

DIXIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

1067



NICE du 20 au 24 MAI 1985

PROGRAMMATION DE L'ALGORITHME DE GAUSS-MARQUARDT POUR LE TRAITEMENT
DE SIGNAUX DISCRETS ISSUS DE SYSTEMES BIOLOGIQUES NON-LINEAIRES.

E. GREMILLET, M. VOUTAY, T. HERRMANN, A. CHAMPAILLER,
M. DECOUSUS, J.C. HEALY.

Laboratoire de Biophysique et d'Informatique Médicale.
30 rue Ferdinand Gambon. 42023 SAINT-ETIENNE Cédex. FRANCE.

RESUME

Plusieurs domaines de la biologie font appel à la modélisation non-linéaire. La réponse à ces problèmes implique de disposer d'un logiciel fiable et souple de régression non-linéaire.

Nous avons mis au point un programme basé sur la méthode de Gauss-Marquardt, sur un ordinateur HP1000.

Après un rappel de l'algorithme de Gauss-Newton, nous précisons les modifications proposées par Marquardt pour limiter les risques de divergence et de "illconditioning".

Les principales options algorithmiques sont ensuite détaillées en particulier l'inversion de matrice par la méthode de Doolittle-Choleski, et l'optimisation itérative de la valeur du scalaire de Marquardt selon un schéma inspiré de Tommasson et Roux qui aboutit à décomposer chaque itération en 2 ou plusieurs sous-itérations et permet théoriquement d'avoir la correction la plus ample non divergente.

Le principal critère d'arrêt des itérations est deux diminutions relatives successives de la somme des carrés des résidus inférieurs à un rapport choisi.

La précision des estimations fournies s'est avérée excellente en comparant les résultats obtenus par fitting d'une polysinoïde de périodes T , $\frac{T}{2}$, $\frac{T}{3}$, et $\frac{T}{4}$ avec ceux obtenus par un programme de TFD sur des données chronobiologiques réelles.

Enfin, les développements actuels du programme sont présentés : passage aux estimations-M "robustes".

SUMMARY

Non-linear models are involved in various fields of biology. Such problems imply having adaptation and reliable non-linear regression program.

Thus, we perfected a program based on Gauss-Newton-Marquardt method, implemented in a HP 1000 computer.

After a recall of Gauss-Newton algorithm, we specify the modifications proposed by Marquardt in order to limit divergence risks and illconditioning problems.

The main algorithmic choices are then specified especially matrix inversion by Doolittle-Choleski's method and iterative optimisation of the value of Marquardt's scalar according to a scheme due to Tommasson et Roux and slightly modified. This scheme splits each iteration into two or more sub-iterations and theoretically allows the greatest non-divergent correction.

The main iteration stopping criteria is : two successive relative decrements of the residual sum of squares lower than a given ratio.

The accuracy of the estimates was found excellent when comparing the results obtained by a polysinoidal fitting (with periods

$T, \frac{T}{2}, \frac{T}{3}, \frac{T}{4}$) with those obtained by a DFT program, about real chronobiologic datas.

Finally current developments of the program are presented : implementation of the M-estimates method, so-called robust estimation.



L'adéquation d'un modèle mathématique analytique à des données expérimentales discrètes est un problème fréquemment rencontré par les biologistes dans des domaines très variés : discrimination de sous-populations cellulaires à partir d'un histogramme de mesures, modélisation de courbes de survie cellulaire, détermination de constantes biochimiques, cinétique de traceurs endogènes ou exogènes, exploitation des courbes d'étalonnage de dosage, ... Les motivations peuvent être différentes :

- attitude explicative : le modèle théorique est connu et on cherche à expliciter ses paramètres dont la signification est explicite ;
- attitude pragmatique : plutôt qu'une modélisation à proprement parler on recherche une "formulation numérique" des données permettant une exploitation rapide et commode des résultats (exemple typique : courbes d'étalonnage) ;
- attitude mixte : plusieurs modèles théoriques sont plausibles et la meilleure adéquation de l'un ou l'autre sera un argument de poids.

