Chapitre 8

Modélisation des données

1. Maximum de vraisemblance
   - Statistique de Poisson
   - Statistique de Gauss
2. Principe général
   - Statistique de Gauss
3. Ajustement d'une droite
4. Ajustement d'une fonction linéaire
5. Ajustement d'une fonction non linéaire
   - Méthode de Levenberg-Marquardt

Maximum de vraisemblance

On désire ajuster une fonction («modèle») $f(x)$ dépendant de certains paramètres a_j $(j = 1, \ldots, M)$:

$$f(x; a_1, \ldots, a_M)$$

sur des données

$$y_j \quad (j = 1, \ldots, N)$$

entachées d'une incertitude

$$\sigma_j \quad (j = 1, \ldots, N)$$

N = nombre de données
M = nombre de paramètres

Si nous supposons *a priori* équiprobables toutes les valeurs que peuvent prendre les paramètres, le modèle le plus probable, *compte tenu des données* y_j, est celui pour lequel la *vraisemblance*

$$P(r \mid q^i . p)$$

est maximale

Dans le cas présent, on a :

$$
\begin{aligned}
p &= \text{les connaissances de base (lois de la physique,...)} \\
r &= \text{le résultat de l'expérience} \\
&= \text{l'ensemble des données } y_j \pm \sigma_j \\
q^{\,i} &= \text{un modèle particulier} \\
&= \text{la fonction } f(x) \text{ pour un choix particulier} \\
&\ \text{des paramètres } a_1, \ldots, a_M
\end{aligned}
$$

Si nous supposons les erreurs sur les différentes mesures y_j indépendantes les unes des autres, nous pouvons écrire :

$$P(\,r \mid q^{\,i} \,.\, p\,) = \prod_{j=1}^{N} P(\,y_j \mid q^{\,i} \,.\, p\,)$$

Or,

$$P(\,y_j \mid q^{\,i} \,.\, p\,) = P(\,y_j \mid f(x_j; a_1^i, \ldots, a_M^i) \,.\, p\,)$$

Remarque :

l'effet du profil instrumental est supposé inclus dans $f(x)$

Cas particuliers :

1. Statistique de Poisson

Soit

$$f_j = f(x_j; a_1^i, \ldots, a_M^i)$$

$$P(\, y_j \mid f_j \,.\, p\,) = \frac{{f_j}^{y_j}}{y_j!}\, e^{-f_j}$$

$$\Rightarrow P(\, r \mid q^{\,i} \,.\, p\,) = \prod_{j=1}^{N} \frac{{f_j}^{y_j}}{y_j!}\, e^{-f_j}$$

$$\Rightarrow \ln P(\, r \mid q^{\,i} \,.\, p\,) = C + \sum_{j=1}^{N} (y_j \ln f_j - f_j)$$

Les paramètres $a_1, \ldots, a_M$ cherchés sont ceux qui maximisent

$$\mathcal{S}_P = \sum_{j=1}^{N} (y_j \ln f_j - f_j)$$

2. Statistique de Gauss

$$P(y_j \mid f_j . p) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(y_j - f_j)^2}{2\sigma_j{}^2}}$$

$$\Rightarrow P(r \mid q^i . p) = \frac{1}{\sqrt{2\pi}} \prod_{j=1}^{N} \frac{1}{\sigma_j} e^{-\frac{(y_j - f_j)^2}{2\sigma_j{}^2}}$$

$$\Rightarrow \ln P(r \mid q^i . p) = C - \frac{1}{2} \sum_{j=1}^{N} \left(\frac{y_j - f_j}{\sigma_j} \right)^2$$

Les paramètres $a_1, \ldots, a_M$ cherchés sont ceux qui maximisent

$$\mathcal{S}_G = -\frac{1}{2} \sum_{j=1}^{N} \left(\frac{y_j - f_j}{\sigma_j} \right)^2$$

et donc qui minimisent

$$\chi^2 = \sum_{j=1}^{N} \left(\frac{y_j - f_j}{\sigma_j} \right)^2$$

Principe général

1. Construire une «fonction de mérite» qui mesure l'accord entre le modèle et les données

(dans notre cas, cette «fonction de mérite» est déduite du maximum de vraisemblance)

