

# Eléments de régression linéaire

Andreas Stahel, Pierre–André Chevalier

Mai 2012

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Détermination des coefficients de la droite</b>	<b>2</b>
2.1	Exemple numérique	2
2.2	Le cas général	4
<b>3</b>	<b>Fondements statistiques</b>	<b>5</b>
3.1	Quelques notions de calcul des probabilités	5
3.1.1	Distribution normale	5
3.1.2	Estimation de la moyenne et de l'écart-type à partir des données	7
3.2	La méthode de Maximum–Likelihood	8
3.3	Estimation des variances de $a$ et $m$	9
3.3.1	Détermination de l'écart-type $\sigma$ des mesures $y_i$	10
3.3.2	Variance d'une somme de variables aléatoires, variances de $a$ et de $m$	10
<b>4</b>	<b>Compléments</b>	<b>11</b>
4.1	Régression linéaire générale, notation matricielle	11
4.2	Exemple : Intensity of light of an LED depending on the angle of observation	13
4.2.1	Description of the problem	13
4.2.2	A first, naive solution and its problems	13
4.2.3	How to eliminate the problem	15
4.2.4	Using the command <code>LinearRegression()</code>	17
4.3	Vérification des équations	18
4.3.1	Construction des équations	18
4.3.2	Notation vectorielle et matricielle	18
4.3.3	Codage efficace, décomposition <b>QR</b> et <i>Octave</i>	19
4.4	Mesures de précisions différentes	21
4.4.1	Droite de régression	21
4.4.2	Régression générale	22
4.5	Adaptation d'une courbe exponentielle	22
4.6	Erreur dans les valeurs de $x$ et $y$	24
4.7	Instruction <i>Octave</i> /MATLAB <code>LinearRegression()</code>	24

# 1 Introduction

Connaissant les positions de  $n$  points  $(x_i, y_i)$ ,  $i = 1 \dots n$  dans le plan, on s'intéresse à déterminer "la" droite d'équation  $y = a + m \cdot x$  qui passe "le mieux possible" par ces points. Il s'agit alors de déterminer les valeurs des coefficients  $a$  et  $m$  qui correspondent à la droite la plus appropriée. Lorsqu'il n'y a que deux points, l'exercice est très facile. En revanche, le problème est plus compliqué lorsque les points sont plus nombreux, car ceux-ci ne sont en général pas exactement alignés sur une droite. Même graphiquement, il est très difficile de trouver la droite optimale lorsque les points sont très dispersés.

x =	0.0	1.0	2.0	3.5	4.0
y =	-0.5	1.0	2.4	2.0	3.1

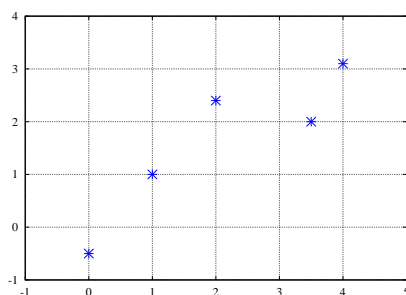


Figure 1: Une distribution de points dans le plan

Le but est de faire passer la droite d'équation

$$f(x) = a + m \cdot x$$

le mieux possible à travers la distribution de points. Une variante de la droite "la meilleure possible" est décrite par les conditions suivantes:

Trouver  $a$  et  $m$  de telle sorte que la somme  $g := \sum_{i=1}^n (f(x_i) - y_i)^2$  soit minimale.

La somme des carrés des distances verticales entre les points et la droite doit être rendue le plus petit possible. Pour cette raison, cette méthode s'appelle parfois la **méthode des moindres carrés**<sup>1</sup>. Le plus souvent, on emploie le terme de **régression linéaire**.

## 2 Détermination des coefficients de la droite

Dans cette section, nous allons développer la procédure de calcul permettant de déterminer les paramètres recherchés  $a$  et  $b$ . Nous allons pour commencer effectuer les calculs sur la base d'un exemple numérique, et ensuite déduire des formules générales. Des calculs analogues peuvent être trouvés dans l'ouvrage de référence [Pres92, p. 662].

### 2.1 Exemple numérique

Pour les quatre points donnés dans l'exemple ci-dessus, la fonction  $g$  est

<sup>1</sup>Dans la littérature, on utilise fréquemment la notation  $\chi^2 = g$  et l'on cherche à minimiser le "Chi-carré".

$$\begin{aligned}
g(a, m) &= \sum_{i=1}^5 (a + m \cdot x_i - y_i)^2 \\
&= (a + 0.5)^2 + (a + m - 1)^2 + (a + m \cdot 2 - 2.4)^2 + (a + m \cdot 3.5 - 2.0)^2 + (a + m \cdot 4 - 3.1)^2
\end{aligned}$$

Cette fonction est à minimiser par rapport aux deux paramètres  $a$  et  $m$ .

**Solution :**

Il s'agit de trouver le minimum de la fonction  $g(a, m)$ . Il est clair que des valeurs très grandes (positives) ou très petites (négatives) de  $a$  et  $m$  conduisent à des valeurs très grandes de  $g(a, m)$ . En outre, il est clair aussi que la fonction  $g$  est dérivable par rapport à ses deux variables  $a$  et  $m$ . Un maximum de cette fonction apparaît lorsque les deux dérivées partielles par rapport à  $a$  et à  $m$  sont nulles. Par conséquent, on a les deux conditions nécessaires

$$\frac{\partial g}{\partial a} = 0 \quad \text{et} \quad \frac{\partial g}{\partial m} = 0$$

Pour l'exemple numérique donné ci-dessus, on obtient les deux équations suivantes:

$$\begin{aligned}
\frac{\partial g}{\partial a} &= 2 ((a + 0.5) + (a + m - 1) + (a + m \cdot 2 - 2.4) + (a + m \cdot 3.5 - 2.0) + (a + m \cdot 4 - 3.1)) \\
&= 2 a \cdot 5 + 2 m (0 + 1 + 2 + 3.5 + 4) - 2 (-0.5 + 1.0 + 2.4 + 2.0 + 3.1) \\
&= 2 (a \cdot 5 + m \cdot 10.5 - 8) = 0 \\
\frac{\partial g}{\partial m} &= 2 (1 (a + m - 1) + 2 (a + m \cdot 2 - 2.4) + 3.5 (a + m \cdot 3.5 - 2.0) + 4 (a + m \cdot 4 - 3.1)) \\
&\quad \text{verwende } 0 + 1 + 2 + 3.5 + 4 = 10.5 \\
&\quad \text{und } 0^2 + 1^2 + 2^2 + 3.5^2 + 4^2 = 33.25 \\
&= 2 a \cdot 10.5 + 2 m \cdot 33.25 - 2 (-0.5 \cdot 0 + 1.0 \cdot 1 + 2.4 \cdot 2.0 + 2.0 \cdot 3.5 + 3.1 \cdot 4) \\
&= 2 (a \cdot 10.5 + m \cdot 33.25 - 25.2) = 0
\end{aligned}$$

pour les deux inconnues  $a$  et  $m$ . Il s'agit d'un système d'équations linéaires, qui peut être écrit sous la forme

$$\begin{aligned}
a \cdot 5 + m \cdot 10.5 &= 8 \\
a \cdot 10.5 + m \cdot 33.25 &= 25.2
\end{aligned}$$

Les solutions de ce système sont

$$\begin{pmatrix} a \\ m \end{pmatrix} = \begin{bmatrix} 5 & 10.5 \\ 10.5 & 33.25 \end{bmatrix}^{-1} \begin{pmatrix} 8 \\ 25.2 \end{pmatrix} = \begin{pmatrix} 0.025 \\ 0.750 \end{pmatrix}$$

et la droite cherchée est donnée par l'équation

$$f(x) = a + m x = 0.025 + 0.75 x$$

Les résultats des calculs ci-dessus sont illustrés graphiquement dans la Figure 2. □

Le problème ci-dessus peut aussi être résolu par l'exécution de quelques lignes de code MATLAB/Octave.

```

Octave
x = [0; 1; 2; 3.5; 4]; y = [-0.5; 1; 2.4; 2.0; 3.1];
F = [ones(length(x),1) x];
[p,y_v,r,p_v] = LinearRegression(F,y)

```

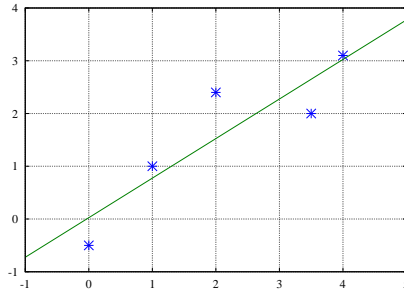


Figure 2: Distribution de points et droite de régression

## 2.2 Le cas général

Dans l'exemple ci-dessus, nous avons construit la fonction

$$g(a, m) = \sum_{i=1}^n (a + m \cdot x_i - y_i)^2$$

que nous avons minimisée par rapport aux paramètres  $a$  et  $m$ .