L'utilisation de plus en plus fréquente de modèles non-linéaires nous a conduit (1) à mettre au point un programme de régression non-linéaire basé sur la méthode de Gauss-Newton et tenant compte des améliorations de Marquardt.

Matériel : Calculateur Hewlett-Packard série 1000, modèle A700.
Langage PASCAL/1000, c'est à dire norme ANSI avec quelques extensions dont en particulier le type LONGREAL (réel double précision) utilisé dans cette application.

Bibliothèque HP-VIS (Vector Instruction Set) permettant une manipulation aisée et rapide des variables structurées.

Principe :

- Gauss-Newton :

On dispose de n couples (x_i, y_i) où x est une variable de contrôle à variance nulle et y une mesure expérimentale.

Le modèle implique p paramètres θ_j (vecteur θ) qui satisfait :

$$y_i, x_i = f(x_i, \theta) + e_i$$

Les e_i étant les erreurs expérimentales suivant une même loi normale $(0, \sigma^2)$.

A partir d'un vecteur θ^0 de départ le développement de Taylor au 1er ordre permet de calculer le vecteur correctif de Gauss-Newton qui minimise la somme des carrés des écarts

$$RSS = \sum_{i=1}^n w_i e_i^2$$

(où les w_i sont les poids éventuels des points : (x_i, y_i)) (2) :

$$B = [A' \cdot W \cdot A]^{-1} A' W C$$

où A = le jacobien

W = la matrice diagonale des w_i

C = n -vecteur des résidus $(y_i - f(x_i, \theta))$

On batit alors une nouvelle estimation $\theta^1 = B + \theta^0$ et le calcul est itéré jusqu'à un critère d'arrêt.

- Marquardt :

Pour éviter les problèmes de divergence rencontrés avec des fonctions hautement non-linéaires (l'approximation linéaire au 1er ordre n'est alors vérifiée qu'en voisinage immédiat de θ^0) Marquardt a proposé de rajouter un scalaire pontif λ à la dia-

gonale de la matrice à inverser (2,3). Le nouveau vecteur de correction B_m a alors un module d'autant plus petit que λ est grand, mais il n'est plus invariant aux changements d'échelle et il faut le normaliser par $D^{-1/2}$ où D est la matrice diagonale de $A'WA$.

$$B_m = D^{-1/2} [D^{-1/2} A' W A D^{-1/2} + \lambda I]^{-1} D^{-1/2} A' W C$$

Cette modification a en outre l'avantage de limiter les problèmes de "illconditioning" (quasi colinéarité de la matrice à inverser rendant l'inversion impossible) qui peuvent aussi être améliorés par le recours à la double précision.

Options algorithmiques :

- Le programme actuel ne propose pas de pondération et W n'intervient pas dans le calcul.
- La complexité apparente du calcul matriciel à effectuer peut être considérablement simplifiée et l'ordre des opérations à l'itération $n^{\circ} k$ est :
 - . A et C sont calculés à partir de θ^{k-1}
 - . seul le triangle supérieur de $A'A$ est calculé, directement à partir de A .
 - . $D^{-1/2}$ est calculé et stocké dans un n -vecteur pour éviter un double adressage lors des calculs suivants
 - . Triangle supérieur de $D^{-1/2} A' A D^{-1/2} = S$ est calculé
 - . $A'C$ est calculé à partir de A et C
 - . S est ajouté à S
 - . S est inversé par la méthode de Doolittle-Choleski (4,5) :

* $S = T'T$ où T , triangulaire supérieure est calculée ligne par ligne selon :

$$t_{i,i} = \sqrt{s_{i,i} - \sum_{k=1}^{i-1} t_{ki}^2}$$

$$t_{i,j} = \frac{1}{t_{i,i}} (s_{i,j} - \sum_{k=1}^{i-1} t_{ki} t_{kj})$$

* T est inversée $\tau = T^{-1}$ où τ , triangulaire supérieure, est calculée ligne par ligne selon

$$\tau_{i,i} = 1/t_{i,i}$$

$$\tau_{i,j} = \left[- \sum_{k=i+1}^n \tau_{ik} t_{kj} \right] / t_{ij}$$

* enfin, le triangle supérieur de $R = S^{-1}$ est obtenu par $\tau \tau'$, calcul qui n'utilise que τ .

