém, como opções, duas modificações do método que tentam anular a referida limitação, pois procuram obter convergência ou acelerar esta, conforme o caso, se o método não modificado falha.

No programa em discussão é permitido, como opção, impedir variações dos parâmetros superiores a 10 % do respectivo valor. Em certos casos, em que não se obtém convergência sem um condicionamento deste tipo, esta opção permite obtê-la porque impede afastamentos bruscos exagerados do conjunto inicial de valores dos parâmetros. Noutros casos, em que a convergência é lenta, o uso desta opção permite acelerá-la. Note-se porém, que foram encontrados casos em que o uso da opção tornou a convergência mais lenta; esta situação, que não é inesperada, ocorre, presumivelmente, ou quando o problema em estudo for suficientemente linear por sua própria natureza, ou quando os valores dos parâmetros iniciais fornecidos formarem um conjunto particularmente favorável quanto a minimização de erros.

Uma segunda opção consiste em minimizar o erro quadrático segundo a distância mínima do ponto ao espectro ajustado (isto é, segundo a perpendicular) e não segundo as ordenadas (fig. 1). Para os lados das bandas, em que as ordenadas variam muito rapidamente com as abcissas, é de esperar que a minimização segundo a perpendicular dê origem a um melhor ajustamento da curva calculada aos pontos experimentais. Esta opção poderá, portanto, ser útil para um refinamento final das soluções encontradas. Nesta opção, o erro segundo as ordenadas é projectado sobre a perpendicular ao espectro no ponto em consideração, o que dá um valor razoavelmente aproximado da distância mínima (fig. 1). Usou-se, no programa, o seguinte procedimento: o espectro na vizinhança imediata do ponto em consideração é aproximado por uma função parabólica (convoluta) [5], ajustada segundo um critério de mínimos quadrados a cinco pontos (dos quais o central é o ponto em consideração) e calcula-se a perpendicular à parábola na abcissa do ponto em consideração; finalmente, projecta-se o erro segundo as ordenadas na direcção desta perpendicular.

Uma outra vantagem deste programa quanto ao anterior [1] é a possibilidade de manter fixos alguns dos parâmetros, ao longo do ajustamento, se assim for indicado nos dados. Uma tal possibilidade é essencial num programa para decomposição de espectros, em que, se há duas bandas suficientemente sobrepostas, é impossível obter os seis parâmetros que as definem (se forem lorentzianas ou gaussianas puras): se se tenta ajustar todos estes parâmetros, a matriz das equações normais é singular (ou quase), sendo impossível a sua inversão para o prosseguimento do cálculo (cálculo das correcções aos parâmetros). Neste caso, podem-se, por exemplo, determinar quatro parâmetros, fixando, por exemplo, a posição e a largura da banda aparentemente predominante e deixando ajustar a sua altura bem como os três parâmetros

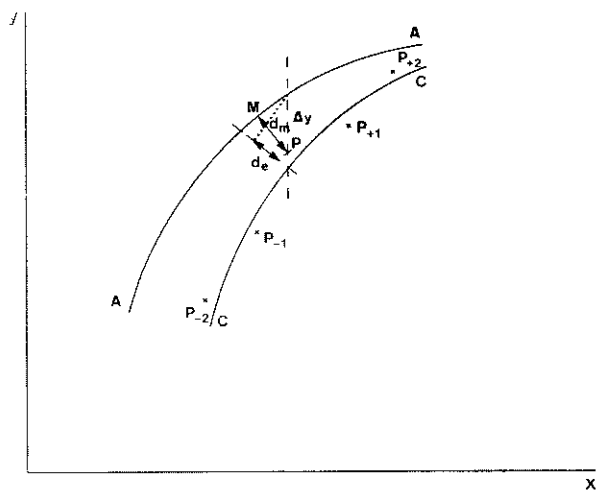


Fig. 1

Como opção, o programa FITGL minimiza $\Sigma(d_m)^2$ em que d_m é a distância dos pontos experimentais ao espectro (PM) em vez de $\Sigma(\Delta y)^2$, em que Δy é a diferença de ordenadas. A distância d_m é aproximada pela distância d_e , projecção de Δy sobre a perpendicular à convoluta CC no ponto cuja abcissa é a de P. A convoluta C é obtida por ajuste de uma parábola aos cinco pontos P_{-2} , P_{-1} , $P_0 \equiv P$, P_{+1} e P_{+2} . AA é o espectro aproximado até à iteração anterior à considerada

de uma segunda banda. Note-se que a impossibilidade da determinação dos seis parâmetros, num caso destes, resulta da existência de correlação entre eles: existem diversos conjuntos diferentes de parâmetros que permitem um bom ajustamento do modelo à curva experimental, pelo que há um certo grau de indeterminação nos valores obtidos.

