Capítulo VI – Ajuste dos Mínimos Quadrados

- Dados ajustáveis a uma linha recta
- Determinação das constantes A e B
- Incerteza nas medidas de y
- Incerteza na determinação de A e B
- Ajuste dos mínimos quadrados a outras curvas:
  - Função polinomial
  - Função exponencial
  - Regressão múltipla

112

Dados ajustáveis a uma linha recta

- Muitas das mais comuns e interessantes experiências de Física estão relacionadas com várias medições de duas grandezas físicas diferentes, a fim de se investigar a relação matemática entre elas. Por exemplo, é o que fazemos quando medimos tempos e alturas para verificar a relação

$$h = \frac{1}{2}gt^2$$

- Comecemos por considerar o caso em que suspeitamos que duas quaisquer variáveis físicas estão relacionadas pela equação de uma recta na forma

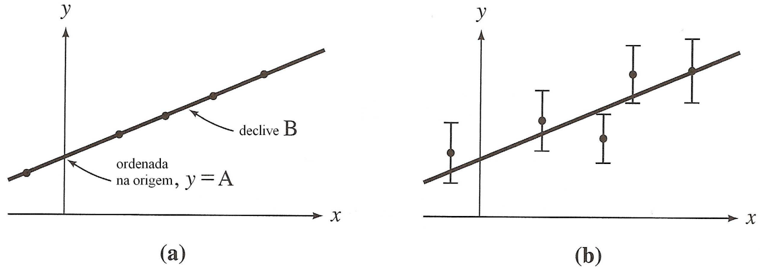
$$y = A + Bx \qquad (6.1)$$

- Num gráfico de y em função de x, devemos obter uma recta com declive B e que intersecta o eixo dos Y em y = A, como sabemos.

- É claro que, se fizéssemos N medidas diferentes de x (x_1, ..., x_N) e dos correspondentes y (y_1, ..., y_N), e se essas medidas não estivessem sujeitas a incertezas, cada um dos pares de pontos (x_i, y_i) estaria exactamente sobre a recta y = A + Bx (figura 6a).

113

Dados ajustáveis a uma linha recta (cont.)



- Na prática, há incertezas e o máximo que podemos esperar é que a distância entre cada ponto (x_i, y_i) e a recta seja razoável, como na figura 6b.
- Quando realizamos este tipo de medidas podemos pôr duas questões:

1. Se temos como garantido que x e y estão linearmente relacionadas, então o problema é encontrar a recta y = A + Bx que melhor se ajusta aos resultados experimentais, ou seja, encontrar as melhores estimativas para as constantes A e B com base nos dados (x_1, y_1), ..., (x_N, y_N).

   Este problema pode ser abordado graficamente e pode também ser tratado analiticamente, aplicando o princípio da máxima probabilidade. O método analítico para encontrar a recta que melhor se ajusta a uma série de dados experimentais é designado por ***regressão linear*** ou ***ajuste dos mínimos quadrados***. É o tema deste capítulo.

114

Dados ajustáveis a uma linha recta (cont.)

2. A 2ª questão que se pode pôr é se os valores medidos (x_1, y_1), ..., (x_N, y_N), estão realmente de acordo com as nossas expectativas de encontrar uma relação linear entre x e y.

   Para responder a esta dúvida, devemos, primeiro, encontrar a recta que melhor se ajusta aos dados experimentais e, depois, arranjar um modo de avaliar quão bem ela se ajusta a esses dados.

   Conhecendo a incerteza nas medidas, podemos fazer a representação gráfica dos dados, representando as incertezas associadas através de barras de erros (figura 6b). Juntamos a melhor recta de ajuste e avaliar visualmente a qualidade do mesmo.

   Contudo, se as incertezas nos dados não são conhecidas ou não estão devidamente estimadas, devemos fazer a avaliação examinando a distribuição dos próprios dados. Esse será o objecto de estudo de um próximo capítulo.

115

Determinação das constantes A e B

- Voltemos então à questão de encontrar a melhor recta y = A + Bx que se ajusta aos ponto medidos (x_1, y_1), ..., (x_N, y_N).

- Para simplificar admitiremos que a incerteza na medida de x é desprezável e que as incertezas nas medidas de y_i têm todas o mesmo valor. (Esta aproximação é razoável em muitas realizações experimentais.) Mais especificamente, assumiremos que cada y_i é governado por uma distribuição Gaussiana e que todas as distribuições têm o mesmo parâmetro largura σ_y.

