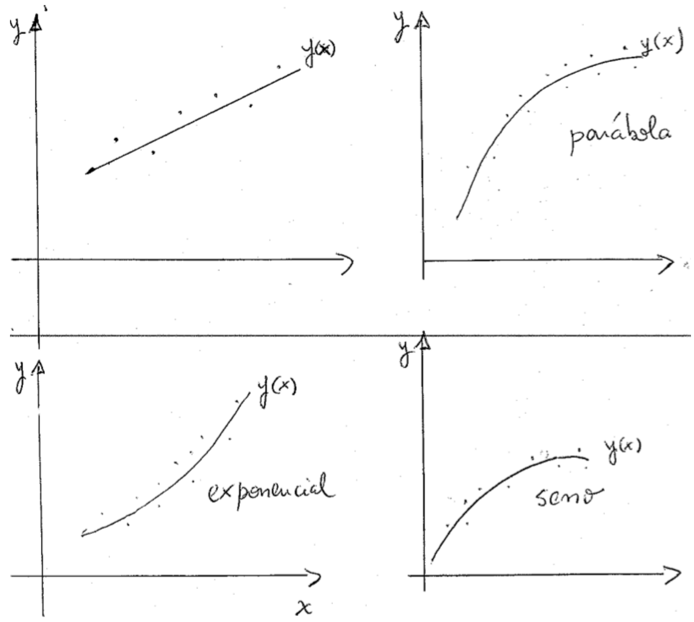

Figura 7.1: O problema do ajuste de funções a um conjunto de dados

Capítulo 7

Aproximação de Funções por Mínimos Quadrados

7.1 Introdução

Dado um conjunto de observações (dados), frequentemente deseja-se condensar os dados ajustando a eles um modelo que depende de parâmetros ajustáveis. Às vezes o modelo é simplesmente uma classe conveniente de funções, tais como polinômios, exponenciais, Gaussianas, etc. e o ajuste visa simplesmente obter os coeficientes apropriados (Fig. 7.1). Em outras situações, os parâmetros do modelo vêm de alguma teoria subjacente que os dados devem necessariamente satisfazer. Por exemplo, os dados podem ser a posição no céu de um asteróide em função do tempo e os parâmetros a serem ajustados os elementos orbitais da órbita do asteróide.

O procedimento de ajuste de dados por uma função envolve definirmos uma *função de mérito*, que mede a concordância entre os dados e o modelo com uma dada escolha de parâmetros. Em estatística frequentista, a função de mérito é em geral escolhida de forma que pequenos valores da função de mérito representam uma boa concordância entre os dados e a função ajustada. Dessa forma, o ajuste de funções é um problema de minimização em múltiplas dimensões.

Há outras questões que vão além de se encontrar o melhor ajuste. Dados são, em geral, inexatos, ou seja, estão associados a erros (ou ruídos, no contexto de processamento de sinais). Assim, dados típicos nunca ajustam exatamente o modelo adotado, mesmo quando o modelo é correto. Necessitamos, assim, de meios para avaliar se o modelo é ou não apropriado, ou seja, precisamos testar a qualidade do ajuste usando-se algum tipo de teste estatístico.

Outra questão é que, além de obtermos os parâmetros do nosso melhor ajuste, devemos também determinar a acurácia com a qual estes parâmetros foram obtidos. Em termos frequentistas, devemos obter os erros associados a estes parâmetros.

Como aproximar uma função (modelo) a um conjunto de pontos?

Suponhamos um conjunto de pontos (x_i, y_i), $i = 1, \ldots, m$, que desejamos aproximar por uma função $y = f(x)$, escrita como uma combinação linear de uma família de funções, f_i, de forma que

$$y(x) = c_1 f_1(x) + c_2 f_2(x)_+ \ldots + c_n f_n(x)\,,$$

onde os c_i, $i = 1, \ldots, n$ são os parâmetros livres do problema. Note que o número de parâmetros livres, n, deve ser menor ou igual ao número de pontos, m, e que escolha da família de funções depende da natureza do problema. A aproximação consiste em achar os parâmetros livres que minimizem uma função de mérito, que, como vimos, é uma função que mede a qualidade da aproximação.