2. Déterminer les paramètres du modèle de façon à minimiser la «fonction de mérite»

3. Estimer dans quelle mesure le modèle est compatible avec les données

4. Estimer l'incertitude sur les paramètres ajustés

Statistique de Gauss :

Minimiser

$$\chi^2 = \sum_{j=1}^{N} \left(\frac{y_j - f_j(a_1, \ldots, a_M)}{\sigma_j} \right)^2$$

χ^2 = somme de carrés de N variables aléatoires normales

Lorsque $a_1, \ldots, a_M$ sont ajustés pour minimiser χ^2, les termes de la somme ne sont plus statistiquement indépendants

Cependant, les différentes valeurs de χ^2 à son minimum suivent une distribution du chi-carré à $\nu = N - M$ degrés de liberté

La probabilité d'obtenir par hasard une valeur du chi-carré au moins aussi élevée que celle que l'on obtient, χ^2, est donnée par

$$Q(\frac{\nu}{2}, \frac{\chi^2}{2}) = 1 - P(\frac{\nu}{2}, \frac{\chi^2}{2})$$

où $P(a, x)$ est la *fonction gamma incomplète* :

$$P(a, x) \equiv \frac{\gamma(a, x)}{\Gamma(a)} \equiv \frac{1}{\Gamma(a)} \int_0^x e^{-t}\, t^{a-1}\, dt$$

Si Q est petit, c'est qu'il est peu probable qu'une fluctuation statistique soit responsable d'une si grande valeur de χ^2

$\Rightarrow$ soit :

(1) le modèle est erroné

(2) les incertitudes de mesure ont été sous-estimées

(3) les erreurs de mesure ne suivent pas une distribution normale

En règle générale, on s'attend à avoir une valeur

$$\chi^2 \simeq \nu = N - M$$

Plus précisément, pour les grandes valeurs de ν, χ^2 suit une distribution normale

– de moyenne ν

– d'écart-type $\sigma = \sqrt{2\nu}$

Ajustement d'une droite

Nous désirons ajuster la fonction

$$f(x; a_1, a_2) = a_1 + a_2\, x$$

sur les points

$$(x_j, y_j \pm \sigma_j) \quad (j = 1, \ldots, N)$$

où nous supposons les x_j connus exactement et les erreurs de mesure sur y_j distribuées normalement

(1) *Estimation des paramètres*

Minimiser :

$$\chi^2(a_1, a_2) = \sum_{j=1}^{N} \left(\frac{y_j - a_1 - a_2\, x_j}{\sigma_j} \right)^2$$

Remarque : si les erreurs ne sont pas distribuées normalement, la minimisation du χ^2 ne correspond pas au maximum de vraisemblance, mais peut néanmoins donner de bons résultats

$$\begin{cases} \dfrac{\partial \chi^2}{\partial a_1} = 0 \Rightarrow a_1 S + a_2 S_x = S_y \\ \\ \dfrac{\partial \chi^2}{\partial a_2} = 0 \Rightarrow a_1 S_x + a_2 S_{xx} = S_{xy} \end{cases}$$

avec

$$S \equiv \sum_{j=1}^{N} \frac{1}{\sigma_j^2}$$

$$S_x \equiv \sum_{j=1}^{N} \frac{x_j}{\sigma_j^2}$$

$$S_y \equiv \sum_{j=1}^{N} \frac{y_j}{\sigma_j^2}$$

$$S_{xx} \equiv \sum_{j=1}^{N} \frac{x_j^2}{\sigma_j^2}$$

$$S_{xy} \equiv \sum_{j=1}^{N} \frac{x_j\, y_j}{\sigma_j^2}$$

Posons

$$\Delta \equiv S\, S_{xx} - S_x^2$$

$$\Rightarrow \begin{cases} a_1 = \dfrac{S_{xx}\, S_y - S_x\, S_{xy}}{\Delta} \\ \\ a_2 = \dfrac{S\, S_{xy} - S_x\, S_y}{\Delta} \end{cases}$$

