

Observability and fisher information matrix in nonlinear regression

Claude Jauffret

▶ **To cite this version:**

Claude Jauffret. Observability and fisher information matrix in nonlinear regression. IEEE Transactions on Aerospace and Electronic Systems, 2007, 43 (2), pp.756-759. 10.1109/TAES.2007.4285368 . hal-01820468

HAL Id: hal-01820468
https://univ-tln.hal.science/hal-01820468

Submitted on 22 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Observability and Fisher Information Matrix in Nonlinear Regression

Claude Jauffret, *Member IEEE*
Université du Sud Toulon-Var
Bâtiment X
BP 20132, 83957 La Garde Cedex, FRANCE
E-mail: jauffret@univ-tln.fr

Abstract

This paper is devoted to the link between the Fisher Information Matrix invertibility and the observability of a parameter to be estimated in a nonlinear regression problem.

Keywords

Observability, identifiability, Cramèr-Rao Lower Bound, Fisher Information Matrix, nonlinear regression.

I. Introduction

In Signal Processing [1], in Target Motion Analysis [2], a large class of measurements can be modelled as

$$X = h(\theta) + \varepsilon \tag{1}$$

where X is the available measurement vector (element of $\mathbb{R}^n$), θ is the unknown deterministic parameter (lying in $\mathbb{R}^d$) and $h(.)$ is a (known) nonlinear mapping from $\mathbb{R}^d$ to $\mathbb{R}^n$. The vector ε represents the additive measurement noise.
Whatever the estimation technique employed (Least Squares, Maximum Likelihood, ...), the observability of parameter θ must be investigated.

Most of the time, the analysis of observability is a tough task, and many authors suggest to declare that θ is observable if the Fisher Information Matrix (FIM) of θ given X, under Gaussian hypothesis concerning ε, is nonsingular at θ (see, for example, [2] page 168). Intuitively (and practically), it turns out that this way is sufficient. But it is legitimate to wonder about the meaning of that analysis: Observability being a deterministic notion, why use a statistical tool (the FIM) to establish its status? Why restrain oneself to the Gaussian law? Would the conclusion be the same if the FIM was computed for another law?
Parts of answers can be found in Jazwinski's book ([3] page 231) but linear Gaussian cases are concerned.
The aim of this paper is to answer these questions by establishing clearly the link between the status of the FIM (singular or nonsingular) under a large class of probability laws and the status of the observability, i.e. in nonlinear and non-Gaussian cases.
The coming section recalls some classical definitions of observability. In the third section,

we give a general form of the FIM. Some pathologic (but still relevant) cases are analyzed in Section IV. The last section gives the main result, after presenting the necessary mathematical tools.

II. Observability Concepts

There are several ways to define the observability concept: it can be a global one (for all the vectors of $\mathbb{R}^d$), or a local one (for a special θ).
We recall the three major definitions of observability.

Definition 1. *The noise-free system $X = h(\theta)$ is simply observable at θ_0 if*

$$\forall \theta \in \mathbb{R}^d, \{\theta \neq \theta_0\} \Rightarrow \{h(\theta) \neq h(\theta_0)\} \tag{2}$$

Definition 2. *The noise-free system is (simply) observable if*

$$\forall \theta, \forall \theta' \in \mathbb{R}^d, \{\theta \neq \theta'\} \Rightarrow \{h(\theta) \neq h(\theta')\} \tag{3}$$

Definition 3. *The noise-free system is locally observable at θ_0 if*

$$\exists U_{\theta_0} \subset \mathbb{R}^d \text{ (open subset containing } \theta_0), \forall \theta \in U_{\theta_0}, \{\theta \neq \theta_0\} \Rightarrow \{h(\theta) \neq h(\theta_0)\} \tag{4}$$

Remark 1.

a) These definitions come from the theory of dynamic systems in which the parameter θ changes in time and must be hence denoted $\theta(t)$. For such systems, there exists some other definitions of observability to take into account the trajectory of $\theta(t)$ [4].

b) When $h(.)$ is a linear mapping (in practice a matrix), local observability and simple observability coincide.