Vamos definir uma função de mérito (que chamaremos resíduo), da seguinte forma

$$r = \sum_{i=1}^{n} r_i = \sum_{i=1}^{n} [y(x_i) - y_i]\,. \tag{7.1}$$

Assim, uma boa aproximação seria obtida minimizando o resíduo, por exemplo impondo que $r = 0$. Analisemos tal função de mérito com um exemplo simples. Suponhamos que foi realizado um experimento em que se obtiveram os pontos (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), mostrados na Figura 7.4. Pode-se observar que todas as retas que foram traçadas na Figura obedecem ao critério $r = 0$, pois os pontos 1 e 2 são simétricos em relação aos pontos 3 e 4 para todas elas. Isso mostra que minimizar o resíduo não é uma boa escolha para se aproximar uma função. Portanto, a função (7.1) não é uma boa função de mérito.

O problema com esta função reside no fato dos resíduos r_i serem tanto positivos quanto negativos. Assim, se deixarmos de considerar o sinal dos erros evitaremos este problema. Uma forma de fazer isso é considerar o módulo dos erros, mais isso introduziria dificuldades matemáticas indesejadas. Outro critério com esta mesma característica, porém com tratamento matemático mais simples, é exigir que

$$\sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} [y(x_i) - y_i]^2 \tag{7.2}$$

seja mínimo. Este é o famoso critério dos mínimos quadrados.

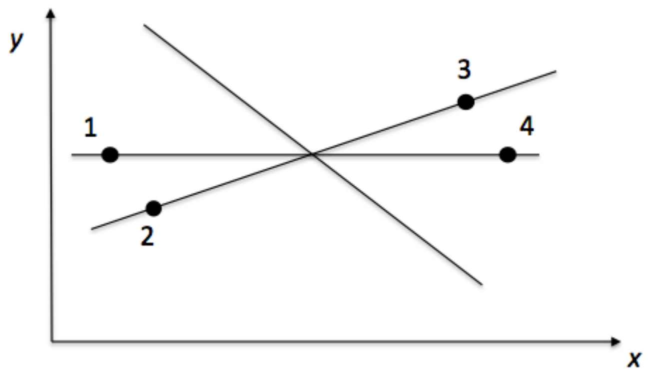

Figura 7.2: Exemplo que ilustra porque o resíduo não é uma boa função de mérito.

7.2 Critério dos Mínimos Quadrados

7.2.1 Sem pesos

Nosso objetivo é aproximar a função geral

$$y(x) = c_1 f_1(x) + c_2 f_2(x)_+ \ldots + c_n f_n(x) \,,$$

com n parâmetros livres, a um conjunto de m pontos usando-se o critério dos mínimos quadrados como função de mérito. Faremos isso minimizando a quantidade

$$\chi^2 \equiv \sum_{i=1}^{m} [y(x_i) - y_i]^2 = \sum_{i=1}^{m} [c_1 f_1(x_i) + c_2 f_2(x_i) + \ldots + c_n f_n(x_i) - y_i]^2 \,, \tag{7.3}$$

chamada de "qui-quadrado".

O χ^2 será mínimo quando as n derivadas parciais

$$\frac{\partial \chi^2}{\partial c_1}, \frac{\partial \chi^2}{\partial c_2}, \ldots, \frac{\partial \chi^2}{\partial c_n} \tag{7.4}$$

se anularem simultaneamente. Vamos escrever de forma explícita esta condição para o j-ésimo coeficiente

$$\frac{\partial \chi^2}{\partial c_j} = 2 \sum_{i=1}^{m} \left[(y(x_i) - y_i) \frac{\partial y(x_i)}{\partial c_j} \right] = 0 \,.$$