(2) *Estimation de l'incertitude sur les paramètres*

Si les erreurs sur les différentes mesures sont non corrélées :

$$\sigma_{a_i}^2 = \sum_{k=1}^{N} \sigma_k^2 \left(\frac{\partial a_i}{\partial y_k}\right)^2$$

$$\frac{\partial a_1}{\partial y_k} = \frac{S_{xx} - x_k\, S_x}{\sigma_k^2\, \Delta}$$

$$\frac{\partial a_2}{\partial y_k} = \frac{x_k\, S - S_x}{\sigma_k^2\, \Delta}$$

$$\Rightarrow \begin{cases} \sigma_{a_1}^2 = \dfrac{S_{xx}}{\Delta} \\[2ex] \sigma_{a_2}^2 = \dfrac{S}{\Delta} \end{cases}$$

Covariance :

$$\mathrm{Cov}(a_1, a_2) = \sum_{k=1}^{N} \sigma_k^2\, \frac{\partial a_1}{\partial y_k}\, \frac{\partial a_2}{\partial y_k}$$

$$\Rightarrow\ \mathrm{Cov}(a_1, a_2) = -\frac{S_x}{\Delta}$$

(3) *Accord entre le modèle et les données*

Calculer :

$$Q = 1 - P(\frac{\nu}{2}, \frac{\chi^2}{2})$$

– Si $Q \gtrsim 0.1$, on peut estimer que l'ajustement est satisfaisant

– Si $Q \gtrsim 0.001$, l'ajustement est peut-être acceptable (les erreurs ne sont peut-être pas distribuées normalement, ou ont peut-être été sous-estimées)

– Si $Q \lesssim 0.001$, on doit se poser des questions : le modèle n'est probablement pas une bonne représentation des données

– Si $Q \simeq 1$, on doit aussi se poser des questions : les erreurs ont peut-être été surestimées (à moins que les données n'aient été « arrangées »...)

Cas particulier : si les σ_j ne sont pas connus

$\Rightarrow$ on peut les estimer à partir de l'ajustement :

on calcule a_1 et a_2 par les formules habituelles, avec

$$\sigma_j = 1, \ \ \forall j$$

l'incertitude sur les données peut alors être estimée par :

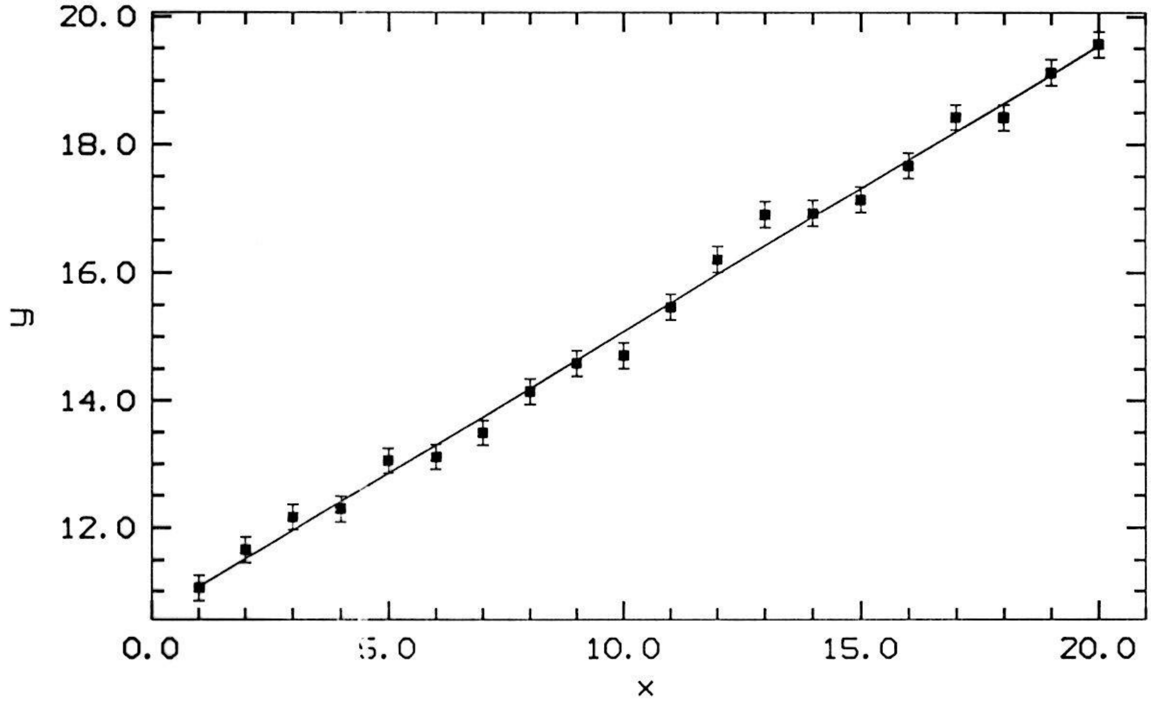
$$\sigma^2 = \frac{1}{N} \sum_{j=1}^{N} (y_j - a_1 - a_2 x_j)^2$$