- Se conhecêssemos as constantes A e B, então, para qualquer valor x_i (que considerámos não ter incerteza associada) poderíamos determinar o verdadeiro valor y_i:

$$(\text{verdadeiro valor de } y_i) = A + Bx_i.$$

- A medida de y_i segue uma distribuição normal centrada no seu verdadeiro valor e com largura σ_y. A probabilidade de obter o valor y_i observado é:

$$\mathrm{Prob}_{A,B}(y_i) \propto \frac{1}{\sigma_y} e^{-(y_i - A - Bx_i)^2/2\sigma_y^2} \qquad (6.2)$$

116

Determinação das constantes A e B (cont.)

$$\mathrm{Prob}_{A,B}(y_i) \propto \frac{1}{\sigma_y} e^{-(y_i - A - Bx_i)^2/2\sigma_y^2} \qquad (6.2)$$

- Os índices A e B lembram que a probabilidade depende dos valores (desconhecidos) de A e B.

- A probabilidade de obter a nossa série completa de medidas y_1, ..., y_N corresponde ao produto:

$$\mathrm{Prob}_{A,B}(y_1,\ldots,y_N) = \mathrm{Prob}_{A,B}(y_1)\ldots\mathrm{Prob}_{A,B}(y_N)$$
$$\propto \frac{1}{\sigma_y^N} e^{-\chi^2/2} \qquad (6.3)$$

onde o expoente é dado por:

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - A - Bx_i)^2}{\sigma_y^2} \qquad (6.4)$$

- Agora, de um modo que já nos é familiar, consideraremos que as melhores estimativas para as constantes desconhecidas A e B são os valores de A e B que maximizam a probabilidade na eq. 6.3, ou seja, minimizam a eq. 6.4 do χ^2.

117

Determinação das constantes A e B (cont.)

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - A - Bx_i)^2}{\sigma_y^2} \qquad (6.4)$$

- Para encontrarmos esses valores de A e B, diferenciamos a eq. 6.4 relativamente a A e B e igualamos as derivadas a zero:

$$\frac{\partial \chi^2}{\partial A} = \frac{-2}{\sigma_y^2} \sum_{i=1}^{N} (y_i - A - Bx_i) = 0 \qquad (6.5)$$

$$\frac{\partial \chi^2}{\partial B} = \frac{-2}{\sigma_y^2} \sum_{i=1}^{N} x_i (y_i - A - Bx_i) = 0 \qquad (6.6)$$

$$\boxed{AN + B\sum x_i = \sum y_i} \quad (6.7) \qquad \boxed{A\sum x_i + B\sum x_i^2 = \sum x_i y_i} \quad (6.8)$$

118

Determinação das constantes A e B (cont.)

- As equações 6.7 e 6.8 – por vezes designadas por *equações normais* – podem ser resolvidas de modo a obtermos as estimativas dos mínimos quadrados para as constantes A e B:

$$\boxed{A = \frac{\sum_i x_i^2 \sum_i y_i - \sum_i x_i \sum_i x_i y_i}{\Delta}} \qquad (6.9)$$

$$\boxed{B = \frac{N\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{\Delta}} \qquad (6.10)$$

onde $$\Delta = N\sum_i x_i^2 - \left(\sum_i x_i\right)^2 \qquad (6.11)$$

- Os resultados 6.9 e 6.10 correspondem às melhores estimativas para as constantes A e B da linha recta y = A + Bx, baseadas nos N pares de ponto medidos (x_1, y_1),..., (x_N, y_N). A recta resultante é designada por **ajuste dos mínimos quadrados** ou **regressão linear de y em x**.

119

Incerteza nas medidas de y

• Relembremos que os números y_1, ..., y_N não são N medidas do mesmo número. Portanto, não ficamos com uma ideia do grau de confiança examinando a sua dispersão.

• Apesar disso, podemos facilmente fazer uma estimativa da incerteza σ_y dos números y_1, ..., y_N. Estamos a considerar que a medida de cada y_i segue uma distribuição normal em volta do seu verdadeiro valor $A+Bx_i$, com largura σ_y. Assim, os desvios $y_i-(A+Bx_i)$ seguem também uma distribuição normal, todas centradas em zero e com a mesma largura σ_y.

• Este facto sugere imediatamente que uma boa estimativa para σ_y seria dada pelo somatório do quadrado dos desvios, na forma familiar:

$$\sigma_y = \sqrt{\frac{1}{N}\sum_i (y_i - A - Bx_i)^2} \qquad (6.12)$$

120

Incerteza nas medidas de y (cont)

$$\sigma_y = \sqrt{\frac{1}{N}\sum_i (y_i - A - Bx_i)^2} \qquad (6.12)$$

• Este resultado pode ser confirmado usando o princípio da máxima probabilidade. A melhor estimativa para o parâmetro σ_y é o valor para o qual a probabilidade de obter os valores observados y_1, ..., y_N (eq. 6.3) é máxima. Diferenciando a eq. 6.3 relativamente a σ_y e igualando a derivada a zero obtém-se a eq. 6.12.