Temos que

$$\frac{\partial y(x_i)}{\partial c_j} = f_j(x_i)$$

portanto

$$\frac{\partial \chi^2}{\partial c_j} = 2 \sum_{i=1}^{m} \left[(y(x_i) - y_i) f_j(x_i) \right] = 0 \,,$$

ou ainda

$$\frac{\partial \chi^2}{\partial c_j} = \sum_{i=1}^{m} [(c_1 f_1(x_i) + c_2 f_2(x_i) + \ldots + c_n f_n(x_i) - y_i) f_j(x_i)] = 0 \,.$$

Fazendo a multiplicação e separando os somatórios, temos, finalmente, a seguinte expressão para o j-ésimo coeficiente

$$c_1 \sum_{i=1}^{m} f_1(x_i) f_j(x_i) + c_2 \sum_{i=1}^{m} f_2(x_i) f_j(x_i) + \ldots + c_n \sum_{i=1}^{m} f_n(x_i) f_j(x_i) = \sum_{i=1}^{m} f_j(x_i) y_i \,.$$

que, vale lembrar, vem do fato de termos imposto que a derivada parcial do chi-quadrado com relação a c_j deve ser zero.

Desta maneira, a condição (7.4) define um sistema de n equações, que deve ser resolvido para encontrar os parâmetros do ajuste:

$$\begin{bmatrix} \sum_{i=1}^{m} (f_1(x_i))^2 & \sum_{i=1}^{m} f_1(x_i) f_2(x_i) & \ldots & \sum_{i=1}^{m} f_1(x_i) f_n(x_i) \\ \sum_{i=1}^{m} f_2(x_i) f_1(x_i) & \sum_{i=1}^{m} (f_2(x_i))^2 & \ldots & \sum_{i=1}^{m} f_2(x_i) f_n(x_i) \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{m} f_n(x_i) f_1(x_i) & \sum_{i=1}^{m} f_n(x_i) f_2(x_i) & \ldots & \sum_{i=1}^{m} (f_n(x_i))^2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} f_1(x_i) y_i \\ \sum_{i=1}^{m} f_2(x_i) y_i \\ \vdots \\ \sum_{i=1}^{m} f_n(x_i) y_i \end{bmatrix} .$$

Este sistema é usualmente chamado de sistema padrão do problema dos mínimos quadrados. Pode-se mostrar que a matriz de tal sistema tem determinante não nulo, mas ele pode ser (e frequentemente o é) mal-condicionado, de forma que a solução pode ser difícil, ou impossível, de se obter.

Exemplo 1: ajuste de uma reta

Se a função a ser ajustada for uma reta, temos que

$$y(x) = c_1 f_1(x) + c_2 f_2(x) \,,$$

onde $f_1(x) = 1$ e $f_2(x) = x$. Os coeficientes serão dados pelo sistema de ordem 2

$$\begin{bmatrix} \sum_{i=1}^{m} 1 & \sum_{i=1}^{m} x_i \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} x_i y_i \end{bmatrix} .$$

Como exemplo, vamos ajustar uma reta ao conjunto de pontos

x	10	20	30	40	50	60	70	80
y	2	5	6	7	10	13	14	15

A tabela abaixo nos ajuda a calcular os coeficientes do sistema padrão:

i	x_i	y_i	$x_i y_i$	x_i^2
1	10	2	20	100
2	20	5	100	400
3	30	6	180	900
4	40	7	280	1600
5	50	10	500	2500
6	60	13	780	3600
7	70	14	980	4900
8	80	15	1200	6400
Σ	360	72	4040	20400

Obtemos o seguinte sistema linar

$$\begin{bmatrix} 8 & 360 \\ 360 & 20400 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 72 \\ 4040 \end{bmatrix} ,$$