σ_{a_1} et σ_{a_2} sont calculés par les formules habituelles, multipliées par $\sqrt{\chi^2/(N-2)}$

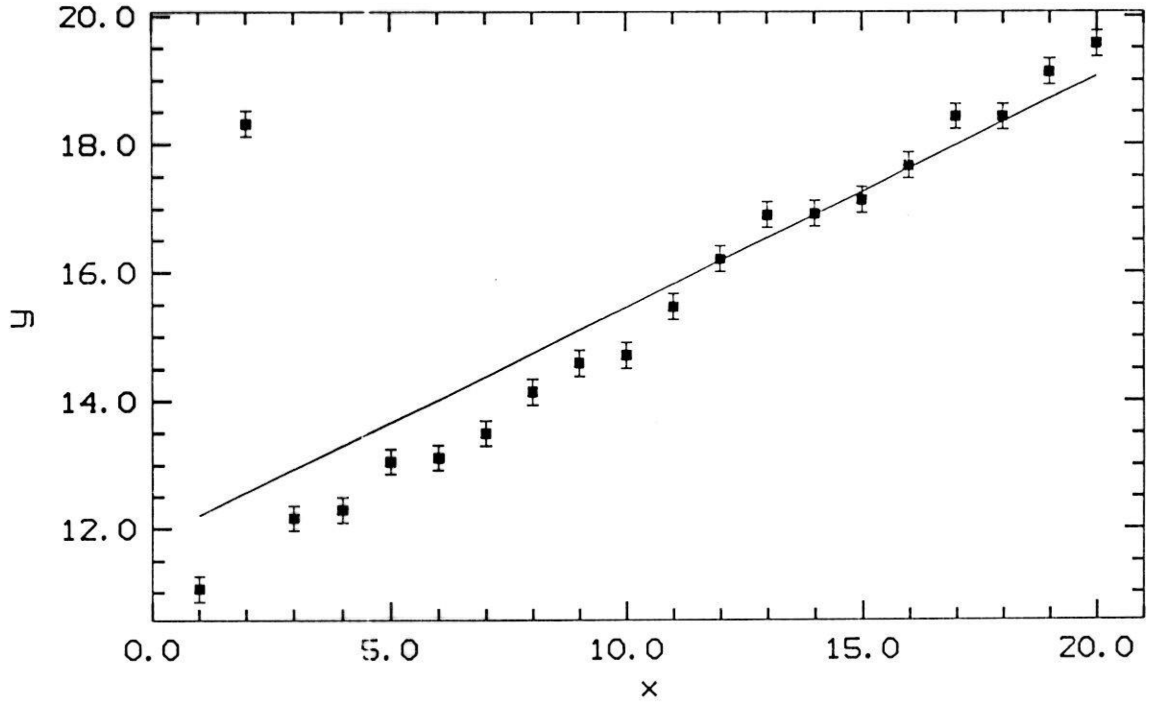
Attention !