### Solution :

Pour trouver le minimum de la fonction  $g(a, m) = \chi^2(a, m)$ , il faut trouver les zéros des deux dérivées partielles.

$$\frac{\partial g}{\partial a} = 0 \quad \text{et} \quad \frac{\partial g}{\partial m} = 0$$

Ceci conduit aux deux équations

$$\begin{aligned} \frac{\partial g}{\partial a} &= 2 \sum_{i=1}^n (a + m x_i - y_i) = 0 \\ \frac{\partial g}{\partial m} &= 2 \sum_{i=1}^n (a + m x_i - y_i) x_i = 0 \end{aligned}$$

pour les deux inconnues  $a$  et  $m$ . Il s'agit d'un système d'équations linéaires, qui peut être écrit sous la forme

$$\begin{aligned} a \cdot n + m \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i \\ a \sum_{i=1}^n x_i + m \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n (y_i \cdot x_i) \end{aligned}$$

Les coefficients de ce système d'équations s'expriment par des sommes que l'on peut calculer à partir des données  $x$  et  $y$ . On peut utiliser les notations suivantes:

$$\begin{aligned} S_x &= \sum_{i=1}^n x_i & S_y &= \sum_{i=1}^n y_i \\ S_{xx} &= \sum_{i=1}^n x_i^2 & S_{xy} &= \sum_{i=1}^n x_i y_i \end{aligned}$$

Ces quatre sommes se calculent facilement si les points de données  $(x_i, y_i)$  sont donnés l'un après l'autre. Dans la plupart des calculatrices de poche, cette fonctionnalité est déjà implémentée. Les coefficients  $a$  et  $m$  peuvent alors s'obtenir par résolution de l'équation matricielle

$$\begin{bmatrix} n & S_x \\ S_x & S_{xx} \end{bmatrix} \begin{pmatrix} a \\ m \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix}$$

Cette  $2 \times 2$ -matrice peut être inversée et la matrice inverse est

$$\begin{bmatrix} n & S_x \\ S_x & S_{xx} \end{bmatrix}^{-1} = \frac{1}{n S_{xx} - S_x^2} \begin{bmatrix} S_{xx} & -S_x \\ -S_x & n \end{bmatrix}$$

und

$$\begin{pmatrix} a \\ m \end{pmatrix} = \begin{bmatrix} n & S_x \\ S_x & S_{xx} \end{bmatrix}^{-1} \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix}$$

Finalement, les paramètres  $a$  et  $m$  sont donnés par

$\begin{aligned} \Delta &= n \cdot S_{xx} - S_x^2 \\ a &= \frac{1}{\Delta} (S_{xx} S_y - S_x S_{xy}) \\ m &= \frac{1}{\Delta} (n \cdot S_{xy} - S_x S_y) \end{aligned}$
---

Nous pouvons maintenant entrer les valeurs des paramètres dans l'équation de la droite et obtenir le résultat attendu. En outre, les calculs ci-dessus peuvent être programmés sans grande difficulté ou complexité.

□

### 3 Fondements statistiques

Dans cette section, nous présenterons une introduction aux fondements statistiques qui apparaissent dans les problèmes de régression. Une discussion complète de cette question n'est toutefois pas possible dans ce contexte, pour des raisons de place et de temps.

#### 3.1 Quelques notions de calcul des probabilités

##### 3.1.1 Distribution normale

La fonction

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(\frac{-1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right)$$

décrit une **distribution normale** de **moyenne**  $\mu$  et d'**écart-type**  $\sigma$ . L'expression  $V = \sigma^2$  s'appelle aussi **variance** de la distribution. Si  $x$  est une variable aléatoire suivant la distribution normale ci-dessus, il est alors possible de déterminer la probabilité avec laquelle une mesure de  $x$  tombe dans un certain intervalle  $a \leq x \leq b$  donné. Si l'on dispose d'une grande série de mesures, alors la probabilité est donnée par

$$\begin{aligned} P(a \leq x \leq b) &= \frac{\text{Nombre de mesures observées entre } a \text{ et } b}{\text{Nombre total de mesures}} \\ &= \int_a^b f(x) dx = \frac{1}{\sigma \sqrt{2\pi}} \int_a^b \exp\left(\frac{-1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right) dx \end{aligned}$$

Dans la Figure 3, on peut voir une distribution normale de moyenne  $\mu = 3.5$  et d'écart-type  $\sigma = 1$ . On peut lire la valeur  $\mu = 3.5$  à la position du maximum et la largeur de cette **courbe de Gauss** en forme de cloche est donnée par l'écart-type  $\sigma$  gegeben. En calculant les valeurs

$$f(\mu) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(\frac{-1}{2} \left(\frac{\mu - \mu}{\sigma}\right)^2\right) = \frac{1}{\sigma \sqrt{2\pi}}$$

$$f(\mu \pm \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(\frac{-1}{2} \left(\frac{\mu \pm \sigma - \mu}{\sigma}\right)^2\right) = \frac{1}{\sigma \sqrt{2\pi}} \frac{1}{\sqrt{e}}$$

on peut remarquer que

$$\frac{f(\mu \pm \sigma)}{f(\mu)} = \frac{1}{\sqrt{e}} \approx 0.6065$$

et que la valeur de  $f(x)$  tombe à environ 60% de la valeur maximale lorsqu'on s'éloigne d'un écart-type à partir de la position du maximum.

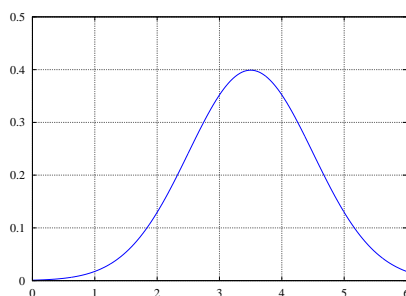


Figure 3: Une distribution normale de moyenne  $\mu = 3.5$  et d'écart-type  $\sigma = 1$

**1 Exemple :** Examinons cette distribution au moyen de quelques exemples de calculs, en prenant comme courbe de référence la courbe de la Figure 3.

- La probabilité qu'une mesure de  $x$  soit plus grande que  $\mu = 3.5$  est donnée par

$$\begin{aligned} P(\mu \leq x) &= \int_{\mu}^{\infty} f(x) dx \\ &= \frac{1}{\sigma \sqrt{2\pi}} \int_{\mu}^{\infty} \exp\left(\frac{-1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right) dx \\ &= \frac{1}{\sigma \sqrt{2\pi}} \int_{3.5}^{\infty} \exp\left(\frac{-1}{2} (x - 3.5)^2\right) dx \\ &= 0.5 \end{aligned}$$

Ainsi, 50% des valeurs mesurées doivent être plus grandes que 3.5.

- La probabilité qu'une mesure de  $x$  soit obtenue entre  $\mu - \sigma$  et  $\mu + \sigma$  est donnée par

$$\begin{aligned} P(\mu - \sigma \leq x \leq \mu + \sigma) &= \int_{\mu - \sigma}^{\mu + \sigma} f(x) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mu - \sigma}^{\mu + \sigma} \exp\left(\frac{-1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right) dx \\ &\approx 0.682689 \end{aligned}$$

Cela signifie qu'environ 68% des mesures s'écartent de  $\mu$  de moins que la distance  $\sigma$ .

- Puisque

$$P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu-2\sigma}^{\mu+2\sigma} \exp\left(\frac{-1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right) dx \approx 0.9545$$

on constate qu'environ 95% des valeurs mesurées s'écartent de la moyenne  $\mu$  de moins que la distance  $2\sigma$ .

- L'intégrale

$$P(-\infty < x < \infty) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{-1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right) dx = 1$$

est toujours égale à 1. Ceci correspond au fait évident que toute mesure doit toujours avoir une valeur comprise entre  $-\infty$  et  $\infty$ . C'est d'ailleurs cette condition qui permet de calculer la constante se trouvant devant l'intégrale.

◇

Pour la distribution normale, on peut effectuer le développement suivant:

$$\begin{aligned} \int_{-\infty}^{\infty} x f(x) dx &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left(\frac{-1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right) dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (s + \mu) \exp\left(\frac{-1}{2} \left(\frac{s}{\sigma}\right)^2\right) ds \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} s \exp\left(\frac{-1}{2} \left(\frac{s}{\sigma}\right)^2\right) ds + \frac{\mu}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{-1}{2} \left(\frac{s}{\sigma}\right)^2\right) ds \\ &= \mu \end{aligned}$$

Lors de la dernière transformation ci-dessus, nous avons utilisé la propriété de symétrie ainsi que le fait que  $P(-\infty < s < \infty) = 1$ . De plus,

$$V = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu)^2 \exp\left(\frac{-1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right) dx$$

Ainsi, la grandeur  $\mu$  peut être interprétée comme la moyenne pondérée de  $x$  et la variance  $V = \sigma^2$  comme moyenne pondérée de l'écart quadratique  $(x - \mu)^2$ . La pondération est donnée par la fonction  $f(x)$  associée à distribution normale.