Note-se que a referida possibilidade de fixar parâmetros tem utilidade mais geral e pode ser usada quando não se obtém convergência a partir de um certo conjunto de valores iniciais fornecidos para os parâmetros, devido a causas diferentes da existência de correlação: neste caso, a fixação de um ou mais dos parâmetros pode fazer convergir os restantes para valores mais próximos dos verdadeiros, que serão usados como ponto de partida num ajustamento final sem fixação do valor de qualquer dos parâmetros. Esta situação pode ocorrer se os valores fornecidos inicialmente para os parâmetros não forem suficientemente próximos dos que minimizam os erros. Esta é uma das causas mais vulgares para a divergência e resulta de não ser válida a linearização que se usa como base do método.

Relativamente ao programa anteriormente usado [1] o programa FITGL é mais geral, pois, conforme opção feita nos dados, permite fazer a decomposição de um espectro não só em bandas gaussianas como também em bandas lorentzianas, ou em bandas lorentzianas modificadas, ou em bandas mistura de gaussianas e lorentzianas (as expressões das funções respectivas são dadas na Tabela 1). Sucede, por vezes, que uma modificação ligeira do modelo adoptado (como a inclusão de um quarto parâmetro por banda, feita nos dois últimos casos) melhora radicalmente a qualidade do ajuste sem ter um efeito pronunciado na posição das bandas; uma tal modificação pode ser preferível à inclusão de mais bandas no modelo (cada uma das quais acrescenta três parâmetros a este), pois que neste caso ocorre correlação entre os parâmetros, como já se referiu.

Note-se que qualquer dos quatro tipos de banda é simétrico; não se incluíram bandas distorcidas por o seu interesse em espectroscopia ser raro [6, 7]. O uso de bandas distorcidas em cromatografia em fase gasosa tem merecido recentemente muita atenção, mas exige derivação numérica das funções [8, 4]. Por esta razão e por não se estar interessado, de momento, nesta técnica, tais bandas não foram incluídas no programa (note-se que a análise das distorções tem sido feita ultimamente por outros processos [9, 12], mas não é de prever que, na prática, estes processos venham a ter maior importância que o desdobramento por mínimos quadrados). Pelas mesmas razões também

Tabela 1

Tipos de banda usados no programa FITGL

Parâmetros da linha base:

- a_1 declive
- a_2 ordenada na origem

Para cada banda (i)

- a_3^i altura
- a_4^i largura a meia altura
- a_5^i posição
- a_6^i parâmetro de mistura ou modificador

a) Funções de Gauss

$$y = a_1(x - x_0) + a_2 + \sum_i a_3^i \exp [-(x - a_5^i)^2 / (a_4^i)^2 \ln 2]$$

b) Funções de Lorentz

$$y = a_1(x - x_0) + a_2 + \sum_i a_3^i / \{1 + [(x - a_5^i) / (a_4^i / 2)]^2\}$$

c) Funções mistura de Gauss e Lorentz

$$y = a_1(x - x_0) + a_2 + \sum_i [a_6^i y_L^i + (1 - a_6^i) y_G^i]$$

em que y_L^i e y_G^i representam, respectivamente, as bandas de Lorentz e Gauss e a_6^i é um parâmetro de mistura, entre 0 e 1.

d) Funções de Lorentz modificadas

$$y = a_1(x - x_0) + a_2 + \sum_i a_3^i \left\{ 1 + [(x - a_5^i) / (a_4^i / 2)]^{a_6^i} \right\}$$

em que a_6^i é um expoente próximo de 2

não se considerou o uso de bandas de forma arbitrária, definidas numericamente [13, 6]. Note-se, porém, que mediante ligeiras alterações no programa se pode fazer o ajuste a qualquer tipo de bandas definidas por um máximo de quatro parâmetros cuja expressão analítica (e suas derivadas em relação aos parâmetros) seja conhecida, como, por exemplo, as propostas por FRASER ou RUSCH [14, 15].

O programa foi escrito em FORTRAN IV, para o computador NCR4100 do Laboratório de Cálculo Automático da Faculdade de Ciências do Porto. O programa, bem como instruções para a sua utilização e amostras dos seus resultados, foram depositados pelos autores junto do Editor desta revista (1).