• Acresce, contudo, que como os números A e B da eq. 6.12, são valores desconhecidos e devem ser substituídos pelas melhores estimativas de A e B, dadas pelas eq.s 6.9 e 6.10, esta substituição reduz ligeiramente o valor de 6.12.

• Pode mostrar-se que esta redução é compensada se substituirmos o factor N do denominador por (N – 2). Assim, o resultado final para a incerteza é:

$$\boxed{\sigma_y = \sqrt{\frac{1}{N-2}\sum_i (y_i - A - Bx_i)^2}} \qquad (6.13)$$

121

Incerteza nas medidas de y (cont)

Alguns comentários sobre o factor (N – 2)

• Se N for grande, a diferença entre N e (N – 2) não é importante.

• É razoável admitirmos o factor (N – 2) se considerarmos apenas dois pontos medidos: (x_1, y_1) e (x_2, y_2). Com apenas dois pontos é sempre possível arranjar uma recta que passe exactamente por esses dois pontos e o ajuste dos mínimos quadrados daria a equação dessa recta. Ou seja, com apenas dois pontos não é possível deduzir nada sobre o grau de confiança nos resultados. Como a recta fica exactamente sobre os dois pontos, os dois termos da soma nas eq.s 6.12 e 6.13 dão zero. Com N no denominador (eq. 6.12) obteríamos a resposta absurda $\sigma_y = 0$. Com (N – 2), obtemos a indeterminação $\sigma_y = 0/0$, o que indica, correctamente, que σ_y não se pode conhecer com apenas duas medidas.

• Na verdade, este factor é uma reminiscência do factor (N – 1) que aparece no desvio padrão de N medidas de uma mesma quantidade x. Nesse caso, para calcularmos σ_x, tivemos primeiro que determinar o valor médio $\overline{x}$. Em certo sentido, este cálculo deixa apenas (N – 1) valores medidos independentes. Podemos assim dizer que, tendo calculado $\overline{x}$, já só restam (N – 1) *graus de liberdade*.

122

Incerteza nas medidas de y (cont)

• No presente caso fizeram-se N medidas mas, antes de calcularmos σ_y, tivemos que determinar duas quantidades, A e B. Restam-nos, então, apenas (N – 2) graus de liberdade.

• Em geral, define-se **nº de graus de liberdade** em qualquer passo de um cálculo estatístico como o **nº de medidas independentes menos o nº de parâmetros calculados a partir desse nº de medidas**.

• Pode demonstrar-se (o que não faremos) que é o nº de graus de liberdade, e não o nº de medidas, que deve aparecer em fórmulas tais como a 6.13.

123

Incerteza nas constantes A e B

- Tendo encontrado a incerteza σ_y dos valores medidos $y_1, \ldots, y_N$, podemos calcular facilmente as incertezas associadas às estimativas de A e B.
- De facto, como as estimativas de A e B são funções bem definidas dos números medidos $y_1, \ldots, y_N$, as incertezas em A e B são obtidas por simples propagação de erros a partir das incertezas associadas aos y_i.
- Fica como exercício mostrar que:

$$\boxed{\sigma_A = \sigma_y \sqrt{\frac{\sum_i x_i^2}{\Delta}}} \qquad (6.14)$$

$$\boxed{\sigma_B = \sigma_y \sqrt{\frac{N}{\Delta}}} \qquad (6.15)$$

onde, como anteriormente, $\Delta = N\sum_i x_i^2 - \left(\sum_i x_i\right)^2$

124

Caso particular da recta que passa na origem: y = Bx

- Pode mostrar-se que, neste caso:

$$B = \frac{\sum_i x_i y_i}{\sum_i x_i^2} \qquad (6.16)$$

$$\sigma_y = \sqrt{\frac{1}{N-1}\sum_i (y_i - Bx_i)^2} \qquad (6.17)$$

$$\sigma_B = \frac{\sigma_y}{\sqrt{\sum_i x_i^2}} \qquad (6.18)$$

125

Incerteza nas constantes A e B (cont.)

• Os resultados apresentados até agora foram baseados na suposição de que só havia incerteza nos valores experimentais de y e que essa incerteza era igual para todos os y_i.