III. Fisher Information Matrix

In the sequel, we assume that the behavior of the vector ε is described by a probability density function (pdf), say p_ε, whose support is $\mathbb{R}^n$. The vector X has its own pdf, denoted p_X. It depends on θ while its support is independent of it [1]. More precisely, the

[1]This assumption is necessary to compute the Fisher Information Matrix.

relationship between them is given by

$$p_X(\nu|\theta) = p_\varepsilon(\nu - h(\theta)). \tag{5}$$

The likelihood function of θ is nothing else than the probability density function of X given θ evaluated at X :

$$L_\theta(X) \triangleq p_X(X|\theta). \tag{6}$$

The Fisher Information Matrix (FIM) is

$$\begin{aligned} \boldsymbol{F}_\theta(X) &\triangleq \boldsymbol{Cov}_\theta\{\nabla_\theta \ln[L_\theta(X)]\} \\ &= \boldsymbol{E}_\theta\left\{\nabla_\theta \ln[L_\theta(X)]\nabla_\theta^T \ln[L_\theta(X)]\right\}. \end{aligned} \tag{7}$$

The Cramèr-Rao Lower Bound (CRLB) of any unbiased estimator of θ is the inverse of $\boldsymbol{F}_\theta(X)$.

If ε is a 0-mean Gaussian vector whose covariance matrix is R_ε (assumed invertible), the FIM can be expressed as follows

$$\boldsymbol{F}_\theta(X) = \nabla_\theta h(\theta) R_\varepsilon^{-1} \nabla_\theta^T h(\theta). \tag{8}$$

Under more general assumptions, we can still give a close form for the FIM .

Theorem 1.

$$\begin{aligned} \boldsymbol{F}_\theta(X) &= \nabla_\theta h(\theta) W_\varepsilon \nabla_\theta^T h(\theta) \\ \textit{where } W_\varepsilon &\triangleq \boldsymbol{E}\{\nabla_\nu \ln[p_\varepsilon(\nu)]_{\nu=\varepsilon} \nabla_\nu^T \ln[p_\varepsilon(\nu)]_{\nu=\varepsilon}\} \end{aligned} \tag{9}$$

Proof :

For the dummy variable z, we have

$$\begin{aligned} \nabla_\theta \ln[p_X(z|\theta)] &= \nabla_\theta \ln[p_\varepsilon(z - h(\theta))] \\ &= \nabla_\theta h(\theta) \nabla_\nu \ln[p_\varepsilon(\nu)]_{\nu=z-h(\theta)}. \end{aligned} \tag{10}$$

Hence

$$\nabla_\theta \ln[L_\theta(X)] = \nabla_\theta h(\theta) \nabla_\nu \ln[p_\varepsilon(\nu)]_{\nu=X-h(\theta)}. \tag{11}$$

As a consequence, the FIM is readily written as

$$\boldsymbol{F}_\theta(X) = \nabla_\theta h(\theta) \boldsymbol{E}_\theta\{\nabla_\nu \ln[p_\varepsilon(\nu)]_{\nu=X-h(\theta)} \nabla_\nu^T \ln[p_\varepsilon(\nu)]_{\nu=X-h(\theta)}\} \nabla_\theta^T h(\theta) \tag{12}$$

The middle term above is

$$\boldsymbol{E}_\theta\{\nabla_\nu \ln\left[p_\varepsilon(\nu)\right]_{\nu=X-h(\theta)} \nabla_\nu^T \ln\left[p_\varepsilon(\nu)\right]_{\nu=X-h(\theta)}\} = \boldsymbol{E}\{\nabla_\nu \ln\left[p_\varepsilon(\nu)\right]_{\nu=\varepsilon} \nabla_\nu^T \ln\left[p_\varepsilon(\nu)\right]_{\nu=\varepsilon}\}. \tag{13}$$

□

In the sequel, W_ε is assumed nonsingular. The following theorem gives us a sufficient condition for that.

Theorem 2. *If $\nabla_\nu p_\varepsilon(\nu)$ is a continuous function then W_ε is nonsingular.*

Proof :

Suppose that W_ε is singular. So there exists a non null vector $u \in \mathbb{R}^d$ such that $u^T W_\varepsilon u = 0$,

$$\begin{aligned} &\text{i.e. } \exists u \neq 0 \text{ s.t. } u^T W_\varepsilon u = 0 \\ &\Leftrightarrow \exists u \neq 0 \text{ s.t. } u^T \left[\int_{\mathbb{R}^d} \nabla_\nu p_\varepsilon(\nu) \nabla_\nu^T p_\varepsilon(\nu) \frac{1}{p_\varepsilon(\nu)} d\nu\right] u = 0 \\ &\Leftrightarrow \exists u \neq 0 \text{ s.t. } \int_{\mathbb{R}^d} \left[\frac{u^T \nabla_\nu p_\varepsilon(\nu)}{\sqrt{p_\varepsilon(\nu)}}\right]^2 d\nu = 0 \end{aligned} \tag{14}$$