(1) Fotocópias dos mesmos serão enviadas (mediante pagamento de 100\$00 para despesas de reprodução e expedição) a quem as solicitar ao Editor, Revista Portuguesa de Química, Instituto Superior Técnico, Lisboa-1.

3 — OBSERVAÇÕES SOBRE O USO DO PROGRAMA

Com o fim de se avaliar as possibilidades e limitações do programa e de se ganhar sensibilidade para a sua utilização, sempre necessária [4] em problemas desta natureza, fez-se uma análise sistemática da decomposição de um espectro constituído por duas bandas gaussianas, da mesma intensidade e largura, mas sobrepostas parcialmente em extensão variável. Com este fim introduziu-se numa variante do programa uma sub-rotina extra geradora de espectros; esta sub-rotina simula um espectro composto por bandas de parâmetros dados, podendo, como opção, os pontos do espectro serem afectados de pequenos desvios, distribuídos aleatoriamente a fim de simular erros experimentais. Gerado o espectro por esta sub-rotina, o programa é usado para recuperar os valores verdadeiros dos parâmetros a partir de valores fornecidos pelo experimentador com uma aproximação previamente fixada. O caso estudado em que as duas bandas gaussianas são idênticas (têm a mesma intensidade e a mesma largura a meia altura) é relativamente mais simples do que os geralmente encontrados na prática mas, por isso mesmo, mais apropriado para um estudo das limitações mais importantes do programa. As bandas foram situadas na região $15\ 000\text{--}20\ 000\text{ cm}^{-1}$ e a largura (a meia altura) era da ordem de grandeza de alguns milhares de cm^{-1} , isto é, os valores usados para os parâmetros foram valores típicos em espectroscopia electrónica. Nos ensaios (cerca de sessenta) variaram-se os diversos factores de que pode depender a qualidade do ajustamento obtido, a fim de se verificar a influência das respectivas variações no resultado. Esses factores são: o afastamento das bandas, a diferença dos valores dos parâmetros iniciais fornecidos ao programa para os verdadeiros, o afastamento entre as abcissas dos pontos do espectro dados, a dispersão das ordenadas do mesmo e a fixação de alguns parâmetros nos valores iniciais fornecidos. O estudo feito permitiu tirar as conclusões que se passam a referir.

À medida que as duas bandas se vão aproximando (para além do limite de detecção visual da existência de duas bandas), a aproximação dos valores estimados dos parâmetros, fornecidos, aos valores verdadeiros, é cada vez mais crítica quanto ao

número de iterações em que se atinge o ajuste e, se a referida aproximação não for suficiente, as iterações não convergem. Nas condições dos ensaios feitos (iguais alturas e larguras das duas bandas) o parâmetro mais crítico a este respeito parece ser a posição das bandas, tendo-se verificado, também, que se as duas posições iniciais fornecidas forem interiores às verdadeiras é mais difícil obter convergência. Em qualquer das situações em que o programa falha pode-se obter convergência por fixação de um mínimo de parâmetros (isto é, pela manutenção dos valores iniciais estimados ao longo de todas as iterações); a qualidade do ajuste obtido dependerá, neste caso, do afastamento entre os valores fixados dos parâmetros e os respectivos valores verdadeiros. Por outro lado verificou-se que, se as duas bandas estão suficientemente próximas, o programa não as distingue de uma só, pois ajusta melhor uma só banda aos dados do que duas; mas isto só se verifica quando se aproximam as bandas bastante para além do limite em que elas deixam de ser visualmente distinguíveis, pelo que há um certo número de situações intermédias em que o ajuste computacional é o único processo capaz de fazer a identificação de duas bandas sobrepostas.

Um outro facto sugerido pelos resultados obtidos é que a diminuição do número de pontos fornecidos ao computador para ajustamento do espectro, isto é, o afastamento das respectivas abcissas não é tão crítico quanto pode parecer quanto à convergência do processo; a sua diminuição, até certos limites, não provoca divergência do ajuste, embora se vá reflectir na sua qualidade, no caso de haver dispersão aleatória dos pontos dados. Por exemplo, no caso de duas bandas gaussianas, sem dispersão aleatória (separadas de $2,2 \times 10^3\text{ cm}^{-1}$) conseguiu-se fazer o ajuste apenas com um número de pontos experimentais duplo do número de parâmetros, sendo necessárias cinco iterações; o aumento até quatro vezes no número de pontos fornecidos não melhorou a qualidade do ajuste, que era bom, mas exigiu progressivamente mais iterações até um número de doze, isto é, aumentou apreciavelmente o tempo de computação requerido. Por outro lado, quando se introduziu dispersão aleatória e se aumentou exageradamente o número de pontos obtidos, observou-se, por vezes, divergência que não ocorria quando o número de pontos era menor.