. Le triangle supérieur de $D^{-1/2} R D^{-1/2}$ est calculé et enfin le résultat final B_m puis θ^k et RSS^k .

- A chaque itération la valeur de λ est adaptée selon un schéma emprunté à Tomassone et Roux (3) et un peu modifié :

. RSS^k est comparé à la valeur précédente RSS^{k-1} dans tous les cas (à moins que RSS^k soit déjà inférieur au seuil choisi) l'itération est reprise à partir de l'addition de λ en utilisant :

* soit λ/ν ($\nu > 1$, arbitraire fixe) si

$RSS^k < RSS^{k-1}$ pour voir si une correction plus ample peut être obtenue sans divergence ; si c'est le cas c'est cette 2ème valeur de B_m^k qui est conservée et λ/ν devient le λ courant pour la $(k+1)$ ème itération ; si la 2ème valeur de RSS^k (obtenue avec λ/ν) est supérieure à RSS^{k-1} , on conserve le 1er résultat B_m^k et λ est inchangé pour la $(k+1)$ ème itération.

PROGRAMMATION DE L'ALGORITHME DE GAUSS-MARQUARDT POUR LE TRAITEMENT DE SIGNAUX DISCRETS ISSUS DE SYSTEMES BIOLOGIQUES NON-LINEAIRES.

* soit λ, ν si $RSS^k > RSS^{k-1}$; si le nouveau RSS^k reste supérieure à RSS^{k-1} on recommence avec λ, ν^2 etc. jusqu'à obtenir $RSS^k < RSS^{k-1}$; la première valeur λ, ν^m satisfaisante devient le λ courant pour la (k+1)ème itération et le résultat correspondant B_m^k est retenu. Cependant si au cours de cette démarche λ, ν^m dépasse une valeur seuil arbitraire (10³⁰) sans que la condition $RSS^k < RSS^{k-1}$ soit satisfaite, on sort de l'itération et le résultat de l'itération précédente θ^{k-1} est retenu comme estimation finale.

. chaque itération comprend en fait 2 ou plusieurs sous-itérations.

. cette optimatisation itérative de λ permet de réduire au maximum les risques de divergence tout en limitant le nombre total d'itérations.

- L'estimation finale est obtenue dès qu'un des critères suivants est atteint :

. $RSS <$ seuil choisi par l'utilisateur mais souvent difficile à évaluer.

. 10p itérations : critère en pratique jamais atteint mais seulement implanté par sécurité.

. 2 diminutions relatives successives de RSS inférieurs à un rapport choisi par l'utilisateur ; ce critère est apparu le plus intéressant en pratique.

. utilisation vaine d'un $\lambda, \nu^m > 10^{30}$ lors d'une sous-itération (cf. plus haut).

- Actuellement, l'utilisateur peut choisir ν et la valeur initiale de λ sinon $\nu=2$ et λ initial = 4.

- Un catalogue de modèles est actuellement implanté mais remis à jour au fil des besoins ; (pour l'instant toutes les fonctions exploitées ont des dérivées partielles analytiquement pas trop compliqués et il n'y a pas été nécessaire de les estimer par différences finies).

- En fin de programme sont calculés :

. RSS final

. critère d'information d'Akaïké :

$$AIC = n \cdot \ln RSS + 2p$$

. la matrice finale A'A = M et son inverse V

- A chaque estimation $\hat{\theta}_j$ est associé

$$m_{jj} = \sum_{i=1}^n \left(\frac{\delta f(x_i, \hat{\theta})}{\delta \theta_j} \right)^2$$

. sa variance = $\frac{RSS}{n-p} \cdot v_{jj}$; il faut rappeler

que ce calcul n'a de signification que si n est grand et RSS très petit, c'est à dire que dans les cas courants il est illusoire de donner un intervalle de confiance aux $\hat{\theta}_j$.

