Cours Méthodes d'optimisation – Algorithmes déterministes

S. Chevalier – Document débuté le 14 Mars 2014

1 Avant Propos

Ce document présente quelques méthodes de base concernant les algorithmes d'optimisation utilisés pour comparer une loi, un modèle à des mesures expérimentales. Cette approche de comparaison théorie/mesures dans le but d'identifier des paramètres physiques est au cœur du propos de ce document. Toutes les méthodes exposées ici sont dîtes d'optimisation sans contraintes. Nous ne traiterons que le cas des fonctions quadratiques telle que définies par les moindres carrés.
Nous nous intéressons tout d'abord aux problèmes linéaires, puis pour les lois/modèles plus compliqués, et donc très souvent non linéaires, nous aborderons les méthodes de plus grandes pente de de gradients conjugués.

2 Méthode des moindres carrés

2.1 Démonstration

Dans cet exemple, nous nous intéressons aux lois polynomiales du type :

$$y = \sum_{n=0}^{\infty} a_n.x^n \tag{1}$$

Le but est de rechercher l'ensemble des coefficients a_n connaissant les valeurs $\tilde{y}$ mesurées à différentes coordonnées x. L'écart entre cette mesure et la loi y s'écrit :

$$J = \frac{1}{2}\sum_{i=1}^{N}(\tilde{y}(x_i) - y(x_i))^2 \tag{2}$$

Ecrit sous forme matricielle, la relation 2 devient :

$$J = \frac{1}{2}\|\tilde{\mathbf{Y}} - \mathbf{Y}\|^2 \tag{3}$$

Où $\tilde{\mathbf{Y}} = [\tilde{y}(x_1)\tilde{y}(x_2)....\tilde{y}(x_n)]^T$ est le vecteur contenant les mesures. La relation 1 peut elle aussi s'écrire sous forme matricielle, ce qui donne :

$$\mathbf{Y} = [\mathbf{X}].(\boldsymbol{\beta}) \tag{4}$$

avec :

$$\mathbf{X} = \begin{bmatrix} x_1^n & x_1^{n-1} & \dots & x_1 & 1 \\ x_2^n & x_2^{n-1} & \dots & x_2 & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ x_m^n & x_m^{n-1} & \dots & x_m & 1 \end{bmatrix} \; et \; \boldsymbol{\beta} = \begin{pmatrix} a_n \\ a_{n-1} \\ \vdots \\ a_1 \\ a_0 \end{pmatrix}$$

L'objectif est ici de trouver les valeurs des coefficients a et b connaissant la relation 4 et des mesures $\tilde{\mathbf{Y}}$. Les valeurs de a et b sont celles qui parviendront à minimiser J qui est l'écart entre la loi et les mesures, et donc à minimiser la relation 2. Autrement dit, nous cherchons :

$$Arg\min(J) = 0 \tag{5}$$

Dans le cas des moindres carrés, cette relation est vérifiée lorsque le gradient de J par rapport au vecteur paramètre β est nulle. Il s'écrit de la manière suivante :

$$\begin{aligned} \nabla_{\beta}.J &= 0 & (6) \\ \frac{1}{2}\nabla_{\beta}.\|\tilde{\mathbf{Y}} - [\mathbf{X}].(\boldsymbol{\beta})\|^2 &= 0 & (7) \\ -\mathbf{X}^T.\|\tilde{\mathbf{Y}} - [\mathbf{X}].(\boldsymbol{\beta})\| &= 0 & (8) \end{aligned}$$

Ce qui permet d'expliciter directement la valeur du vecteur $\boldsymbol{\beta}$ telle que :

$$\boldsymbol{\beta} = [\mathbf{X}^T.\mathbf{X}]^{-1}.[\mathbf{X}^T.\tilde{\mathbf{Y}}] \quad (9)$$

2.2 Exemple avec un polynôme d'ordre 1

Soit la loi suivante :

$$y = a.x + b = [\mathbf{X}].(\boldsymbol{\beta}) \quad (10)$$

Nous recherchons les coefficients a et b qui permette de minimiser l'écart quadratique entre la relation 10 et les mesures de la figure (?). Nous considérons ici 10 points de mesures répartis de 0 à 1, les matrices sont alors les suivantes :

$$\mathbf{X} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_{10} & 1 \end{bmatrix} \; et \; \boldsymbol{\beta} = \begin{pmatrix} a \\ b \end{pmatrix}$$

Nous prendrons les valeurs suivantes : $a = -4$ et $b = 4$. Les mesures sont générées à partir de la relation 10 auquel on a rajouté un bruit normal de 15%. Les résultats sont présentés sur la figure (?), nous obtenons les coefficients suivants :

$$\boldsymbol{\beta} = \begin{pmatrix} -3,73 \\ 3,72 \end{pmatrix}$$

A noter que ces valeurs ne sont pas rigoureusement égales à celles qui ont servies à générées les mesures. L'ajout du bruit provoque une certaine incertitude lors de l'identification des mesures. Ceci illustre donc la nécessité de minimiser le bruit de mesure afin de garantir une identification fiable des paramètres. La même démarche peut être réalisée pour tous les polynôme d'ordre n.