### 3.1.2 Estimation de la moyenne et de l'écart-type à partir des données

Lors une expérience, on sait que des valeurs mesurées sont distribuées selon la loi normale, mais ni  $\mu$  ni  $\sigma$  ne sont connus. Effectuant une série de mesures, on obtient  $n$  valeurs  $x_1, x_2, \dots, x_n$ . Il s'agit alors de trouver des estimations de la moyenne  $\mu$  et de l'écart-type  $\sigma$  à partir de ces données. Les développements ci-dessus conduisent immédiatement à une formule pour la moyenne:

$$\mu = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{S_x}{n}$$

La variance est estimée par

$$V = \sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Le facteur  $(n - 1)$  (au lieu de  $n$ ) situé devant la somme peut étonner, mais il peut être justifié par des arguments statistiques. La somme peut encore être écrite sous la forme

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - n\bar{x}^2 = S_{xx} - \frac{S_x^2}{n} \\ &= \frac{n S_{xx} - S_x^2}{n} \end{aligned}$$

Lors de la dernière transformation ci-dessus, on utilise le fait que

$$\bar{x}^2 = \left( \frac{1}{n} \sum_{j=1}^n x_j \right)^2 = \frac{S_x^2}{n^2}$$

On obtient donc les formulations suivantes:

$$\begin{aligned} \mu &= \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{S_x}{n} \\ V &= \frac{n S_{xx} - S_x^2}{n(n-1)} \\ \sigma &= \sqrt{\frac{n S_{xx} - S_x^2}{n(n-1)}} \end{aligned}$$

Dans le cas où le nombre  $n$  de mesures est très grand, la formule de somme ci-dessus pour la variance ne devrait pas être utilisée, car elle peut engendrer des erreurs d'arrondis. On utilise plutôt la formule suivante:

$$V = \frac{1}{n-1} \left( \sum_{i=1}^n (x_i - \bar{x})^2 - \frac{1}{n} \left( \sum_{i=1}^n (x_i - \bar{x}) \right)^2 \right)$$

Au cas où  $\bar{x}$  est déterminé exactement, le second terme est nul. Dans le cas contraire, ce terme diminue l'influence des erreurs d'arrondis ([Pres92, p. 613]).

### 3.2 La méthode de Maximum-Likelihood

Etant donné une série de mesures  $(x_i, y_i)$ , nous avons construit dans les sections ci-dessus une droite  $y = mx + a$  qui satisfait la condition intuitive

Trouver  $a$  et  $m$  tels que la somme  $\sum_{i=1}^n (f(x_i) - y_i)^2$  soit minimale.

Cette condition permet de calculer les valeurs des paramètres  $a$  et  $m$  et d'obtenir la droite "la meilleure possible". Nous montrerons ici en ce que signifie l'expression "la meilleure possible". Cette section est adaptée de [Pres92, p. 657].

Il n'y a qu'une seule droite "juste", et toutes les autres sont "fausses" ou "moins bonnes". La question décisive est donc la suivante: *Si les vraies valeurs des paramètres  $a$  et  $m$  sont connues, quelle est la*



probabilité que les valeurs mesurées  $(x_i, y_i)$  apparaissent effectivement ? Les vraies valeurs de  $a$  et  $m$  sont alors caractérisées par le fait que, pour celles-ci, la densité de probabilité est maximale.

Pour la suite des développements, nous accepterons les hypothèses ci-dessous, qui peuvent être justifiées lorsqu'on dispose de mesures concrètes:

1. Les valeurs  $x_i$  sont mesurées de manière exacte.
2. On choisit des valeurs pour  $a$  et  $m$ .
3. Les valeurs mesurées  $y_i$  sont distribuées selon la loi normale autour des valeurs  $a + m x_i$  avec un écart-type indépendant de  $x_i$ .
4. On détermine alors la probabilité (plus précisément: la densité de probabilité), avec laquelle les valeurs mesurées  $y_i$  apparaissent effectivement, pour les valeurs supposées de  $a$  et  $m$ .
5. Pour une série de mesures données, on cherche finalement les valeurs de  $a$  et  $m$  pour lesquelles la densité de probabilité est **maximale**.

Au lieu de calculer la probabilité avec laquelle les valeurs  $a$  et  $m$  sont "justes", on calcule la probabilité d'apparition des mesures déjà effectuées, qui doit être maximale. Dans cette situation, on parle de la **méthode du Maximum-Likelihood**. La résolution de cette tâche revient, au bout du compte, à résoudre le problème de minimum exposé ci-dessus.

### 3.3 Estimation des variances de $a$ et $m$

Une nouvelle série de mesures conduirait à de nouvelles estimations de  $a$  et de  $m$ . Il s'agit alors de pouvoir estimer, d'une certaine manière, quelle est la fiabilité des valeurs de  $a$  et de  $m$ . Pour cela, il n'est **pas adéquat** de considérer plusieurs séries de mesures, mais il est préférable de travailler avec une série de mesures unique (voir à ce sujet la Figure 4). Seule la première série de mesures sera prise en compte, les autres séries n'étant qu'hypothétiques. On montrera par la suite que les valeurs de  $a$  et  $m$  sont aussi distribuées selon une loi normale. De ce fait, la variance (ou le carré de l'écart-type) est une mesure significative pour les fluctuations attendues.

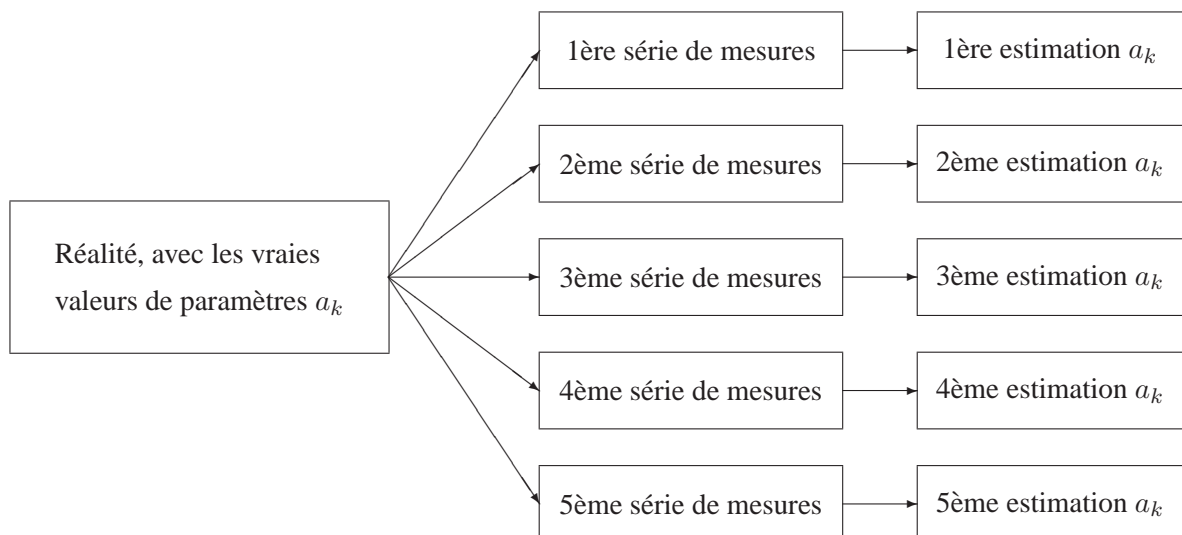


Figure 4: Différentes séries de mesures à des estimations différentes des paramètres

### 3.3.1 Détermination de l'écart-type $\sigma$ des mesures $y_i$

Parfois, l'écart-type  $\sigma$  des mesures peut être estimé à d'informations sur l'ordonnement des mesures, sans que les données soient analysées. Lorsque cela n'est pas possible, on peut essayer d'obtenir une estimation à l'aides des mesures  $y_i$ .

On part du principe que les deux paramètres  $a$  et  $m$  de la droite sont déjà connus. Dans ce cas, les valeurs mesurées  $y_i$  doivent être distribuées selon la loi normale autour des valeurs "justes"  $a + m x_i$ , c'est-à-dire: les différences  $y_i - a - m x_i$  doivent être distribuées selon la loi normale avec une moyenne nulle. Ceci conduit à l'estimation

$$\sigma^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - a - m x_i)^2$$

Le dénominateur est ici  $(n-2)$ , car deux paramètres ont déjà été déterminés.