As experiências feitas confirmaram resultados anteriores [4] segundo os quais um número de pontos experimentais triplo ou quádruplo do número de parâmetros, com excepção de casos desfavoráveis, é suficiente para determinar os valores dos parâmetros com precisão.

4 — CONCLUSÕES

A experiência colhida sugere que se deve usar um processo de tentativa e erro na utilização do programa. Dum modo geral, deve-se começar por um ajuste sem fixação de qualquer parâmetro (a não ser a linha base, se esta for bem definida) e sem impor qualquer limitação às variações dos parâmetros. Se no decorrer do ajuste se verificar que estes são corrigidos para valores sem sentido, deve repetir-se o cálculo com a opção de impedir variações de parâmetros superiores a 10 % o seu valor. Se não se obtiver convergência, podem-se fixar os valores de parâmetros de estimativa mais fácil, deixando o programa actuar sobre os outros. Neste caso, para avaliar a extensão da correlação, deve-se repetir o cálculo com valores diferentes dos parâmetros fixados; pode-se, aliás, logo numa primeira fase, usar o programa com diversos conjuntos de valores estimados diferentes pois a ocorrência ou não ocorrência de convergência depende da proximidade entre os valores estimados e os valores a que corresponde erro mínimo.

Finalmente, não se deve esquecer que obter convergência é um problema matemático e apenas uma das partes do problema estatístico. A obtenção de convergência, quer seja fácil ou difícil, nada garante quanto, por exemplo, à validade do modelo adoptado ou, se este for o verdadeiro mas existir correlação dos parâmetros, quanto aos valores exactos destes, que se podem situar em intervalos mais ou menos amplos. Em consequência, o utilizador de qualquer programa deste tipo deve ter, dum modo geral, uma atitude de cepticismo quanto ao verdadeiro significado dos resultados.

AGRADECIMENTOS

J. A. N. F. Gomes e A. A. S. C. Machado agradecem ao Instituto de Alta Cultura, Lisboa, a concessão de subsídios de investigação pelo Projecto PQ3, em cujo âmbito este trabalho se insere. J. A. N. F. Gomes agradece ao Prof. Doutor Alberto Amaral numerosas discussões sobre o assunto, na fase inicial do trabalho. Os autores agradecem, também, ao Director do Laboratório de Cálculo Automático da Faculdade de Ciências do Porto a generosa concessão de tempo de computação e outras facilidades

Recebido 4. Março. 1975.

BIBLIOGRAFIA

- [1] AMARAL, A. M. S. C., *Rev. Port. Quim.*, **13**, 73 (1971).
- [2] MACHADO, A. A. S. C., «Ph. D. Thesis», Univ. London, 1971.
- [3] FRASER, R. D. B. e SUZUKI, E., *Anal. Chem.*, **38**, 1771 (1966).
- [4] ROBERTS, S. M., WILKINSON, D. H. e WALKER, L. R., *Anal. Chem.*, **42**, 886 (1970).
- [5] SAVITZKY, A. e GOLAY, M. J. E., *Anal. Chem.*, **36**, 1627 (1964).
- [6] KELLER, W. D., LUSEBRINK, T. R. e SEDERHOLM, C. H., *J. Chem. Phys.*, **44**, 782 (1966).
- [7] SIANO, D. B. e METZLER, D. E., *J. Chem. Phys.*, **51**, 1856 (1969).
- [8] GLADNEY, H. M., DOWDEN, B. F., e SWALEN, J. D., *Anal. Chem.*, **41**, 883 (1969).
- [9] GRUSHKA, E., MYERS, M. N., SCHETTLER, P. D. e GIDDINGS, J. C., *Anal. Chem.*, **41**, 889 (1969).
- [10] GRUSHKA, E., MYERS, M. N. e GIDDINGS, J. C., *Anal. Chem.*, **42**, 21 (1970).
- [11] GRUSHKA, E. e MONACELLI, G. C., *Anal. Chem.*, **44**, 484 (1972).
- [12] GRUSHKA, E., *Anal. Chem.*, **44**, 1733 (1972).
- [13] ANDERSON, A. H., GIBB, T. C. e LITTLEWOD, A. B., *Anal. Chem.*, **42**, 434 (1970).
- [14] FRASER, R. D. B. e SUZUKI, E., *Anal. Chem.*, **41**, 37 (1969).
- [15] RUSCH, P. F. e LELIEUR, J. P., *Anal. Chem.*, **45**, 1541 (1969).