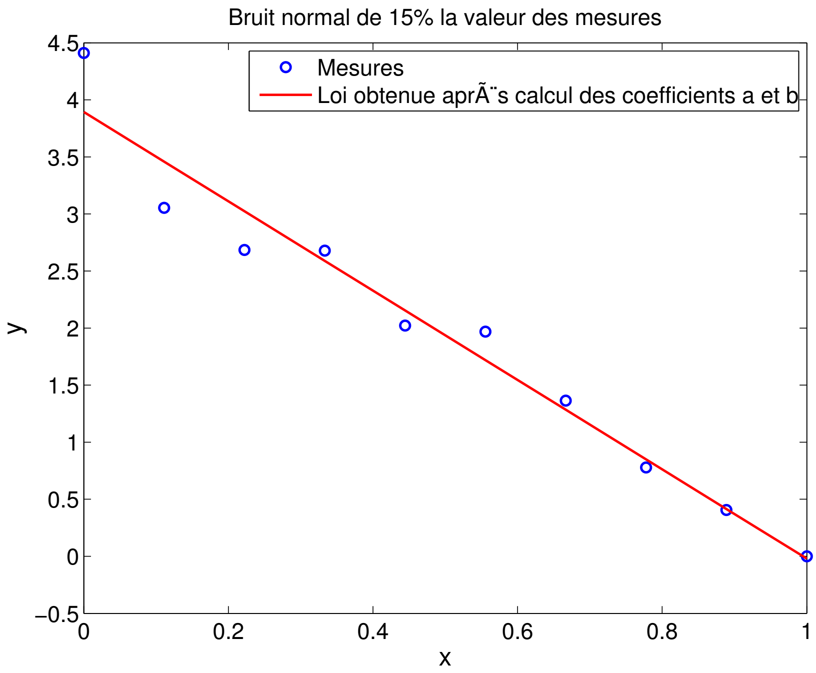

FIGURE 1 – Résultats de l'optimisation de l'équation 4 avec les mesures par la méthode des moindres carrés.

3 Méthode de la plus grande pente

3.1 Présentation générale

La section précédente vient de présenter comment la méthode des moindres carrés permettait le calcul des coefficients de polynômes, c'est-à-dire dans les cas linéaires. Nous nous intéressons maintenant aux lois non-linéaires par rapport aux paramètres. Prenons la loi suivante, possédant n coefficients à déterminer :

$$y = f(a^n, x) \tag{11}$$

Nous cherchons ici à obtenir les coefficients a^n d'après les données $\tilde{\mathbf{Y}}$. Le critère au sens des moindres carrés s'écrit toujours :

$$J = \frac{1}{2}.\|\tilde{\mathbf{Y}} - \mathbf{Y}.(\boldsymbol{\beta})\|^2 \tag{12}$$

Nous cherchons donc à trouver les paramètres a^n qui minimisent le critère J, soit :

$$\boldsymbol{\beta} = Arg\min(J) \tag{13}$$

Où $\boldsymbol{\beta}$ est le vecteur paramètre. Ils sont obtenus pour :

$$\nabla_{\boldsymbol{\beta}}.J = 0 \tag{14}$$

La méthode consiste donc à identifier le vecteur $\boldsymbol{\beta}$ vérifiant l'équation 14. Il est déterminé par itération de la manière suivante :

$$(\boldsymbol{\beta})^{k+1} = (\boldsymbol{\beta})^k + r_{opt}^k.d^k \tag{15}$$

Où d^k est la direction de descente et r_{opt}^k la profondeur de descente optimale. L'équation 15 vérifie donc :

$$J^{k+1} < J^k \tag{16}$$

Dans le cas de la méthode de la plus grande pente, la direction de descente correspond à la direction opposée du gradient, et la profondeur de descente est la longueur optimale dans la direction d^k. Nous obtenons ainsi :

$$d^n = -\nabla_\beta.J = \begin{pmatrix} -\frac{\partial J}{\partial a^0} \\ -\frac{\partial J}{\partial a^1} \\ \vdots \\ -\frac{\partial J}{\partial a^n} \end{pmatrix} \tag{17}$$

r^k se détermine ainsi :

$$r^k_{opt} = Arg\min(J((\boldsymbol{\beta})^k + r^k.d^k)) \tag{18}$$

L'identification de la profondeur de descente revient à effectuer une optimisation monodimensionnelle. Celle-ci peut être effectuée au moyen des méthodes classiques de la dichotomie, de Newton ou encore du point fixe.

3.2 Exemple d'application

A terminer...