Observabilité et inversibilité de la matrice d'information de Fisher dans les problèmes de régression non-linéaire

Claude Jauffret

Université de Toulon et du Var
MS/GESSY
BP 132,- 83 957 LA GARDE CEDEX
Email : jauffret@isitv.univ-tln.fr

Résumé – On établit le lien entre la notion déterministe d'observabilité et la matrice d'information de Fisher. Plus précisemment, on montre que si cette dernière est inversible alors le paramètre concerné est observable localement, quelle que soit la loi du bruit additif dans le modèle non lineaire.

Abstract – The link between the determistic observabilty and the Fisher Information Matrix is established. More precisely, we prove taht if that matrix is non singular, then the parameter of consideration is locally observable, whatever the statistical law of noise be, in the nonlinear model.

1. Formulation du problème

Considérons le problème de la régression non-linéaire qui consiste à estimer un paramètre déterministe inconnu θ (élément de $\mathbb{R}^d$) lorsque l'on dispose d'un vecteur de mesure X (élément de $\mathbb{R}^n$), problème défini par l'équation de mesure

$$X = h(\theta) + \varepsilon \qquad (\Sigma_\varepsilon).$$

Le vecteur ε représente le bruit additif de mesure ; le modèle liant les mesures à l'état θ est donné sous la forme de la fonction déterministe h.

La tentative d'estimation du paramètre θ ne peut être lancée que si auparavant on a su répondre à la question de l'observabilité de ce paramètre à partir de mesures non-bruitées ; plus précisemment, si l'on dispose de M, vecteur de mesures non-bruitées, et du modèle déterministe (ou fonction) h tels que

$$M = h(\theta) \qquad (\Sigma),$$

la question de l'observabilité de θ se pose en termes d'injectivité de h. D'où une première définition :

<u>*Definition 1*</u>

Le paramètre θ est (simplement) observable en θ_0 à partir de M si

$$\forall \theta' \in \mathbb{R}^d \quad (\theta' \neq \theta_0) \Rightarrow h(\theta') \neq h(\theta_0).$$

Par extension, on définit aussi la notion d'observabilité simple en tout point :

<u>*Definition 2*</u>

Le paramètre θ est (simplement) observable à partir de M si

$$\forall \theta, \forall \theta' \in \mathbb{R}^d \quad (\theta \neq \theta') \Rightarrow h(\theta) \neq h(\theta')$$

Remarque :
Une condition nécessaire évidente à l'observabilité simple est que $d \le n$; cette condition est désormais supposée remplie dans le reste du papier.

Il est licite de poser la définition en termes d'injectivité locale de la fonction h. Dans ce cas, on aboutit à une autre définition :

<u>*Definition 3*</u>

Le paramètre θ est dit localement faiblement observable en θ_0 si

$$\exists U_{\theta_0} \subset \mathbb{R}^d (\textit{ouvert contenant } \theta_0), \forall \theta' \in U_{\theta_0}.$$
$$\theta' \neq \theta_0 \Rightarrow h(\theta') \neq h(\theta_0)$$

Remarques :

a) Ces définitions sont issues de la caractérisation des systèmes dynamiques pour lesquels il existe d'autres notions d'observabilité lorsque le paramètre θ varie dans le temps et qu'il faut tenir compte de toute la trajectoire $\theta(t)$ [7].

b) Lorsque h est une fonction linéaire (en pratique une matrice), il y a confusion des deux définitions :

(i) θ est simplement observable observable en θ_0.

(ii) θ est localement faiblement observable en θ_0.

La plupart du temps, l'analyse de l'observabilité (simple ou locale) est une tâche délicate et il est dit que l'observabilité est garantie dès l'instant où la matrice d'information de Fisher, calculée sous l'hypothèse gaussienne en ce qui concerne ε, est inversible [10]. Nous allons montrer que cette assertion est fausse.

2. Matrice d'information de Fisher (MIF)

Le vecteur ε est désormais supposé admettre une densité de probabilité p_ε de support $\mathbb{R}^n$. Ainsi le vecteur X admet lui même une densité de probabilité p_X qui, bien que paramétrée par θ, a un support ne dépendant pas de θ. Cette dernière est donc définie par

$$\begin{aligned} p_X &: \mathbb{R}^n \to \mathbb{R}^+ \\ &\xi \mapsto p_X(\xi) = p_\varepsilon(\xi - h(\theta)) \end{aligned}$$

La vraisemblance de l'échantillon X notée $L(X;\theta)$ n'est donc rien d'autre que la composition de la fonction (certaine) p_X avec le vecteur aléatoire *X*.