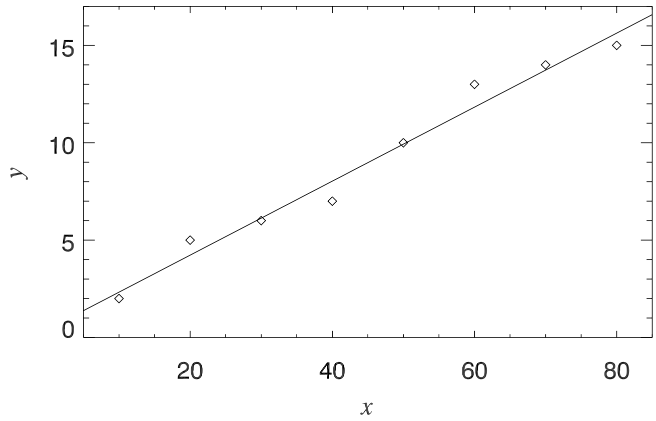

Figura 7.3: Ajuste de uma reta aos pontos listados no exemplo 1.

cuja solução é

$$\begin{aligned} c_1 &= 0.428 \\ c_2 &= 0.190\,. \end{aligned}$$

Desta forma, o melhor ajuste para os pontos dados pelo critério dos mínimos quadrados é a reta $y(x) = 0.428 + 0.190x$.

Exemplo 2: ajuste de uma exponencial

Como ajustar funções não lineares? Algumas funções podem ser rescritas de forma a obter uma outra expressão equivalente e linear. Um exemplo é a função exponencial com dois parâmetros livres

$$y(x) = ae^{-bx}\,,$$

que pode ser reescrita como

$$\ln[y(x)] = \ln(a) - bx \equiv c + dx\,,$$

que é linear nos seus parâmetros c e d.

Importante: preste atenção nos "não-parâmetros", como no caso da função

$$y(x) = ae^{-bx+c}\,.$$

Neste caso, os parâmetros a e c são indistinguíveis. Neste caso, o sistema normal será singular, ou seja, não terá solução (ou terá infinitas soluções).

Exercício:

Ajuste os pontos abaixo a uma exponencial do tipo $y(x) = ae^{-bx}$.

x	0	0.5	1	2
y	2.48	0.901	0.292	3.75×10^{-2}

Resposta: $a = 2.48$ e $b = 2.1$.

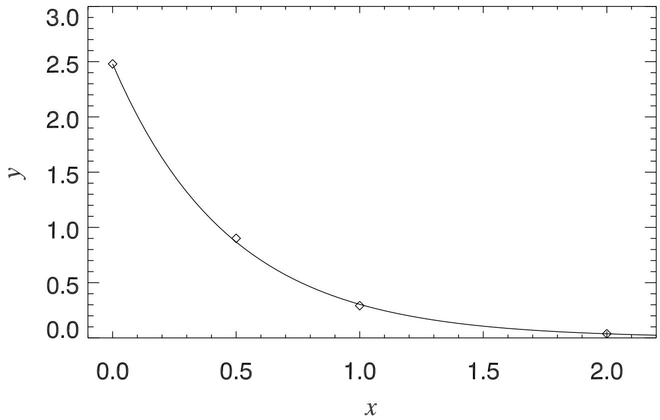

Figura 7.4: Ajuste de uma exponencial aos pontos listados no exemplo 2.

7.2.2 Com pesos

Vamos novamente considerar o caso em que estamos ajustando m pontos observacionais (x_i, y_i), $i = 1, \ldots, m$, com um modelo que tem n parâmetros ajustáveis (c_j, $j = 1, \ldots, n$). Consideremos a questão: "Para um determinado conjunto de parâmetros, qual a probabilidade de que aqueles pontos observacionais em particular tivessem ocorrido?". Se os valores y_i pertencem ao domínio dos reais, estão esta probabilidade é nula, a não ser que tivéssemos adicionado na questão acima a seguinte frase "mais ou menos algum Δy, pequeno e fixo". Vamos, abaixo, sempre considerar tal frase implicitamente. Se a probabilidade de se obter o conjunto de parâmetros for pequena, então podemos concluir que os parâmetros são "improváveis". Ao contrário, intuitivamente sabemos que o conjunto de parâmetros não deveria ser improvável se a escolha tiver sido correta.