Tout ceci présuppose que le modèle fournit une bonne représentation des données

$\Rightarrow$ pas de test indépendant de l'accord entre le modèle et les données



$$a_1 = 10.63 \pm 0.09 \quad a_2 = 0.446 \pm 0.008 \quad \chi^2 = 19.7$$



$$a_1 = 11.85 \pm 0.09 \quad a_2 = 0.360 \pm 0.008 \quad \chi^2 = 999.6$$

Ajustement d'une fonction linéaire

Soit

$$f(x; a_1, \ldots, a_M)$$

une fonction qui dépend linéairement des paramètres $a_1, \ldots, a_M$

Exemples :

$$f(x) = a_1 + a_2\, x + a_3\, x^2 + a_4\, x^3 + \cdots$$

$$f(x) = a_1 \cos x + a_2 \ln x$$

Forme générale :

$$f(x) = \sum_{i=1}^{M} a_i\, F_i(x)$$

où $F_1(x), \ldots, F_M(x)$ sont des fonctions arbitraires de x

(« fonctions de base »)

(1) *Estimation des paramètres*

Minimiser

$$\chi^2 = \sum_{j=1}^{N} \left[\frac{y_j - \Sigma_{i=1}^{M} a_i \, F_i(x_j)}{\sigma_j} \right]^2$$

$$\frac{\partial \chi^2}{\partial a_k} = 0 \Leftrightarrow \sum_{j=1}^{N} \frac{1}{\sigma_j^2} \left[y_j - \sum_{i=1}^{M} a_i \, F_i(x_j) \right] F_k(x_j) = 0$$

Posons

$$\alpha_{ki} \equiv \sum_{j=1}^{N} \frac{F_i(x_j) \, F_k(x_j)}{\sigma_j^2}$$

$$\beta_k \equiv \sum_{j=1}^{N} \frac{y_j \, F_k(x_j)}{\sigma_j^2}$$

$$\Rightarrow \sum_{i=1}^{M} \alpha_{ki} \, a_i = \beta_k \quad (k = 1, \ldots, M)$$

système linéaire de M équations à M inconnues

Soit C_{ik} la matrice inverse de α_{ik}

$$\Rightarrow a_i = \sum_{k=1}^{M} C_{ik} \, \beta_k$$

(2) *Estimation de l'incertitude sur les paramètres*

$$\sigma_{a_i}^2 = \sum_{j=1}^{N} \sigma_j^2 \left(\frac{\partial a_i}{\partial y_j} \right)^2$$

$$\frac{\partial a_i}{\partial y_j} = \sum_{k=1}^{M} C_{ik} \frac{\partial \beta_k}{\partial y_j}$$

$$\frac{\partial \beta_k}{\partial y_j} = \frac{F_k(x_j)}{\sigma_j^2}$$

$$\Rightarrow \sigma_{a_i}^2 = \sum_{j=1}^{N} \sigma_j^2 \sum_{k=1}^{M} C_{ik} \frac{F_k(x_j)}{\sigma_j^2} \sum_{l=1}^{M} C_{il} \frac{F_l(x_j)}{\sigma_j^2}$$

$$\Rightarrow \sigma_{a_i}^2 = \sum_{k=1}^{M} \sum_{l=1}^{M} C_{ik}\, C_{il}\, \alpha_{kl}$$

Or,

$$\sum_{k=1}^{M} C_{ik}\, \alpha_{kl} = \delta_{il}$$

$$\Rightarrow \sigma_{a_i}^2 = C_{ii}$$

Les éléments diagonaux de C sont les variances des paramètres ajustés

On montre de même que les autres éléments sont les covariances

Remarque :

La solution directe du problème de l'ajustement linéaire par moindre carrés est assez susceptible d'être sensible aux erreurs numériques

Des méthodes différentes permettent d'éviter ces problèmes

(voir le chapitre 14 de *Numerical Recipes*, par W.H. Press, B.P. Flannery, S.A. Teukolsky et W.T. Vetterling, *Cambridge University Press*)