D'un point de vue analytique, on suppose que p_ε et h sont partout différentiables sur $\mathbb{R}^n$ et $\mathbb{R}^d$, respectivement. Nous définissons l'opérateur Jacobien $\frac{\partial}{\partial\theta}$ par

$$\frac{\partial h(\theta)}{\partial\theta} \triangleq \begin{bmatrix} \frac{\partial h_1(\theta)}{\partial\theta_1} & \cdots & \frac{\partial h_1(\theta)}{\partial\theta_d} \\ \vdots & & \vdots \\ \frac{\partial h_n(\theta)}{\partial\theta_1} & \cdots & \frac{\partial h_n(\theta)}{\partial\theta_d} \end{bmatrix}.$$

La matrice d'information de Fisher relative au paramètre θ à estimer, évaluée au point θ_0, se définit comme la matrice de variance-covariance du Jacobien de la fonction log-vraisemblance, i.e.

$$F_X(\theta_0) \triangleq Cov_{\theta_0}\left\{\left[\frac{\partial Ln[L(X;\theta)]}{\partial\theta}\right]^T_{\theta=\theta_0}\right\} = E_{\theta_0}\left\{\left[\frac{\partial Ln[L(X;\theta)]}{\partial\theta}\right]^T_{\theta=\theta_0}\left[\frac{\partial Ln[L(X;\theta)]}{\partial\theta}\right]_{\theta=\theta_0}\right\}.$$

Remarquons d'ailleurs que si ε est un vecteur Gaussien centré de matrice de variance-covariance R_ε (supposée non-singulière), alors

$$F_X(\theta_0) = \left[\frac{\partial h(\theta)}{\partial\theta}\right]^T_{\theta=\theta_0} R_\varepsilon^{-1}\left[\frac{\partial h(\theta)}{\partial\theta}\right]_{\theta=\theta_0}$$

Nous nous proposons de donner une expression analogue pour des densités de probabilité de ε quelconques (pourvu que leurs supports soient $\mathbb{R}^n$).

Nous établissons dans un premier temps le résultat suivant :

Théorème 1

$$F_X(\theta_0) = \left[\frac{\partial h(\theta)}{\partial\theta}\right]^T_{\theta=\theta_0} W_\varepsilon\left[\frac{\partial h(\theta)}{\partial\theta}\right]_{\theta=\theta_0}$$

$$\text{où } W_\varepsilon \triangleq Cov\left\{\left[\frac{\partial \ln[p_\varepsilon(t)]}{\partial t}\right]^T_{t=\varepsilon}\right\}.$$

Démonstration :

D'après la définition de la densité de probabilité de X, on a

$$\frac{\partial Ln[p_X(z;\theta)]}{\partial\theta} = \frac{\partial Ln[p_\varepsilon(z-h(\theta))]}{\partial\theta} = \left[\frac{\partial Ln[p_\varepsilon(v)]}{\partial v}\right]_{v=z-h(\theta)}\frac{\partial h(\theta)}{\partial\theta}.$$

La dérivée de la fonction log-vraisemblance relativement au paramètre θ est donc

$$\frac{\partial Ln[L(X;\theta)]}{\partial\theta} = \left[\frac{\partial Ln[p_\varepsilon(v)]}{\partial v}\right]_{v=X-h(\theta)}\frac{\partial h(\theta)}{\partial\theta}.$$

Donc

$$F_X(\theta_0) = \left[\left[\frac{\partial h}{\partial\theta}\right]^T E_\theta\left\{\left[\frac{\partial Ln[p_\varepsilon(v)]}{\partial v}\right]^T\left[\frac{\partial Ln[p_\varepsilon(v)]}{\partial v}\right]_{v=X-h(\theta)}\right\}\left[\frac{\partial h}{\partial\theta}\right]\right]_{\theta=\theta_0}.$$