Para sermos mais quantitativos, suponhamos que cada ponto y_i tenha um erro observacional (ou de medida) que seja independente das outras observações e que possua uma distribuição Gaussiana (normal) em torno do modelo "verdadeiro" $y(x)$ com um dado desvio padrão σ_i. Desta forma, a probabilidade de, em um determinado experimento, se fazer uma medida de valor y_i com desvio padrão σ_i, para um dado x_i, é

$$P_i(c_1, ..., c_n) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{y_i - y(x_i)}{\sigma_i}\right]^2\right)$$

onde o termo $1/\sigma_i\sqrt{2\pi}$ vem do fato de que a integral da probabilidade entre $-\infty$ e ∞ deve ser 1. Recorde que, na expressão acima, $y(x_i)$ representa o valor verdadeiro, ou esperado, para a observação.

A justificativa para assumirmos uma Gaussiana como função de distribuição está no Teorema do Limite Central que afirma (grosso modo) que a densidade de probabilidade de uma variável assume forma Gaussiana se a variável é ela mesma resultante de um grande número de subvariáveis aditivas independentes (ref. Callen).

A probabilidade de obter um conjunto de m pontos observacionais é o produto das probabilidades de cada observação:

$$P(c_1, ..., c_n) = \prod_{i=1}^{m} P_i(c_1, ..., c_n) = \left[\prod_{i=1}^{m}\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right)\right] \exp\left(-\frac{1}{2}\sum_{i=1}^{m}\left[\frac{y_i - y(x_i)}{\sigma_i}\right]^2\right), \quad (7.5)$$

onde o produto de exponenciais foi expresso como a soma dos argumentos. Nesses produtos e somas, quantidades como $1/\sigma_i^2$ atuam como pesos, expressando a contribuição relativa de cada ponto observacional para o resultado final: quando menor for o erro (desvio padrão) maior a importância do dado para o resultado do ajuste.

Vamos assumir que os dados observados são mais prováveis de serem obtidos a partir da distribuição "verdadeira" do que qualquer outra distribuição semelhante com diferentes parâmetros c_j e, portanto, a probabilidade da Eq. (7.5) deve ser máxima. Assim, a estimativa de máxima verossimilhança ("*maximum likehood*") para os c_j são os valores que maximizam a probabilidade da equação (7.5). Pelo fato de o primeiro fator produtório da equação (7.5) ser constante, independente dos valores de c_j, maximizar a probabilidade $P(c_1, ..., c_n)$ é equivalente a minimizar a soma do argumento da exponencial, ou seja, devemos minimizar o χ^2, definido como

$$\chi^2 = \sum_{i=1}^{m} \left[\frac{y_i - y(x_i)}{\sigma_i} \right]^2 .$$

Lembrando que $y(x)$ é escrito de forma geral como $y(x) = c_1 f_1(x) + c_2 f_2(x)_+ \ldots + c_n f_n(x)$, obtemos

$$\chi^2 = \sum_{i=1}^{m} \left[\frac{y_i - c_1 f_1(x) - c_2 f_2(x) - \ldots - c_n f_n(x)}{\sigma_i} \right]^2 . \tag{7.6}$$

Assim, a tarefa de ajuste dos dados será a de encontrar valores c_j que minimizem a soma ponderada dos quadrados dos termos de χ^2.

De forma geral, os valores de χ^2 são afetados por:

- Flutuação nos valores medidos y_i.

- Valores das incertezas σ_i: valores incorretos de σ_i levarão a valores incorretos de χ^2.

- A seleção da função analítica $y(x)$.