- La précision du programme a été testée par l'implantation d'un modèle polysinusoïdale de périodes T, $\frac{T}{2}$, $\frac{T}{3}$, $\frac{T}{4}$; les résultats obtenus sur des données réelles (chronobiologie avec T = 24 h) sont absolument identiques à ceux obtenus par un programme de TFD utilisant l'algorithme de Goétzel.

Perspectives :

Comme nous l'avons dit, le programme n'inclue pas de pondération des points. Actuellement est développée une technique de pondération itérative consistant à affecter chaque point d'un poids w_i à chaque itération.

A l'itération n^e k on calcule d'habitude les résidus

$$y_i - f(x_i, \theta^{k-1}) \text{ et on forme } u_i = \frac{y_i - f(x_i, \theta^{k-1})}{\kappa \cdot s(\theta^{k-1})}$$

avec :

. $s(\theta^{k-1})$: fonction positive de θ , estimant l'échelle des résidus avec θ^{k-1} ; il a été proposé de prendre la médiane des

$$|y_i - f(x_i, \theta^{k-1})|$$

. κ : réel positif arbitraire choisi tel qu'une fraction donnée (par exemple 10%) des points soient tels que :

$$|y_i - f(x_i, \theta^{k-1})| > \kappa \cdot s(\theta^{k-1})$$

Puis chaque point est affecté d'un poids :

$$w_i = \frac{\Psi(u_i)}{u_i}$$

où Ψ est une fonction réelle inspirée en général des fonctions de :

. Huber : $\Psi(u) = u$ si $|u| \leq 1$
 $\Psi(u) = \text{sgn}(u)$ si $|u| > 1$

. Tukey : $\Psi(u) = u(1-u^2)^2$ si $|u| \leq 1$
 $\Psi(u) = 0$ si $|u| > 1$

Ces fonctions sont telles que les points satisfaisant

$$|y_i - f(x_i, \theta^{k-1})| < \kappa \cdot s(\theta^{k-1}) \text{ ont un poids } = 1$$

ou proche de 1 mais que les points les plus "écartés" du modèle (avec θ^{k-1}) sont affectés de poids = 0 ou tendant vers 0.

Cette technique appelée estimation-M(2) est dite estimation "robuste" car elle est supposée insensible aux points aberrants et aux distributions non-normales d'erreurs. Cette méthode prometteuse réclame encore beaucoup de mise au point pour optimiser les choix de κ , $s(\theta)$ et Ψ en fonction des applications.

Références :

1 : GREMILLET E. : Logiciel de régression non-linéaire selon la méthode de Gauss-Marquardt. Mémoire pour l'Attestation d'Etudes Relatives aux Applications à la Biologie Médicale des Radioéléments Artificiels. Lyon I, 1984, LBIM UER de Médecine de St Etienne, 46 p, photocopié.
 2 : JENNRICH R.I., RALSTON M.L. : Fitting non linear models to data. Ann. Rev. Biophys. Bioeng., 1979, 8, 195-238.
 3 : VILA J.P. : Méthodes d'identification des modèles dynamiques. pp 171-195 in Lebreton J.D., Millier C. : Modèles dynamiques déterministes en biologie. Masson, Paris, 1982, 204p.
 4 : ANGOT A. : Compléments de mathématiques à l'usage des ingénieurs de l'électrotechnique et des télécommunications. Masson, Paris, 6ème édition, 1972, 869 p.
 5 : NAKACHE J.P., CHEVALIER A., MORICE V. : Exercices commentés de mathématiques pour l'analyse statistique des données. Bordas, Paris, 1981, 212 p.