Quant au terme médian,

$$E_{\theta_0}\left\{\left[\frac{\partial Ln[p_\varepsilon(v)]}{\partial v}\right]^T\left[\frac{\partial Ln[p_\varepsilon(v)]}{\partial v}\right]\right\} = \int_{\mathbb{R}^n}\left[\left[\frac{\partial Ln[p_\varepsilon(v)]}{\partial v}\right]^T\left[\frac{\partial Ln[p_\varepsilon(v)]}{\partial v}\right]\frac{1}{p_\varepsilon(v)}\right]_{v=z-h(\theta_0)} dz.$$

Le changement de variable $t = z - h(\theta_0)$ conduit à l'expression annoncée :

$$\int_{\mathbb{R}^n}\left[\left[\frac{\partial Ln[p_\varepsilon(v)]}{\partial v}\right]^T\left[\frac{\partial Ln[p_\varepsilon(v)]}{\partial v}\right]\frac{1}{p_\varepsilon(v)}\right]_{v=z-h(\theta_0)} dz$$

$$= \int_{\mathbb{R}^n}\left[\frac{\partial Ln[p_\varepsilon(t)]}{\partial t}\right]^T\left[\frac{\partial Ln[p_\varepsilon(t)]}{\partial t}\right]\frac{1}{p_\varepsilon(t)}dt = Cov\left\{\left[\frac{\partial Ln[p_\varepsilon(t)]}{\partial t}\right]^T_{t=\varepsilon}\right\}.$$

Remarque :
Si ε est gaussien, il est facile de vérifier que $W_\varepsilon = R_\varepsilon^{-1}$.

3. Quelques cas pathologiques

a) Premier contre-exemple :
Considérons la mesure monodimensionnelle

$$X = \theta^3 + \varepsilon \quad (\Sigma_\varepsilon)$$

(ici, n=d=1).
L'équation de mesure sans bruit est donc

$$M = \theta^3 \quad (\Sigma)$$

et le paramètre θ est (simplement) observable en tout point.
Sous l'hypothèse gaussienne, la MIF est

$$F_X(\theta) = 9\theta^4$$

(on suppose que ε est centré réduit).
Donc $F_X(\theta)$ n'est pas inversible en $\theta = 0$.
Commentaires:

1) Dans des cas semblables, on peut montrer qu'il n'existe pas d'estimateur sans biais : la borne de Cramèr-Rao ne peut donc pas être l'inverse de la MIF.
2) Ce cas se rencontre en traitement d'antenne lorsqu'il s'agit d'estimer l'angle d'arrivée d'un source située dans l' « end-fire ». Il est d'ailleurs d'usage d'estimer les cosinus des angles (ou fréquences spatiales) pour lesquels l'information de Fisher n'est jamais nulle [11].
3) La singularité de la MIF en certains points de $\mathbb{R}^d$ pose problème lorsque l'on calcule l'estimé (pour des mesures gaussiennes) par la procédure numérique de Gauss-Newton dans laquelle le Hessien est approché par la MIF évaluée au point de l'itération courante. Le palliatif est alors l' «augmentation » de la MIF selon la méthode de Levenberg-Marquardt [12].

b) Second contre-exemple (tiré de [8]) :
Considérons le vecteur de mesure bidimensionnel

$$X \triangleq \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} a\theta - \sin\theta \\ \cos\theta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}, \quad 0 < a < 1.$$

Nous supposons encore que $Cov(\varepsilon) = I_2$
Sous l'hypothèse gaussienne, on a

$$F_X(\theta) = 2a\left(\frac{a+1}{2a} - \cos\theta\right).$$

Donc la MIF n'est jamais nulle et pourtant les paires (θ, θ') définies par

$$\begin{cases} \theta = 2k\pi + \tau \\ \theta' = 2k\pi - \tau \\ avec \quad \tau = \dfrac{\sin\tau}{a} \end{cases}$$

donnent la même mesure sans bruit. Le paramètre θ n'est donc pas simplement observable. Il est toutefois localement faiblement observable.

4. Analyse

Pour mener à bien l'analyse du lien entre l'observabilité (en un certain sens) et l'inversibilité de la matrice d'information de Fisher, on a besoin de quelques outils basiques d'algèbre linéaire et de géométrie différentielle.

<u>*Théorème 2*</u>
Soit A une $(n \times d)$ matrice $(d \le n)$. Sont équivalents :
(i) $A^T A$ est inversible.
(ii) $\forall S$ une $(n \times n)$ matrice réelle symétrique non-singulière, $A^T S A$ est inversible.
(iii) $Rang(A) = d$.