Ajustement d'une fonction non linéaire

Soit

$$f(x; a_1, \ldots, a_M)$$

une fonction qui dépend non linéairement des paramètres $a_1, \ldots, a_M$

Exemples :

$$f(x) = a_1 \cos(a_2 x)$$

$$f(x) = a_1 e^{-a_2 x}$$

⇨ Méthode <u>itérative</u> nécessaire

Posons

$$(a_1, \ldots, a_M) = \vec{a}$$

$$\chi^2(a_1, \ldots, a_M) = S(\vec{a})$$

1^{er} cas : si on est près du minimum $\vec{a}_{\min}$

$$\Rightarrow S(\vec{a}) \simeq \gamma - \vec{d} \cdot \vec{a} + \frac{1}{2} \vec{a} \, D \, \vec{a} \qquad (1)$$

où $\vec{d}$ = vecteur à M dimensions (- dérivée $1^{\text{ère}}$)

D = matrice $M \times M$ (dérivée 2^{de})

Supposons (1) exacte

$$\Rightarrow \nabla S(\vec{a}) = D \, \vec{a} - \vec{d} \qquad (2)$$

Au minimum, on a :

$$\nabla S(\vec{a}_{\min}) = 0$$

$$\Rightarrow D \, \vec{a}_{\min} = \vec{d} \qquad (3)$$

En un point $\vec{a}_0 \neq \vec{a}_{\min}$, on a :

$$\Rightarrow D \, \vec{a}_0 = \nabla S(\vec{a}_0) + \vec{d} \qquad (4)$$

$(3) - (4)$: $$D \, (\vec{a}_{\min} - \vec{a}_0) = -\nabla S(\vec{a}_0) \qquad (5)$$

$$\Rightarrow \vec{a}_{\min} = \vec{a}_0 - D^{-1} \, [\nabla S(\vec{a}_0)] \qquad (6)$$

2^e^ cas : si on est loin du minimum :

(1) peut être une très mauvaise approximation de $S(\vec{a})$

$\Rightarrow$ tout ce qu'on peut faire pour se rapprocher de $\vec{a}_{\text{min}}$ est de se déplacer dans la direction opposée au gradient :

$$\Rightarrow \vec{a}_1 = \vec{a}_0 - \eta \nabla S(\vec{a}_0) \tag{7}$$

où η = constante positive à déterminer

Calcul de $\vec{d}$ et D

Soit

$$f = f(x; \vec{a})$$

le modèle à ajuster

$$\chi^2(\vec{a}) = \sum_{j=1}^{N} \left[\frac{y_j - f(x_j; \vec{a})}{\sigma_j} \right]^2$$

$$\nabla S = \left(\frac{\partial \chi^2}{\partial a_1}, \ldots, \frac{\partial \chi^2}{\partial a_M} \right)$$

$$\frac{\partial \chi^2}{\partial a_k} = -2 \sum_{j=1}^{N} \frac{y_j - f(x_j; \vec{a})}{\sigma_j^2} \frac{\partial f(x_j; \vec{a})}{\partial a_k}$$

$$\frac{\partial^2\chi^2}{\partial a_k \partial a_l} =$$

$$= 2\sum_{j=1}^{N}\frac{1}{\sigma_j^2}\left[\frac{\partial f(x_j;\vec{a})}{\partial a_k}\frac{\partial f(x_j;\vec{a})}{\partial a_l} - [y_j - f(x_j;\vec{a})]\frac{\partial^2 f(x_j;\vec{a})}{\partial a_k \partial a_l}\right]$$

$$\simeq 2\sum_{j=1}^{N}\frac{1}{\sigma_j^2}\left[\frac{\partial f(x_j;\vec{a})}{\partial a_k}\frac{\partial f(x_j;\vec{a})}{\partial a_l}\right]$$

Raisons qui permettent de négliger la dérivée seconde :