• Começando pelo último aspecto, o que acontece se as incertezas σ_i dos y_i forem diferentes? Usamos o método dos mínimos quadrados pesados: $w_i = \frac{1}{\sigma_i^2}$

$$A = \frac{\sum_i w_i x_i^2 \sum_i w_i y_i - \sum_i w_i x_i \sum_i w_i x_i y_i}{\Delta} \qquad \sigma_A = \sqrt{\frac{\sum_i w_i x_i^2}{\Delta}} \qquad (6.19)$$

$$B = \frac{\sum_i w_i \sum_i w_i x_i y_i - \sum_i w_i x_i \sum_i w_i y_i}{\Delta} \qquad \sigma_B = \sqrt{\frac{\sum_i w_i}{\Delta}} \qquad (6.20)$$

$$\Delta = \sum_i w_i \sum_i w_i x_i^2 - \left(\sum_i w_i x_i\right)^2 \qquad (6.21)$$

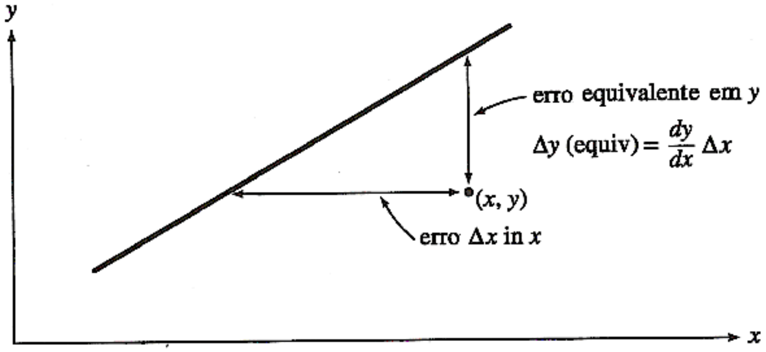
126

Incerteza nas constantes A e B (cont.)

• Analisemos agora o caso em que ambas as variáveis x e y vêm afectadas de incertezas.

• Comecemos por admitir que x vem afectado de erro mas y não, e tomemos um ponto particular medido (x, y) – figura 6-c). O ponto não está sobre a recta devido à incerteza Δx na medida de x.

• Vê-se, a partir do gráfico, que se admitíssemos que o valor de x era exacto e tomássemos antes a incerteza em y, obteríamos um resultado equivalente: o mesmo afastamento do ponto à recta. Podemos então escrever:



(c)

$$\Delta y(\text{equivalente}) = \frac{dy}{dx}\Delta x \qquad (6.22)$$

127

Incerteza nas constantes A e B (cont.)

- O desvio padrão σ_x é a raiz quadrada de Δx que resultaria de repetirmos a medida de x muitas vezes. Assim, podemos substituir Δx e Δy e escrever:

$$\sigma_y(\text{equivalente}) = \frac{dy}{dx}\sigma_x \qquad (6.23)$$

- Este resultado é verdadeiro qualquer que seja a curva y(x) mas é particularmente simples se se tratar de uma linha recta, porque o declive dy/dx é apenas a constante B. Assim, para uma linha recta:

$$\sigma_y(\text{equivalente}) = B\sigma_x \qquad (6.24)$$

- Em particular, se todas as incertezas σ_x são iguais, o mesmo é verdade para as incertezas equivalentes σ_y(equivalentes). Portanto, o problema de ajustar uma recta aos pontos (x_i, y_i) com iguais incertezas em x e sem incertezas em y é equivalente ao problema de ter iguais incertezas em y mas nenhuma em x.

(Na prática, os pontos não ficam exactamente sobre a recta e os dois problemas equivalentes não dão exactamente a mesma recta. Contudo, as linhas concordam dentro das incertezas dadas por 6.14 e 6.15.

128

Incerteza nas constantes A e B (cont.)

- Podemos agora estender o argumento ao caso em que ambos x e y têm incertezas.

- A incerteza em x é equivalente à incerteza em y dada por 6.24. Além disso, y também está sujeito à sua própria incerteza, σ_y. Estas duas incertezas são independentes e os seus quadrados devem ser adicionados.

- Assim, o problema original com incertezas em x e y, pode ser substituído por um problema equivalente no qual apenas y tem incertezas dadas por:

$$\sigma_y(\text{equivalente}) = \sqrt{\sigma_y^2 + (B\sigma_x^2)} \qquad (6.25)$$

- Se as incertezas σ_x forem iguais entre si e as incertezas σ_y forem também iguais entre si, então as incertezas resultantes da eq. 6.25 são também iguais e podemos usar as fórmulas 6.9 a 6.15.