Minimização de χ^2: ajuste de uma reta

Sem perda de generalidade, vamos considerar o caso particular em que aproximamos os pontos por uma reta, ou seja, $y(x) = c_1 + c_2 x$. Como feito na seção anterior, para encontrar os parâmetros c_1 e c_2 que fornecem o mínimo valor para χ^2 igualamos a zero as derivadas parciais de χ^2 com respeito a cada um deles:

$$\frac{\partial \chi^2}{\partial c_1} = \frac{\partial}{\partial c_1} \sum_{i=1}^{m} \left[\frac{1}{\sigma_i^2} (y_i - c_1 - c_2 x_i)^2 \right] = -2 \sum_{i=1}^{m} \left[\frac{1}{\sigma_i^2} (y_i - c_1 - c_2 x_i) \right] = 0 \,,$$

$$\frac{\partial \chi^2}{\partial c_2} = \frac{\partial}{\partial c_2} \sum_{i=1}^{m} \left[\frac{1}{\sigma_i^2} (y_i - c_1 - c_2 x_i)^2 \right] = -2 \sum_{i=1}^{m} \left[\frac{x_i}{\sigma_i^2} (y_i - c_1 - c_2 x_i) \right] = 0 \,.$$

Essas equações podem ser rearranjadas como um sistema de equações lineares

$$\begin{aligned} c_1 \sum_{i=1}^{m} \frac{1}{\sigma_i^2} + c_2 \sum_{i=1}^{m} \frac{x_i}{\sigma_i^2} &= \sum_{i=1}^{m} \frac{y_i}{\sigma_i^2} \\ c_1 \sum_{i=1}^{m} \frac{x_i}{\sigma_i^2} + c_2 \sum_{i=1}^{m} \frac{x_i}{\sigma_i^2} &= \sum_{i=1}^{m} \frac{x_i y_i}{\sigma_i^2} \,. \end{aligned}$$

Usando a regra de Cramer

$$c_1 = \frac{1}{\Delta}\begin{vmatrix} \sum_{i=1}^m \frac{y_i}{\sigma_i^2} & \sum_{i=1}^m \frac{x_i}{\sigma_i^2} \\ \sum_{i=1}^m \frac{x_i y_i}{\sigma_i^2} & \sum_{i=1}^m \frac{x_i}{\sigma_i^2} \end{vmatrix} = \frac{1}{\Delta}\left(\sum_{i=1}^m \frac{x_i}{\sigma_i^2}\sum_{i=1}^m \frac{y_i}{\sigma_i^2} - \sum_{i=1}^m \frac{x_i}{\sigma_i^2}\sum_{i=1}^m \frac{x_i y_i}{\sigma_i^2}\right),$$

$$c_2 = \frac{1}{\Delta}\begin{vmatrix} \sum_{i=1}^m \frac{1}{\sigma_i^2} & \sum_{i=1}^m \frac{y_i}{\sigma_i^2} \\ \sum_{i=1}^m \frac{x_i}{\sigma_i^2} & \sum_{i=1}^m \frac{x_i y_i}{\sigma_i^2} \end{vmatrix} = \frac{1}{\Delta}\left(\sum_{i=1}^m \frac{1}{\sigma_i^2}\sum_{i=1}^m \frac{x_i y_i}{\sigma_i^2} - \sum_{i=1}^m \frac{x_i}{\sigma_i^2}\sum_{i=1}^m \frac{y_i}{\sigma_i^2}\right),$$

onde

$$\Delta = \begin{vmatrix} \sum_{i=1}^m \frac{1}{\sigma_i^2} & \sum_{i=1}^m \frac{x_i}{\sigma_i^2} \\ \sum_{i=1}^m \frac{x_i}{\sigma_i^2} & \sum_{i=1}^m \frac{x_i^2}{\sigma_i^2} \end{vmatrix} = \sum_{i=1}^m \frac{1}{\sigma_i^2}\sum_{i=1}^m \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^m \frac{x_i}{\sigma_i^2}\right)^2.$$

Para o caso particular onde todas as incertezas são iguais $\sigma_i = \sigma,\ i = 1, \ldots, m$ as fórmulas se reduzem às já estudadas sem pesos.