– elle est nulle dans le cas linéaire

– si l'ajustement est bon, son coefficient multiplicatif est l'erreur de mesure sur y_j ; ces erreurs de mesure s'annulent approximativement en moyenne :

$$\sum_{j=1}^{N}\frac{1}{\sigma_j^2}[y_j - f(x_j;\vec{a})] \simeq 0$$

– une approximation sur le calcul des dérivées secondes ne modifie en rien le résultat, qui est déterminé par la condition que la dérivée première s'annule ; elle ne modifie que la manière de parvenir au résultat $\Rightarrow$ la convergence

Posons

$$\beta_k \equiv -\frac{1}{2}\frac{\partial \chi^2}{\partial a_k} = +\frac{1}{2}\, d_k$$

$$\alpha_{kl} \equiv \frac{1}{2}\,\frac{\partial^2 \chi^2}{\partial a_k \partial a_l} = \frac{1}{2}\, D_{kl}$$

α_{kl} = matrice de courbure

Près du minimum :

(5) devient :

$$\sum_{l=1}^{M} \alpha_{kl}\, \delta a_l = \beta_k \tag{8}$$

Loin du minimum :

(7) devient :

$$\delta a_k = \eta\, \beta_k \tag{9}$$

Méthode de Levenberg-Marquardt

On voudrait connaître l'ordre de grandeur de η

Or, le gradient ne donne aucune information sur η ; il n'indique que la direction de la pente, mais pas jusqu'où elle s'étend

Marquardt $\Rightarrow$ les composantes de la matrice de courbure contiennent de l'information quant à l'ordre de grandeur de η :

χ^2 est sans dimension

β_k a la dimension de $\frac{1}{a_k}$

$\Rightarrow$ η doit avoir la dimension de a_k^2

Or, il n'y a dans la matrice de courbure que $\frac{1}{\alpha_{kk}}$ qui ait la dimension de a_k^2

$\Rightarrow$ on suppose que $\frac{1}{\alpha_{kk}}$ définit l'échelle de η

Mais il faut se garder de choisir η trop grand, sinon on pourrait dépasser le minimum

⇒ on choisit

$$\eta_k = \frac{1}{\lambda\, \alpha_{kk}} \qquad (\lambda > 0)$$

⇒ (9) devient :

$$\delta a_k = \frac{1}{\lambda\, \alpha_{kk}}\, \beta_k$$

ou encore :

$$\lambda\, \alpha_{kk}\, \delta a_k = \beta_k \tag{10}$$

⇒ on peut combiner (8) et (10) en écrivant :

$$\sum_{l=1}^{M} \alpha'_{kl}\, \delta a_l = \beta_k \tag{11}$$

avec

$$\alpha'_{kk} = \alpha_{kk}\, (1 + \lambda)$$

et

$$\alpha'_{kl} = \alpha_{kl} \qquad (k \neq l)$$

Lorsque λ est grand, on se déplace selon la plus grande pente

Lorsque λ est petit, on retombe sur la méthode quadratique (métrique variable, *quasi-Newton*)

Algorithme :

(1) choisir une valeur initiale de $\vec{a}$

(2) calculer $\chi^2(\vec{a})$

(3) prendre une valeur initiale de λ (ex.: $\lambda = 0.001$)

(4) résoudre (11) et calculer $\chi^2(\vec{a} + \delta\vec{a})$

(5) si $\chi^2(\vec{a} + \delta\vec{a}) > \chi^2(\vec{a})$, augmenter λ d'un facteur 10 et retourner en (4)

(6) si $\chi^2(\vec{a} + \delta\vec{a}) < \chi^2(\vec{a})$, diminuer λ d'un facteur 10, se déplacer de $\vec{a}$ en $\vec{a} + \delta\vec{a}$ et retourner en (4)

Critère d'arrêt :

– s'arrêter lorsque χ^2 décroît d'une valeur négligeable (ex.: 0.001)

– ne jamais s'arrêter après une augmentation de λ

Une fois le minimum atteint :

Poser $\lambda = 0$ puis calculer la matrice de covariance

$$C = \alpha^{-1}$$