### 3.3.2 Variance d'une somme de variables aléatoires, variances de $a$ et de $m$

Soient des grandeurs aléatoires  $y_i$  **indépendantes** distribuées selon la loi normale, de variances respectives  $V(y_i)$ . Toute combinaison linéaire des grandeurs  $y_i$  est aussi une grandeur aléatoire distribuée selon la loi normale, et l'on a pour les variances les formules

$$\begin{aligned} V(y_1 + y_2) &= V(y_1) + V(y_2) \\ V(\alpha_1 y_1) &= \alpha_1^2 V(y_1) \\ V\left(\sum \alpha_i y_i\right) &= \sum \alpha_i^2 V(y_i) \end{aligned}$$

Ces règles fondamentales pour le calcul des variances peuvent être vérifiées à l'aide d'intégrales.

Les valeurs des  $x_i$  étant connues, on peut aussi vérifier les résultats suivants:

$$S_y = \sum_{i=1}^n y_i \implies V(S_y) = \sum_{i=1}^n V(y_i) \quad \text{et} \quad V(S_x S_y) = S_x^2 \sum_{i=1}^n V(y_i)$$

Examinons maintenant les sommes qui apparaissent dans les formules pour  $a$  et  $m$ :

$$\begin{aligned} S_{xx} S_y - S_x S_{xy} &= \sum_{i=1}^n (S_{xx} - S_x x_i) y_i \implies V(S_{xx} S_y - S_x S_{xy}) = \sum_{i=1}^n (S_{xx} - S_x x_i)^2 V(y_i) \\ n S_{xy} - S_x S_y &= \sum_{i=1}^n (n x_i - S_x) y_i \implies V(n S_{xy} - S_x S_y) = \sum_{i=1}^n (n x_i - S_x)^2 V(y_i) \end{aligned}$$

Nous supposons ici que les écarts-types de toutes les valeurs  $y_i$  sont identiques et données par  $\sigma$ , c'est-à-dire:  $V(y_i) = \sigma^2$ . On a alors:

$$\begin{aligned} V(S_{xx} S_y - S_x S_{xy}) &= \sum_{i=1}^n (S_{xx}^2 - 2 S_{xx} S_x x_i + S_x^2 x_i^2) \sigma^2 \\ &= (n S_{xx}^2 - 2 S_{xx} S_x^2 + S_x^2 S_{xx}) \sigma^2 = (n S_{xx}^2 - S_{xx} S_x^2) \sigma^2 \\ V(n S_{xy} - S_x S_y) &= \sum_{i=1}^n (n^2 x_i^2 - 2 n x_i S_x + S_x^2) \sigma^2 \\ &= (n S_{xx} - 2 n S_x S_x + n S_x^2) \sigma^2 = (n^2 S_{xx} - n S_x^2) \sigma^2 \end{aligned}$$

Il suit des formules fondamentales

$$a = \frac{1}{\Delta} (S_{xx} S_y - S_x S_{xy}) \quad , \quad m = \frac{1}{\Delta} (n \cdot S_{xy} - S_x S_y) \quad \text{où} \quad \Delta = n \cdot S_{xx} - S_x^2$$

les résultats immédiats

$$\sigma_a^2 = V(a) = \frac{1}{\Delta^2} (n S_{xx}^2 - S_{xx} S_x^2) \sigma^2 = \frac{S_{xx}}{n \cdot S_{xx} - S_x^2} \sigma^2$$

$$\sigma_m^2 = V(m) = \frac{1}{\Delta^2} (n^2 S_{xx} - n S_x^2) \sigma^2 = \frac{n}{n \cdot S_{xx} - S_x^2} \sigma^2$$

Grâce à cela, nous pouvons estimer les écarts-types des paramètres  $a$  et  $m$  de la droite cherchée, à partir des valeurs mesurées  $(x_i, y_i)$  et de l'écart-type  $\sigma$  des valeurs  $y$  de la droite.

Des parties des développements ci-dessus sont repris de [Bevi69, p. 113]. D'autres explications analogues peuvent être trouvées dans [Pres92].

## 4 Compléments

### 4.1 Régression linéaire générale, notation matricielle

Etant donné une série de mesures  $(x_i, y_i)$ , nous avons construits ci-dessus le système d'équations

$$\begin{bmatrix} n & S_x \\ S_x & S_{xx} \end{bmatrix} \begin{pmatrix} a \\ m \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix}$$

Nous allons maintenant transformer ces équations à l'aide d'une matrice  $\mathbf{F}$  et d'un vecteur  $\vec{y}$  contenant les données:

$$\mathbf{F} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \text{et} \quad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}$$

On a:

$$\begin{bmatrix} n & S_x \\ S_x & S_{xx} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \mathbf{F}^T \cdot \mathbf{F}$$

et

$$\begin{pmatrix} S_y \\ S_{xy} \end{pmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & \cdots & x_n \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \mathbf{F}^T \cdot \vec{y}$$

Le système ci-dessus peut aussi être écrit sous la forme

$$\mathbf{F}^T \cdot \mathbf{F} \cdot \vec{p} = \mathbf{F}^T \cdot \vec{y}$$

Dans le cas d'une droite de régression, ces changements de notations n'apportent aucun avantage. En revanche, cette nouvelle notation permet de décrire des problèmes plus compliqués.

Les idées développées ci-dessus peuvent être utilisées non seulement pour la recherche d'une droite dans une distribution de points, mais elles permettent aussi la prise en compte de plusieurs variables indépendantes et de courbes autres que des droites (ou des plans lorsqu'il y a deux variables). Les difficultés principales se trouvent alors dans la notation utilisée et dans les calculs plus longs, mais ceux-ci peuvent être effectués sans problèmes par un ordinateur.

Considérons donc plusieurs valeurs des variables indépendantes  $x_1, x_2, \dots, x_n$  et des fonctions de la forme

$$\Phi(x) = \sum_{j=1}^k p_j \cdot f_j(x)$$

où les fonctions  $f_j(x)$  sont données. On cherche à calculer les paramètres  $p_1, \dots, p_k$ . Pour les points de données  $(x_i, y_i)$ , l'expression

$$\chi^2 := \sum_{i=1}^n (\Phi(x_i) - y_i)^2 = \sum_{i=1}^n \left( \left( \sum_{j=1}^k p_j \cdot f_j(x_i) \right) - y_i \right)^2$$

doit être minimisée.

La condition que toutes les dérivées partielles par rapport aux  $p_j$  soient nulles conduit à la construction d'un système linéaire de  $k$  équations, pour les  $k$  inconnues  $p_j$ . On va examiner les détails en section 4.3.

**2 Exemple :** Prenant les valeurs particulières

$$k = 2 \quad p_1 = a \quad p_2 = m \quad f_1(x) = 1 \quad f_2(x) = x \quad \Phi(x) = a + mx$$

on obtient la situation de la section précédente: une droite qui doit être adaptée à des valeurs mesurées.  $\diamond$

**3 Exemple :** Prenant les paramètres

$$\begin{array}{lll} p_1 = a & p_2 = b & p_3 = c \\ f_1(x) = x^2 & f_2(x) = x & f_3(x) = 1 \end{array}$$

on recherche une parabole d'équation  $\Phi(x) = ax^2 + bx + c$  qui s'adapte le mieux possible aux valeurs mesurées.

Dans la section 4.3, nous verrons que dans ce cas la matrice  $\mathbf{F}$  est donnée par

$$\mathbf{F} = \begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \\ \vdots & \vdots & \vdots \\ x_n^2 & x_n & 1 \end{bmatrix}$$

et la matrice  $\mathbf{M} = \mathbf{F}^T \cdot \mathbf{F}$  dans le système  $\mathbf{F}^T \cdot \mathbf{F} \cdot \vec{p} = \mathbf{F}^T \cdot \vec{y}$  est une matrice  $3 \times 3$ .  $\diamond$

Il est important de souligner que la fonction  $f_i(x)$  peut tout à fait être non linéaire. La linéarité fait référence aux paramètres  $p_j$ . On trouve dans la table 1 quelques exemples de telles fonctions.

Type de fonctions	Paramètres	Régression linéaire / non linéaire
$y = a + m x$	$a, m$	linéaire
$y = a x^2 + b x + c$	$a, b, c$	linéaire
$y = a e^{c x}$	$a, c$	non linéaire
$y = a e^{c x}$	$a$	linéaire
$y = a \sin(\omega t + \delta)$	$a, \omega, \delta$	non linéaire
$y = a \cos(\omega t) + b \sin(\omega t)$	$a, b$	linéaire

Tableau 1: Exemples de régressions linéaires ou non linéaires

## 4.2 Example : Intensity of light of an LED depending on the angle of observation

### 4.2.1 Description of the problem

The intensity of the light emitted by an LED will depend on the angle  $\alpha$  of observation. The data sheets of the supplier should show this information. A sample of real data is stored in the file `LEDdata.txt` and the results is shown in Figure 5.