<u>*Définition 2*</u> [9]
Pour $(d \le n)$, $h : \mathbb{R}^d \to \mathbb{R}^n$ est une immersion en θ_0 si le rang de $\left[\dfrac{\partial h(\theta)}{\partial \theta}\right]_{\theta=\theta_0}$ est égal à d.

<u>*Théorème 3*</u> [9]
$h : \mathbb{R}^d \to \mathbb{R}^n$ est une immersion en θ_0 s'il existe un ouvert U_{θ_0} contenant θ_0 tel que le rang de $\left[\dfrac{\partial h}{\partial \theta}\right]_\theta$ est égal à d quel que soit θ dans U_{θ_0}

<u>*Corollaire 1*</u> [9]
Si *$h : \mathbb{R}^d \to \mathbb{R}^n$ est une immersion en θ_0, alors h est localement injective, c'est-à-dire qu'il existe un ouvert U_{θ_0} contenant θ_0 tel que* $h_{U_{\theta_0}} : U_{\theta_0} \subset \mathbb{R}^d \to \mathbb{R}^n$ est une injection.

Remarques :
Evidemment, si *h* est une application linéaire,

$$\left[\frac{\partial h(\theta)}{\partial\theta}\right]_{\theta=\theta_1}=\left[\frac{\partial h(\theta)}{\partial\theta}\right]_{\theta=\theta_2},\forall\theta_1,\theta_2\in\mathbb{R}^d.$$

Donc, h est injective en un point, elle est injective partout.

5. Conclusion

Combinant les arguments précédemment évoqués, nous pouvons conclure par le

Corollaire 2

Pourvu que le support de p_ε soit $\mathbb{R}^n$ et que la matrice $Cov\left\{\left[\frac{\partial\ln[p_\varepsilon(t)]}{\partial t}\right]^T_{t=\varepsilon}\right\}$ soit non-singulière, si la matrice d'information de Fisher est inversible en θ_0 - ou de façon équivalente si $\left[\frac{\partial h(\theta)}{\partial\theta}\right]^T_{\theta=\theta_0}\left[\frac{\partial h(\theta)}{\partial\theta}\right]_{\theta=\theta_0}$ est inversible – alors θ est localement faiblement observable en θ_0.

Références

[1] Nardone, S.C. and Aidala, V.J. (1981)
Observability criteria for bearings-only target motion analysis.
IEEE Transaction on Aerospace and Electronic Systems, AES-17, 2 (Mar. 1981), 162-166.

[2] Payne, A.N. (1988)
Observability conditions for angles-only tracking.
In Processings of 22nd Asilomar Conference on Signals, Systems, and Computers, Oct. 31-Nov. 2, 1988.

[3] Jauffret, C. and Pillon, D.
Observability in passive target motion analysis
IEEE Transaction on Aerospace and Electronic Systems, AES-32, 4 (Oct. 1996), 1290-1300.

[4] Le Cadre, J.P. and Jauffret, C.
Discrete time observability and estimability analysis for bearings only target motion analysis.
IEEE Transaction on Aerospace and Electronic Systems, AES-33, 1 (Jan. 1997), 178-201.

[5] Bar-Shalom, Y. and Li, X.R.
Estimation and tracking : Principles, Techniques ans Software, p 168.
Artech House, Boston London (1993).

[6] Arnold, J.F., Bar-Shalom, Y., Estrada, R. and Mucci, R.A.
Target parameter estimation using measurements acquired with a small number of sensors.
IEEE Journal of Oceanic Engineering, OE-8, 3 (jul. 1983), 163-172.

[7] Hermann, R. and Krener, A.J.
Non-linear controllability and observability.
IEEE Transaction on Automatic Control, AC-22, 5 (oct. 1997), 728-740.

[8] Isidori, A.
Nonlinear Control Systems.
Springer (1989), 479-480.

[9] Schwartz, L.
Calcul différentiel et équations différentielles.
Hermann (1992).

[10] Bar-shalom, Y. et Li, X.R.
Estimation and Tracking : Principles, Techniques and Software
Artech House (1993), 168-169.

[11] Hinich, M.J.
Estimating bearing when the source is endfire to an array.
Jasa, 65 n°3 (1979), 845-846.

[12] Dennis, J.E.
User's Guide to Nonlinear Optimization Algorithms.
Proceedings of IEEE, vol 72, n°12 (déc. 1984), 1765-1776.