As incertezas dos parâmetros são dadas por (Bevington & Robinson)

$$\sigma_{c_1}^2 = \frac{1}{\Delta}\sum_{i=1}^N \frac{x_i^2}{\sigma_i^2}$$

$$\sigma_{c_2}^2 = \frac{1}{\Delta}\sum_{i=1}^N \frac{1}{\sigma_i^2}.$$

Exemplo 3: ajuste de uma reta

Seja o conjunto de pontos

x	10	20	30	40	50	60	70	80
y	2	12	6	7	10	13	7	15
σ_y	0.5	2.8	0.4	0.6	0.6	0.6	3.5	0.7

Neste conjunto, os pontos (20, 12) e (70, 7) são *outliers*, ou seja, estão fora da tendência geral indicada pelos outros pontos. Por este motivo, têm um erro associado muito maior que os demais. Se fizermos um ajuste destes pontos sem considerar os erros, obtemos o resultado mostrado na Figura 7.5-a. Vemos que este ajuste não descreve bem a tendência geral esperada quando consideramos apenas os pontos com menor erro. O procedimento correto é fazermos o ajuste levando em conta os erros. O resultado é mostrado na Figura 7.5-b. Os valores obtidos para os parâmetros livres são

$$c_1 = -0.10 \pm 0.32$$

$$c_2 = 0.199 + / - 0.004\,.$$

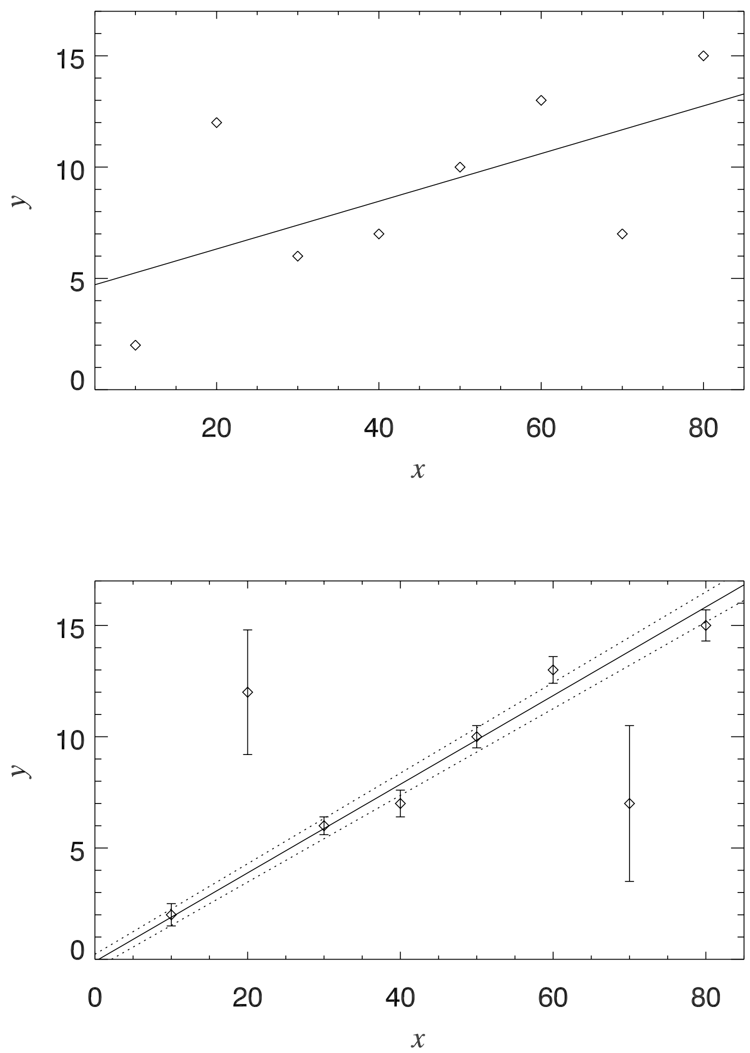

Figura 7.5: a) Acima: ajuste sem considerar os erros dos dados. b) Abaixo: ajuste feito considerando os erros. As linhas pontilhadas correspondem aos limites do erro estimado para o ajuste, e foram traçadas considerando os maiores e menores valores de c_1 e c_2 dentro da incerteza.