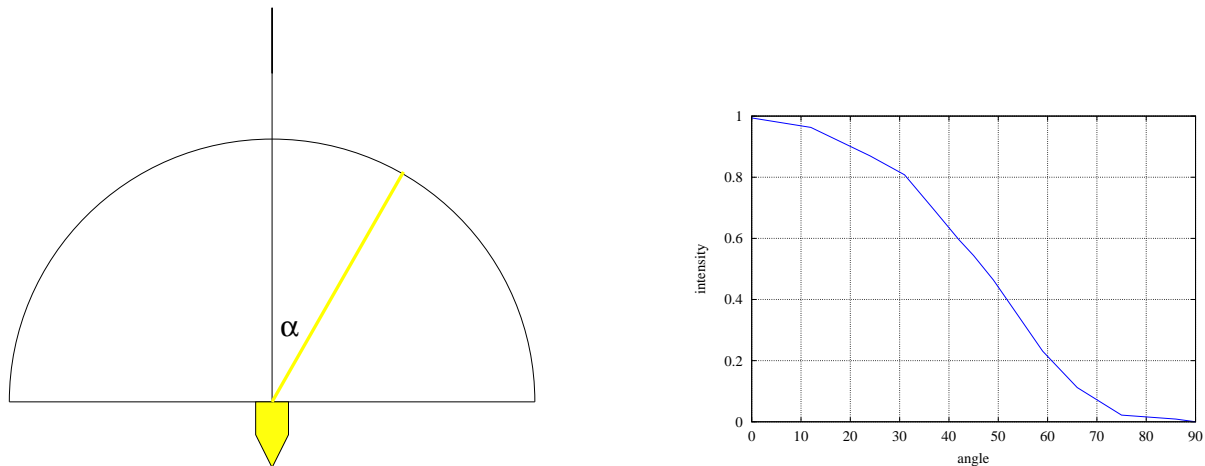


Figure 5: Intensity of light as function of the angle

### 4.2.2 A first, naive solution and its problems

To do further analysis it can be useful to have a formula for the intensity as function of the angle and linear regression is one of the options on how to obtain such a formula. We try to fit a polynomial of degree 4 through the given points, i.e. we seek parameters  $p_j$  such that

$$\Phi(x) = p_1 1 + p_2 x + p_3 x^2 + p_4 x^3 + p_5 x^4$$

In this case the matrix  $\mathbf{F}$  is given by

$$\mathbf{F} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 \\ 1 & x_3 & x_3^2 & x_3^3 & x_3^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & x_n^4 \end{bmatrix}$$

The following code will fit a polynomial of degree 4 through those points and then display the result in Figure 6. The algorithm is performing the following steps:

- Read the data.
- Construct the matrix  $\mathbf{F}$  with the correct columns
- Solve the linear system  $\mathbf{F}^T \cdot \mathbf{F} \cdot \vec{p} = \mathbf{F}^T \cdot \vec{y}$  of 5 linear equations for the 5 unknown components of the parameter vector  $\vec{p}$

```

                                Octave
angle      =[-0 12 24 31 37 42 45 49 54 59 66 75 86 90]';
intensity = [ 0.994 0.963 0.87 0.808 0.693 0.596...
              0.544 0.464 0.348 0.231 0.112 0.022 0.009 0]';

F = ones(length(angle),5);
for k = 1:5
    F(:,k) = angle.^(k-1);
end

M = F'*F;
p = M\(F'*intensity)

```

The resulting parameters point towards an intensity function

$$\text{Intensity}(\alpha) = 124.28 + 0.1111\alpha - 4.0576 \cdot 10^{-3}\alpha^2 + 5.0299 \cdot 10^{-5}\alpha^3 - 2.1087 \cdot 10^{-7}\alpha^4$$

The code below generates Figure 6.

```

                                Octave
al = (0:1:90)'; % consider angles from 0 to 90 degree

Inew = polyval(flipud(p),al);
plot(angle,intensity,al,Inew);
grid on
xlabel('angle');ylabel('intensity')

```

The result in Figure 6 is useless since the estimated variances of the parameters are of the same order of magnitude as the values<sup>2</sup>. We need to find the reason for the failure and how to eliminate the problems.

The previous implementation of the linear regression algorithm has to solve a system of equations with the matrix  $\mathbf{F}' \cdot \mathbf{F}$ .

With the help of

```

                                Octave

```

<sup>2</sup>Previous releases of MATLAB and Octave generated numbers which were obviously wrong. This was very useful for didactical purposes, but less so for real world problems.

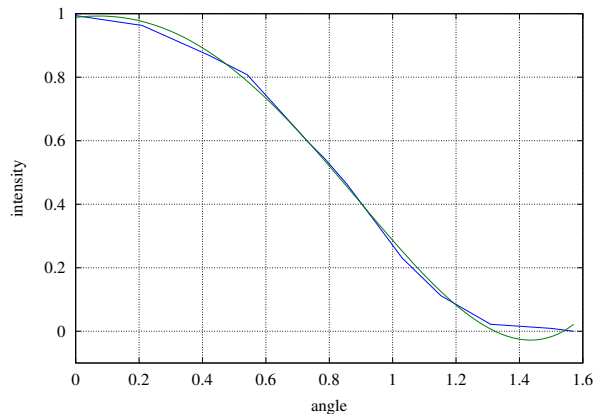


Figure 6: Intensity of light as function of the angle and a first regression curve

```
F' * F
ans =
 1.4000e+01  6.7000e+02  4.1114e+04  2.8161e+06  2.0767e+08
 6.7000e+02  4.1114e+04  2.8161e+06  2.0767e+08  1.6113e+10
 4.1114e+04  2.8161e+06  2.0767e+08  1.6113e+10  1.2949e+12
 2.8161e+06  2.0767e+08  1.6113e+10  1.2949e+12  1.0666e+14
 2.0767e+08  1.6113e+10  1.2949e+12  1.0666e+14  8.9414e+15
```

we see that the matrix contains numbers of the order 1 and of the order  $10^{16}$  and one should not be surprised by trouble when solving such a system of equations. Mathematically speaking we have a **condition number** that is very large ( $10^{16}$ ) and thus we will lose many digits of precision. Entries of vastly different sizes are an indication for large condition numbers, but other effects also matter and you will have to consult specialized literature or a mathematician to obtain more information.

### 4.2.3 How to eliminate the problem

There are different measures to be taken to avoid the problem. For a reliable solution they should all be taken into account.

#### 1. Rescaling

For a polynomial of degree 4 and angles of  $90^\circ$  the matrix  $\mathbf{F}$  will contain numbers of the size 1 and  $90^4$ . Thus  $\mathbf{F}' \cdot \mathbf{F}$  will contain numbers of the size  $90^8 \approx 100^8 = 10^{16}$ . If we switch to radians instead of degrees this will be reduced to  $(\frac{\pi}{2})^8 \approx 40$  and thus should avoid the problem. The code below will generate a reasonable solution.

```
Octave
load 'LEDdata.txt'
scalefactor = 180/pi;
newangle = angle/scalefactor;

F = ones(length(newangle),5);
for k = 1:5
    F(:,k) = newangle.^(k-1);
end

M = F' * F;
p = M \ (F' * intensity)

al = ((0:1:90)')/scalefactor; % consider angles from 0 to 90 degree
Inew = polyval(flipud(p),al);
```

```

plot(newangle*scalefactor,intensity,al*scalefactor,Inew);
grid on
xlabel('angle');ylabel('intensity')

```

## 2. Better choice of basis functions

Since the intensity function  $I(\alpha)$  has to be symmetric with respect to  $\alpha$ , i.e.  $I(-\alpha) = I(\alpha)$ , there can be no contributions of the form  $\alpha$  or  $\alpha^3$ . Thus we seek a good fit for a function of the type

$$I(\alpha) = p_1 + p_2 \alpha^2 + p_3 \alpha^4$$

The code below leads to the result in Figure 7. The condition number of  $\mathbf{F}' \cdot \mathbf{F}$  is approximately 200 and thus poses no problem. The result in Figure 7 is now useful for further investigations and the computations indicate that the intensity is approximated by

$$I(\alpha) = 1.02951 - 0.95635 \alpha^2 + 0.21890 \alpha^4$$

The code is a slight modification of the previous code.

```

Octave
angle      = [ -0 12 24 31 37 42 45 49 54 59 66 75 86 90]';
intensity  = [ 0.994 0.963 0.87 0.808 0.693 0.596...
              0.544 0.464 0.348 0.231 0.112 0.022 0.009 0]';
scalefactor = 180/pi;
newangle   = angle/scalefactor;

F = ones(length(angle),3);
F(:,1) = ones(length(angle),1);
F(:,2) = newangle.^2;
F(:,3) = newangle.^4;

M = F'*F;
p = M\(F'*intensity)

al = ((0:1:90)')/scalefactor; % consider angles from 0 to 90 degree
Inew = polyval([p(3) 0 p(2) 0 p(1)],al);
plot(newangle*scalefactor,intensity,al*scalefactor,Inew);
grid on
xlabel('angle');ylabel('intensity')

```

This point is by far the most important aspect to consider when using the linear regression method.

