

Moindres carrés non linéaires: Méthode de Levenberg-Marquardt

ISIMA 2F4

J. Koko

La méthode de Levenberg-Marquardt est la méthode de minimisation standard pour les problèmes de moindres carrés non linéaires. Elle combine astucieusement la robustesse de la méthode de la plus forte pente (lorsqu'on est loin de l'optimum) et l'efficacité de la méthode de Newton (au voisinage de l'optimum).

1 Moindres carrés non linéaires

Soit un modèle à ajuster $y = y(x; \mathbf{a})$, dépendant non linéairement de n paramètres $\mathbf{a} = (a_1, \dots, a_n)^T$. On dispose de p mesures (x_i, y_i) , $i = 1, \dots, p$. Soit f la fonction, de \mathbb{R}^n dans \mathbb{R}^p , définie par

$$f(\mathbf{a}) = (f_1(\mathbf{a}), \dots, f_p(\mathbf{a}))^T$$

avec $f_i(\mathbf{a}) = y(x_i; \mathbf{a}) - y_i$, pour $i = 1, \dots, p$.

L'erreur, en norme Euclidienne, entre les mesures et le modèle est alors donnée par

$$e(\mathbf{a}) = \frac{1}{2} f(\mathbf{a})^T f(\mathbf{a}) = \frac{1}{2} \|f(\mathbf{a})\|_{\mathbb{R}^p}^2.$$

L'ajustement consiste donc à trouver le paramètre \mathbf{a} qui minimise la fonction $e(\mathbf{a})$, i.e. résoudre le problème d'optimisation

$$\min_{\mathbf{a} \in \mathbb{R}^n} e(\mathbf{a}).$$

2 La méthode de Levenberg-Marquardt

On note $J_f(\mathbf{a})$ la matrice jacobienne de la multi-application f , i.e. la matrice $p \times n$ suivante

$$J_f(\mathbf{a}) = \nabla f(\mathbf{a}) = \left(\frac{\partial f_i}{\partial a_j}(\mathbf{a}) \right)_{i,j} \quad (1)$$

Avec (1), le gradient de e est donné par

$$\nabla e(\mathbf{a}) = J_f(\mathbf{a})^T f(\mathbf{a}) \quad (2)$$

Dans la méthode de Levenberg-Marquardt, la matrice Hessienne $H(\mathbf{a})$ de e est approchée par

$$H(\mathbf{a}) = J_f(\mathbf{a})^T J_f(\mathbf{a}), \quad (3)$$

beaucoup plus facile à évaluer que la matrice Hessienne exacte.

On considère maintenant un processus de minimisation itératif avec une mise à jour de la

$$\mathbf{a}^{k+1} = \mathbf{a}^k + \mathbf{d}^k, \quad k = 0, 1, \dots,$$

où \mathbf{d} est la direction de descente. Si on utilise l'algorithme de la plus *forte pente*, on a

$$\mathbf{d}^k = -\lambda_k \nabla e(\mathbf{a}^k), \quad (4)$$

où $\lambda_k > 0$ est le pas de déplacement. Dans le cas où la matrice Hessienne (3) est définie positive (ce qui n'est pas toujours le cas!!), on peut prendre \mathbf{d}^k comme solution du système linéaire

$$H(\mathbf{a}^k) \mathbf{d}^k = -\nabla e(\mathbf{a}^k). \quad (5)$$

L'idée principale consiste à utiliser (4) lorsqu'on est loin de l'optimum, puis de lui substituer (5) à l'approche de l'optimum. On remplace alors (4) et (5) par une formule unique

$$(H(\mathbf{a}^k) + \lambda_k \mathbb{I}_n) \mathbf{d}^k = -\nabla e(\mathbf{a}^k), \quad (6)$$

où \mathbb{I}_n est la matrice identité d'ordre n . Le paramètre λ_k doit être ajuster pour que \mathbf{d}^k solution de (6) soit toujours une direction de descente.

3 L'algorithme de Levenberg-Marquardt

Initialisation $k = 0$, \mathbf{a}^0 donné, $\lambda_0 = 0.001$

Repéter $k \geq 0$

Calculer \mathbf{d}^k solution de

$$(J_f(\mathbf{a}^k)^T J_f(\mathbf{a}^k) + \lambda_k \mathbb{I}_n) \mathbf{d}^k = -J_f(\mathbf{a}^k)^T f(\mathbf{a}^k)$$

Si $e(\mathbf{a}^k + \mathbf{d}^k) < e(\mathbf{a}^k)$ (descente)

$$\lambda^{k+1} = \lambda^k / 10$$

$$\mathbf{a}^{k+1} = \mathbf{a}^k + \mathbf{d}^k$$

$$k = k + 1$$

Sinon (pas de descente)

$$\lambda_{k+1} = 10 \lambda_k$$

$$\mathbf{a}^{k+1} = \mathbf{a}^k$$

$$k = k + 1$$

fin si

Jusqu'à $\|J_f(\mathbf{a}^k)^T f(\mathbf{a}^k)\| < \varepsilon$

Pour le test d'arrêt, on prendra $0.0001 \leq \varepsilon \leq 0.01$. Le fait d'augmenter λ quand $e(\mathbf{a}^k + \mathbf{d}^k) \geq e(\mathbf{a}^k)$ signifie qu'on renforce la diagonale de la matrice pour s'éloigner de la matrice singulière.

4 Le TP

Programmer la méthode de Levenberg-Marquardt pour les problèmes de moindres carrés non linéaires. La mise au point du programme se fera sur l'échantillon bidimensionnel suivant

```
x 0.01 1.93 2.95 3.26 4.18 5.73 6.29 7.70 8.91 9.12
y 0.98 0.84 0.80 0.78 0.82 0.78 0.80 0.85 0.90 0.95
```

On fait l'hypothèse que le modèle non linéaire à ajuster est une somme de M gaussiennes, i.e.

$$y(x; \mathbf{a}) = \sum_{l=1}^M a_{3l-2} \exp \left[- \left(\frac{x - a_{3l-1}}{a_{3l}} \right)^2 \right].$$

Les paramètres à trouver sont donc a_{3l-2} , a_{3l-1} et a_{3l} pour $l = 1, \dots, M$. Ajuster le modèle pour différentes valeurs de $M \leq 5$. Tracer, pour chaque valeur de M , l'échantillon et la courbe d'ajustement.

Mêmes questions avec l'échantillon

```
x 0.9 1.3 1.9 2.1 2.6 3.0 3.9 4.4 4.7 5.0 6.0 7.0 8.0 9.2 10.5 11.3 11.6 12.0 12.6 13.0 13.3
y 1.3 1.5 1.85 2.1 2.6 2.7 2.4 2.15 2.05 2.1 2.25 2.3 2.25 1.95 1.4 0.9 0.7 0.6 0.5 0.4 0.25
```