- Se, pelo contrário, as incertezas em x ou em y não são iguais, a eq. 6.25 ainda é válida mas teremos que usar o método dos mínimos quadrados pesados.

129

Ajuste dos mínimos quadrados a outras curvas: Função Polinomial

$$\boxed{y = A + Bx + Cx^2 + \ldots + Hx^n} \qquad (6.26)$$

- Tomemos, para simplificar: $y = A + Bx + Cx^2$ (6.27) $\left(\text{Ex}: \ y = y_0 + v_0 t - \frac{1}{2} g t^2 \right)$
- Como anteriormente, suponhamos que temos uma série de medidas (x_i, y_i), i=1, …, N, com os x_i sem incertezas e os y_i igualmente incertos.
- Para cada x_i, o correspondente valor verdadeiro y_i é dado pela eq. 6.27, onde A, B e C são ainda desconhecidos.
- Consideramos que os y_i seguem distribuições normais, centradas no verdadeiro valor correspondente, e todas com a mesma largura σ_y.
- A probabilidade de obter os N valores observados $y_1, \ldots, y_N$ é :

$$\text{Prob}(y_1, \ldots, y_N) \propto e^{-\chi^2/2}$$

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(y_i - A - Bx_i - Cx_i^2\right)^2}{\sigma_y^2} \qquad (6.28)$$

130

Função Polinomial

- Aplicando o princípio da máxima probabilidade, derivando o expoente e igualando a zero obtemos as seguintes equações:

$$\begin{aligned} AN + B\sum_i x_i + C\sum_i x_i^2 &= \sum_i y_i \\ A\sum_i x_i + B\sum_i x_i^2 + C\sum_i x_i^3 &= \sum_i x_i y_i \\ A\sum_i x_i^2 + B\sum_i x_i^3 + C\sum_i x_i^4 &= \sum_i x_i^2 y_i \end{aligned} \qquad (6.29)$$

- Extraindo os parâmetros A, B e C obtemos o ajuste polinomial dos mínimos quadrados ou a regressão polinomial.
- Em princípio, um método semelhante pode ser aplicado a qualquer função f(x) que dependa de parâmetros desconhecidos A, B, etc. Embora alguns casos possam ser complexos, são sempre resolúveis aqueles que correspondem a funções que dependem linearmente dos parâmetros A, B, etc, como é o caso, por exemplo, de y = A sin x + B cos x.

131

Ajuste dos mínimos quadrados a outras curvas: Função Exponencial

$$\boxed{y = Ae^{Bx}} \qquad (6.30)$$

• Para podermos determinar as constantes A e B da forma resolúvel que conhecemos, transformamos a eq. 6.30, não linear, numa relação linear entre x e y, à qual possamos aplicar o ajuste dos mínimos quadrados:

$$\ln y = \ln A + Bx$$

$$z = \ln A + Bx$$

$$(x_i, y_i) \longrightarrow (x_i, z_i)$$

O método os mínimos quadrados dá-nos a melhor estimativa para as constantes ln A (de onde extraímos A) e B.

132

Função Exponencial

• Na verdade, quando as incertezas em y são iguais, as incertezas em z não o são:

$$\sigma_z = \left|\frac{dz}{dy}\right|\sigma_y = \frac{\sigma_y}{y}$$

Em rigor, deveríamos usar, portanto, o método dos mínimos quadrados pesados.

• Na prática, muitas vezes a dispersão das incertezas em z não é grande e, portanto, assumem-se incertezas iguais e aplica-se o método normal.

133

Ajuste dos mínimos quadrados a outras curvas: Múltipla regressão

- Até aqui considerámos apenas a relação entre 2 variáveis. E se existirem mais? (Exemplo: PV=nRT)

- Tomemos $z = A + Bx + Cy$ e consideremos que numa série de medidas (x_i, y_i, z_i), com i = 1, …, N, os x_i e y_i são exactos e os z_i têm associadas iguais incertezas.

- Aplicando o princípio da máxima probabilidade como anteriormente, concluímos que as constantes A, B e C podem ser determinadas a partir de:

$$AN + B\sum_i x_i + C\sum_i y_i = \sum_i z_i$$
$$A\sum_i x_i + B\sum_i x_i^2 + C\sum_i x_i y_i = \sum_i x_i z_i$$
$$A\sum_i y_i + B\sum_i x_i y_i + C\sum_i y_i^2 = \sum_i y_i z_i$$

- Este método é conhecido por *regressão múltipla.*

134