Choose your basis function for linear regression very carefully,  
based on information about the system to be examined.

There are many software packages (*Mathematica*, *MATLAB*, *Octave*, *Excel*, ...) that allow to perform linear regression with polynomials of high degrees. This author is not aware of **one single problem** where a polynomial of high degree lead to useful information. All software behave according to the **GIGO**<sup>3</sup> principle.

## 3. QR decomposition instead of $\mathbf{F}' \cdot \mathbf{F}$ matrix

Idea and consequences of this change in algorithm is given in the next section. Any serious implementation of a linear regression method should use this modification. The provided *Octave*-code `LinearRegression()` does take this point into account.

<sup>3</sup>Garbage In, Garbage Out



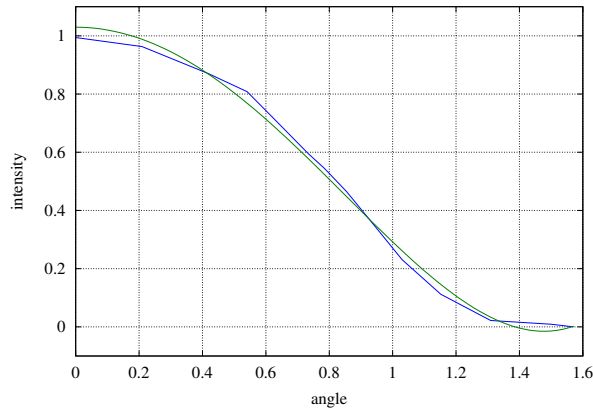


Figure 7: Intensity of light as function of the angle and regression with an even function

#### 4.2.4 Using the command `LinearRegression()`

At the usual place you find the *Octave* script `LinearRegression.m`. Within *Octave* you may type `help LinearRegression` to access the builtin help information. The command will perform a general linear regression and also determine the estimated standard deviation for the parameters, see section 3.3.2 (page 10).

```

Octave
angle      = [ -0 12 24 31 37 42 45 49 54 59 66 75 86 90]';
intensity  = [ 0.994 0.963 0.87 0.808 0.693 0.596...
              0.544 0.464 0.348 0.231 0.112 0.022 0.009 0]';
scalefactor = 180/pi;
newangle   = angle/scalefactor;

F = ones(length(angle),3);
F(:,1) = ones(length(angle),1);
F(:,2) = newangle.^2;
F(:,3) = newangle.^4;

[p,y_var,r,p_var] = LinearRegression(F,intensity);
p
p_stddev = sqrt(p_var)

Inew = F*p;
plot(angle,intensity,angle,Inew)

```

Dans les notations ci-dessus, et grâce à l'utilisation de la commande `LinearRegression()` dans *Octave*, une grande variété de fonctions peuvent être examinées, dans l'idée de les adapter aux mesures. Il faut toutefois souligner ici que **le choix du type de fonction à utiliser ne doit être effectué qu'en relation avec l'application voulue.**

Si l'on sait que la courbe cherchée est une parabole (dans l'exemple), il est alors possible de déterminer les trois paramètres de cette parabole au moyen d'une régression linéaire. Il est toutefois très inhabituel d'effectuer une régression linéaire pour déterminer des polynômes de degré élevé.

### 4.3 Vérification des équations

Pour des points donnés  $(x_i, y_i)$  avec  $1 \leq i \leq n$  et un choix adapté des fonctions de base  $f_j(x)$  avec  $1 \leq j \leq k$ , on cherche des valeurs optimales pour les paramètres  $p_j$ , tels que la fonction

$$\Phi(x) = \sum_{j=1}^k p_j f_j(x)$$

s'adapte le mieux possible aux points donnés. Construire le système d'équations pour les inconnues  $p_j$ , en reprenant l'idée de base dans la section 2.

**Solution :** Les calculs sont un peu plus longs, et la solution sera décomposée en plusieurs étapes intermédiaires.

#### 4.3.1 Construction des équations

La fonction

$$\chi^2(\vec{p}) := \sum_{i=1}^n (\Phi(x_i) - y_i)^2 = \sum_{i=1}^n \left( \left( \sum_{j=1}^k p_j \cdot f_j(x_i) \right) - y_i \right)^2$$

est à minimiser relativement aux variables  $p_l$ , pour  $l = 1, \dots, k$ . Ses dérivées partielles sont

$$\frac{\partial \chi^2(x_i)}{\partial p_l} = \frac{\partial}{\partial p_l} \sum_{j=1}^k p_j \cdot f_j(x_i) = f_l(x_i) \quad \text{pour } l = 1, 2, \dots, k$$

Il s'agit donc de résoudre les équations

$$\frac{\partial \chi^2(\vec{p})}{\partial p_l} = 2 \sum_{i=1}^n \left( \sum_{j=1}^k p_j \cdot f_j(x_i) - y_i \right) \cdot f_l(x_i) = 0 \quad \text{pour } l = 1, \dots, k$$

Ces équations peuvent aussi s'écrire sous la forme

$$\sum_{j=1}^k p_j \left( \sum_{i=1}^n f_j(x_i) \cdot f_l(x_i) \right) = \sum_{i=1}^n f_l(x_i) \cdot y_i \quad \text{pour } l = 1, \dots, k$$

Il est tout à fait possible de résoudre ces équations directement. Toutefois, le danger est grand que le code devienne incompréhensible du fait des nombreuses sommes et des nombreux indices qui y apparaissent. Pour simplifier cela, le traitement de ce système d'équations sous forme matricielle est beaucoup plus clair.

#### 4.3.2 Notation vectorielle et matricielle

Pour comprendre ce qui suit, il est nécessaire d'être familiarisé avec les notations des vecteurs, des matrices, et de leurs multiplications. Le but est de formuler les expressions de la section précédente sous forme matricielle. Soient donc les vecteurs et la matrice

$$\vec{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{F} = \begin{bmatrix} f_1(x_1) & f_2(x_1) & \dots & f_k(x_1) \\ f_1(x_2) & f_2(x_2) & \dots & f_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_n) & f_2(x_n) & \dots & f_k(x_n) \end{bmatrix}$$

Il est facile de voir que

$$(\mathbf{F} \cdot \vec{p})_i = \sum_{j=1}^k f_j(x_i) \cdot y_j \quad \text{pour } i = 1, 2, \dots, n$$

Les calculs ci-dessus peuvent être écrits au moyen de cette "nouvelle" notation, et l'on obtient pour  $\chi^2(\vec{p})$  l'expression

$$\begin{aligned} \chi^2(\vec{p}) &= |\mathbf{F} \cdot \vec{p} - \vec{y}|^2 = (\mathbf{F} \cdot \vec{p} - \vec{y})^T \cdot (\mathbf{F} \cdot \vec{p} - \vec{y}) \\ &= (\vec{p}^T \cdot \mathbf{F}^T - \vec{y}^T) \cdot (\mathbf{F} \cdot \vec{p} - \vec{y}) \\ &= \vec{p}^T \mathbf{F}^T \mathbf{F} \vec{p} - \vec{y}^T \mathbf{F} \vec{p} - \vec{p}^T \mathbf{F}^T \vec{y} + \vec{y}^T \vec{y} \end{aligned}$$

Grâce au fait que  $\vec{a}^T \vec{b} = \langle \vec{a}, \vec{b} \rangle$  (c'est-à-dire: le produit d'un vecteur-ligne avec un vecteur-colonne revient au même que le produit scalaire des deux vecteurs), nous avons:

$$\chi^2(\vec{p}) = \langle \mathbf{F} \cdot \vec{p} - \vec{y}, \mathbf{F} \cdot \vec{p} - \vec{y} \rangle = \langle \mathbf{F} \vec{p}, \mathbf{F} \vec{p} \rangle - 2 \langle \mathbf{F} \vec{p}, \vec{y} \rangle + \langle \vec{y}, \vec{y} \rangle$$

Cela donne dans les notations ci-dessus:

$$\begin{aligned} (\mathbf{F}^T \cdot \vec{y})_l &= \sum_{i=1}^n f_l(\vec{x}_i) \cdot y_i \\ (\mathbf{F}^T \cdot \mathbf{F})_{l,j} &= \sum_{i=1}^n f_l(\vec{x}_i) \cdot f_j(\vec{x}_i) \\ (\mathbf{F}^T \cdot \mathbf{F} \cdot \vec{p})_l &= \sum_{j=1}^m (\mathbf{F}^T \cdot \mathbf{F})_{l,j} p_j \end{aligned}$$

De cette manière, le système d'équation

$$\sum_{j=1}^m p_j \left( \sum_{i=1}^n f_j(\vec{x}_i) \cdot f_l(\vec{x}_i) \right) = \sum_{i=1}^n f_l(\vec{x}_i) \cdot y_i \quad l = 1, \dots, m$$

peut aussi s'écrire sous la forme raccourcie

$$\mathbf{F}^T \cdot \mathbf{F} \cdot \vec{p} = \mathbf{F}^T \cdot \vec{y} \quad .$$

Dans cette dernière expression, le vecteur  $\vec{y}$  et la matrice  $\mathbf{F}$  sont construits à partir des données  $(x_i, y_i)$  et des fonctions  $f_j$ , et il s'agit de calculer le vecteur  $\vec{p}$ .

### 4.3.3 Codage efficace, décomposition QR et Octave

Pour une matrice qui contient plus de lignes que de colonnes, la décomposition QR est définie par le produit

$$\mathbf{F} = \mathbf{Q} \cdot \mathbf{R}$$

où  $\mathbf{Q}$  est une matrice orthogonale ( $\mathbf{Q}^{-1} = \mathbf{Q}^T$ ) et où  $\mathbf{R}$  est une matrice triangulaire supérieure:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_o \\ \mathbf{0} \end{bmatrix} \quad \text{et} \quad \mathbf{Q} = \begin{bmatrix} \mathbf{Q}_l & \mathbf{Q}_r \end{bmatrix}$$

Dans ces matrices, le bloc supérieur  $\mathbf{R}_o$  et le bloc gauche  $\mathbf{Q}_l$  sont des matrices inversibles. La multiplication d'un vecteur avec la matrice orthogonale  $\mathbf{Q}$  correspond à une rotation de ce vecteur, dont la

longueur ne change pas. Il s'agit donc de résoudre le problème:

$$\begin{aligned}
 \mathbf{F} \cdot \vec{p} - \vec{y} &= \vec{r} \text{ de longueur minimale} \\
 \mathbf{Q} \cdot \mathbf{R} \cdot \vec{p} - \vec{y} &= \vec{r} \text{ de longueur minimale} \\
 \mathbf{R} \cdot \vec{p} - \mathbf{Q}^T \cdot \vec{y} &= \mathbf{Q}^T \cdot \vec{r} \\
 \begin{bmatrix} \mathbf{R}_o \\ \mathbf{0} \end{bmatrix} \cdot \vec{p} - \begin{bmatrix} \mathbf{Q}_l^T \\ \mathbf{Q}_r^T \end{bmatrix} \cdot \vec{y} &= \begin{bmatrix} \mathbf{Q}_l^T \\ \mathbf{Q}_r^T \end{bmatrix} \cdot \vec{r} \\
 \begin{bmatrix} \mathbf{R}_o \cdot \vec{p} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{Q}_l^T \cdot \vec{y} \\ \mathbf{Q}_r^T \cdot \vec{y} \end{bmatrix} &= \begin{bmatrix} \mathbf{Q}_l^T \cdot \vec{r} \\ \mathbf{Q}_r^T \cdot \vec{r} \end{bmatrix}
 \end{aligned}$$

Le vecteur  $\vec{p}$  n'a pas d'influence sur la partie inférieure gauche du système. La longueur du vecteur du côté gauche est donc minimale si  $\mathbf{Q}_l^T \cdot \vec{r} = \vec{0}$ . Grâce à cela, on peut utiliser les premières équations du système pour déterminer  $\vec{p}$  de manière unique par l'équation simplifiée:

$$\begin{aligned}
 \mathbf{R}_o \cdot \vec{p} &= \mathbf{Q}_l^T \cdot \vec{y} \\
 \vec{p} &= \mathbf{R}_o^{-1} \cdot \mathbf{Q}_l^T \cdot \vec{y}
 \end{aligned}$$

La décomposition QR est avantageuse pour la résolution du système  $\mathbf{F}^T \cdot \mathbf{F} \cdot \vec{p} = \mathbf{F}^T \cdot \vec{y}$ , car elle permet de diminuer la condition de la matrice, et ne pose que peu de problèmes numériques.

Pour l'exemple introductif de la première page, on a

$$\vec{x} = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3.5 \\ 4 \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} -0.5 \\ 1 \\ 2.4 \\ 2.0 \\ 3.1 \end{pmatrix} \quad \text{pour} \quad \mathbf{F} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3.5 \\ 1 & 4 \end{bmatrix}$$

La décomposition QR de la matrice  $\mathbf{F}$  est donnée par

$$\mathbf{F} = \mathbf{Q} \cdot \mathbf{R} = \begin{bmatrix} -0.447 & -0.627 & -0.421 & -0.350 & -0.326 \\ -0.447 & -0.329 & 0.034 & 0.504 & 0.661 \\ -0.447 & -0.030 & 0.846 & -0.197 & -0.211 \\ -0.447 & 0.418 & -0.219 & 0.562 & -0.512 \\ -0.447 & 0.568 & -0.240 & -0.519 & 0.388 \end{bmatrix} \cdot \begin{bmatrix} -2.236 & -4.696 \\ 0 & 3.347 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

La résolution effective du problème est donc

$$\begin{aligned}
 \begin{bmatrix} -2.236 & -4.696 \\ 0 & 3.347 \end{bmatrix} \cdot \begin{pmatrix} a \\ m \end{pmatrix} &= \begin{bmatrix} -0.447 & -0.447 & -0.447 & -0.447 & -0.447 \\ -0.627 & -0.329 & -0.030 & 0.418 & 0.568 \end{bmatrix} \begin{pmatrix} -0.5 \\ 1 \\ 2.4 \\ 2.0 \\ 3.1 \end{pmatrix} \\
 &= \begin{pmatrix} -3.578 \\ 2.510 \end{pmatrix}
 \end{aligned}$$

Ce système est facilement résoluble de bas en haut, et sa solution est  $(a, m) = (0.025, 0.750)$ . Le code Octave/MATLAB-Code ci-dessous exécute ces calculs.

### Octave

```
x = [0;1;2;3.5;4]; y = [-0.5;1;2.4;2.0;3.1];

F = ones(length(x),2); % construct the matrix F
F(:,2) = x; % values of x coordinates as second column

[Q,R] = qr(F) % factorization
p = R(1:2,1:2)\(Q(:,1:2)'*y) % solve the upper block of R*p=Q'*y
```

Au cas où l'on n'a pas besoin des matrices complètes **R** et **Q**, les deux dernières lignes du code peuvent être remplacées par le code ci-dessous:

### Octave

```
[Q,R] = qr(F,0);
p = R\ (Q'*y)
```

□

## 4.4 Mesures de précisions différentes

### 4.4.1 Droite de régression

On dispose d'une série de mesures  $(x_i, y_i)$  dont les valeurs  $x_i$  sont connues de manière exacte, et dont l'écart-type des valeurs  $y_i$  est  $\sigma_i$ . Dans les sections précédentes, nous avons supposé que les écarts-types  $\sigma_i$  étaient tous les mêmes, mais ceci n'est pas toujours le cas. Un écart-type  $\sigma_i$  petit signifie que la valeur correspondante  $y_i$  est connue avec une grande précision, et l'on devrait attribuer à une telle valeur un poids plus grand qu'à une valeur dont l'écart-type est plus important, donc la précision plus faible. Ceci peut être réalisé par la fonction suivante, à minimiser:

$$\chi^2 := \sum_{i=1}^n \left( \frac{a + m x_i - y_i}{\sigma_i} \right)^2$$

Pour que la valeur de cette fonction soit minimale, les dérivées partielles par rapport à  $a$  et à  $m$  doivent s'annuler, ce qui conduit au système d'équations linéaires:

$$\begin{bmatrix} S & S_x \\ S_x & S_{xx} \end{bmatrix} \begin{pmatrix} a \\ m \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix}$$

où

$$\begin{aligned} S &= \sum_{i=1}^n \frac{1}{\sigma_i^2} & S_x &= \sum_{i=1}^n \frac{x_i}{\sigma_i^2} & S_y &= \sum_{i=1}^n \frac{y_i}{\sigma_i^2} \\ S_{xx} &= \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} & S_{xy} &= \sum_{i=1}^n \frac{x_i y_i}{\sigma_i^2} \end{aligned}$$

La solution unique de ce système est

$$\begin{aligned} \Delta &= S \cdot S_{xx} - S_x^2 \\ a &= \frac{1}{\Delta} (S_{xx} S_y - S_x S_{xy}) \\ m &= \frac{1}{\Delta} (S \cdot S_{xy} - S_x S_y) \end{aligned}$$

En plus de ces explications, on peut aussi mettre en œuvre la méthode du **Maximum-Likelihood**, de manière à justifier les résultats.

Les formules ci-dessus sont reprises de [Pres92, §15.3, p. 661].

#### 4.4.2 Régression générale

La formule

$$\chi^2 = \langle \mathbf{F} \cdot \vec{p} - \vec{y}, \mathbf{F} \cdot \vec{p} - \vec{y} \rangle = \langle \mathbf{F}\vec{p}, \mathbf{F}\vec{p} \rangle - 2 \langle \mathbf{F}\vec{p}, \vec{y} \rangle + \langle \vec{y}, \vec{y} \rangle$$

de la section 4.3.2 pour la variance avec pondération égale peut être modifiée, au cas où l'on attribue à chaque point de mesure la pondération  $1/\sigma_i^2$ . On peut alors utiliser la matrice de pondération  $\mathbf{W}$  suivante:

$$\mathbf{W} = \begin{bmatrix} 1/\sigma_1 & & & & \\ & 1/\sigma_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & 1/\sigma_m \end{bmatrix}$$

On a alors

$$\chi^2 = \langle \mathbf{W}\mathbf{F} \cdot \vec{p} - \vec{y}, \mathbf{W}\mathbf{F} \cdot \vec{p} - \vec{y} \rangle = \langle \mathbf{W}\mathbf{F}\vec{p}, \mathbf{W}\mathbf{F}\vec{p} \rangle - 2 \langle \mathbf{W}\mathbf{F}\vec{p}, \vec{y} \rangle + \langle \vec{y}, \vec{y} \rangle$$

et le système matriciel correspondant est

$$(\mathbf{W}\mathbf{F})^T \cdot \mathbf{W}\mathbf{F} \cdot \vec{p} = (\mathbf{W}\mathbf{F})^T \cdot \vec{y}$$

ou encore

$$\mathbf{F}^T \cdot \mathbf{W}^2 \cdot \mathbf{F} \cdot \vec{p} = \mathbf{F}^T \cdot \mathbf{W} \cdot \vec{y} \quad .$$

Le codage en MATLAB ou en Octave peut se faire de manière comparable à la section 4.3.3.

#### 4.5 Adaptation d'une courbe exponentielle

**4 Problème :** Des points de mesures  $(x_i, y_i)$  devraient se situer sur une courbe d'équation

$$f(x) = a \cdot x^b$$

et l'on cherche à déterminer les paramètres  $a$  et  $b$ .

x	0.6	0.7	0.8	0.9	1.	1.1	1.2	1.3
y	2.61445	2.78728	2.69749	2.89113	3.0522	3.22576	3.4746	3.68019

**Solution :**

- Ce problème **n'est pas** une application simple des calculs précédents, car la fonction  $f(x) = a \cdot x^b$  ne peut pas être exprimée sous la forme  $f(x) = p_1 \cdot f_1(x) + p_2 \cdot f_2(x)$ .
- En appliquant des deux côtés de l'équation  $f(x) = a \cdot x^b$  la fonction  $\ln$ , on construit une nouvelle fonction

$$g(x) = \ln f(x) = \ln (a \cdot e^{b \ln x}) = \ln a + b \ln x$$

Cette nouvelle fonction  $g$  dépend linéairement des deux constantes  $p_1 = \ln a$  et  $p_2 = b$ , et l'on peut utiliser les développements de la section précédente. Il s'agit alors de calculer  $p_1, p_2$  de manière à ce que la somme

$$\chi^2 = \sum_{i=1}^n (\ln a + b \ln x_i - \ln y_i)^2$$

soit minimale. Cette simplification est basée sur le fait que, sur une échelle logarithmique log-log, une fonction de la forme  $a \cdot x^b$  se transforme en une droite. De cette façon, les paramètres  $a$  et  $b$  peuvent être déterminés grâce à une régression linéaire, et en utilisant Octave ou MATLAB.

### Octave

```
x = [0.6 ; 0.7 ; 0.8 ; 0.9 ; 1. ; 1.1 ; 1.2 ; 1.3];  
y = [2.61445; 2.78728; 2.69749; 2.89113; 3.0522; 3.22576; 3.4746; 3.68019];  
lnx = log(x); lny = log(y);  
F = [ones(size(x)) lnx];  
p = LinearRegression(F, lny)
```

Avec les résultats numériques  $p = (p_1, p_2)^T = (1.14483, 0.43151)^T$ , on obtient

$$\begin{aligned}\ln(y) &= 1.14483 + 0.43151 \cdot \ln(x) \\ y &= e^{1.14483} x^{0.43151} = 3.1419 x^{0.43151}\end{aligned}$$

(c) Les valeurs ci-dessus des paramètres  $a$  et  $b$  ne minimisent **pas** la fonction

$$\sum_{i=1}^n (a \cdot x_i^b - y_i)^2$$

Pour trouver quelle est l'expression qui est effectivement minimisée, on doit regarder la fonction  $\chi^2$  un peu plus attentivement et utiliser l'approximation  $\ln(1+z) \approx z$  pour  $|z| \ll 1$ :

$$\begin{aligned}\chi^2 &= \sum_{i=1}^n (\ln a + b \ln x_i - \ln y_i)^2 = \sum_{i=1}^n \left( \ln \frac{a \cdot x_i^b}{y_i} \right)^2 \\ &= \sum_{i=1}^n \left( \ln \left( 1 + \frac{a \cdot x_i^b - y_i}{y_i} \right) \right)^2 \approx \sum_{i=1}^n \left( \frac{a \cdot x_i^b - y_i}{y_i} \right)^2\end{aligned}$$

Un minimum de  $\chi^2$  correspond donc à un minimum de la somme des carrés des **erreurs relatives**  $(ax_i^b - y_i)/y_i$ .

La recherche d'un minimum de la fonction

$$\xi^2 = \sum_{i=1}^n (a \cdot x_i^b - y_i)^2$$

revient à résoudre un problème d'interpolation non linéaire. Les résultats peuvent être très différents.  $\square$

**5 Exemple :** Pour trouver une courbe de la forme

$$f(x) = a \cdot e^{b \cdot x}$$

qui passe "le mieux possible" à travers une distribution de points, on utilise la transformation

$$\ln y = \ln a + b \cdot x$$

ce qui revient à trouver le minimum de la fonction

$$\sum_{i=1}^n (\ln a - b \cdot x_i - \ln y_i)^2$$

De même que dans l'exemple précédent, on utilise pour cela le code

## Octave

```
F = [ones(size(x)) x];  
p = LinearRegression(F, lny)
```

On obtient à partir des valeurs numériques  $p = (0.64732, 0.48645)^T$  le résultat

$$\begin{aligned}\ln(y) &= 0.64732 + 0.48645 \cdot x \\ y &= e^{0.64732} \cdot e^{0.43151 x} = 1.9104 e^{0.43151 x}\end{aligned}$$

◇

### 4.6 Erreur dans les valeurs de $x$ et $y$

Dans tous les calculs ci-dessus, nous avons pris l'hypothèse que les valeurs de  $x$  sont connues avec exactitude. Dans le cas contraire, il est nécessaire de modifier les calculs de manière conséquente (voir [Pres92, §15.3, p. 666] ou [Bevi69]).

### 4.7 Instruction *Octave*/MATLAB `LinearRegression()`

L'auteur de la présente notice (A. Stahel) a écrit une instruction *Octave* nommée `LinearRegression()` qui a les capacités suivantes:

- La régression linéaire peut être exécutée avec n'importe quelles fonctions de base. Les paramètres optimaux sont alors déterminés.
- La variance des grandeurs indépendantes et la longueur du vecteur résiduel sont calculées.
- La variance des paramètres est estimée.
- Trouver la documentation online ci-dessous. Trouver plus de documentation en [Octave07]

## Octave

```
octave:1> help LinearRegression  
general linear regression  
  
[p,y_var,r,p_var] = LinearRegression(F,y)  
[p,y_var,r,p_var] = LinearRegression(F,y,weight)  
  
determine the parameters p_j (j=1,2,...,m) such that the function  
f(x) = sum_(i=1,...,m) p_j*f_j(x) fits as good as possible to the  
given values y_i = f(x_i)  
  
parameters  
F n*m matrix with the values of the basis functions at the support points  
  in column j give the values of f_j at the points x_i (i=1,2,...,n)  
y n column vector of given values  
weight n column vector of given weights  
  
return values  
p m vector with the estimated values of the parameters  
y_var estimated variance of the error  
r weighted norm of residual  
p_var estimated variance of the parameters p_j
```

Une documentation détaillée et des exemples d'application sont inclus dans la documentation relative à MATLAB et *Octave*.



## Références

- [Bevi69] P. R. Bevington. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw–Hill, New York, 1969.
- [Pres92] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C, The Art of Scientific Computing*. Cambridge University Press, second edition, 1992.
- [Octave07] A. Stahel. Octave at the HTI Biel. lecture notes, 2007.