Since $\nabla_\theta p_\varepsilon(\theta)$ is a continuous function, this last statement is equivalent to

$$\begin{aligned} &\exists u \neq 0 \text{ s.t. } u^T \nabla_\nu p_\varepsilon(\nu) = 0, \forall \nu \in \mathbb{R}^d \\ &\Leftrightarrow \bigcap_{\nu \in \mathbb{R}^d} \ker\{\nabla_\nu p_\varepsilon(\nu) \nabla_\nu^T p_\varepsilon(\nu)\} \neq \{\vec{0}\}. \end{aligned} \tag{15}$$

Let r be the dimension of the vector space $\bigcap_{\nu \in \mathbb{R}^d} \ker\{\nabla_\nu p_\varepsilon(\nu) \nabla_\nu^T p_\varepsilon(\nu)\}$. In a suitable basis, the last $(d-r)$ components of $\nabla_\nu p_\varepsilon(\nu)$ will be null, i.e.

$$\nabla_\nu p_\varepsilon(\nu) = \left[\frac{\partial p_\varepsilon(\nu)}{\partial \nu_1}, \frac{\partial p_\varepsilon(\nu)}{\partial \nu_2}, \cdots, \frac{\partial p_\varepsilon(\nu)}{\partial \nu_r}, 0, \cdots, 0\right]^T \forall \nu \in \mathbb{R}^d \tag{16}$$

which means that in that basis, $p_\varepsilon(\nu) = p_\varepsilon(\nu_1, \nu_2, ..., \nu_r)$. This contradicts the fact that $\int_{\mathbb{R}^d} p_\varepsilon(\nu) d\nu = 1$. □

Remark 2.

If ε is Gaussian of covariance matrix R, it is readily shown that $W_\varepsilon = R_\varepsilon^{-1}$.

IV. Pathologic cases

A. Case I :

We consider the two-dimensional measurement ($n = d = 2$)

$$X \triangleq \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} \theta_1^3 + \theta_2 \\ \theta_1^3 - \theta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}. \tag{17}$$

Obviously, the associated noise-free system is simply observable at any $\theta = (\theta_1, \theta_2)^T$. Under the assumption that $\varepsilon_i \propto G(0,1)$ and independent, the FIM is

$$\boldsymbol{F}_\theta(X) = \begin{bmatrix} 9\theta_1^4 & 0 \\ 0 & 2 \end{bmatrix} \tag{18}$$

which is singular at any θ such that $\theta_1 = 0$.

Remark 3.

a) In similar cases, we can prove that no unbiased estimator of θ exists since for such estimators the CRLB is the inverse of the FIM. This fact is met in array processing for the estimation of the end-fire bearing [5].

b) The singularity of the FIM at some points of $\mathbb{R}^d$ can cause some problems during the Gauss Newton routine for which the Hessian is approximated by the FIM evaluated at the point of the current iteration. The palliative is the augmentation of the FIM by some αId as suggested in the Levenberg-Marquardt method [6].

B. Case II

This counter-example comes from [7] p. 479. Let's consider the two-dimensional measurement vector

$$X \triangleq \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} a\theta - \sin\theta \\ cos\theta \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}, \text{ with } a \in]0,1[\tag{19}$$

$$\boldsymbol{F}_\theta(X) = 2a\left(\frac{a+1}{2a} - \cos\theta\right) \tag{20}$$

Hence, the FIM is never equal to 0, but still the pairs (θ_1, θ_2) defined by

$$\begin{cases} \theta_1 \triangleq 2k\pi + \tau \\ \theta_2 \triangleq 2k\pi - \tau \end{cases} \tag{21}$$

are undistinguishable, τ being the root of the equation $\tau = \dfrac{\sin\tau}{a}$. The parameter θ is not simply observable, but locally observable.

V. Analysis

A. Mathematical Tools

We need two types of tools : one from the linear algebra theory and the second one from the differential calculus.

Theorem 3. *Let A a $(n \times d)$ matrix $(d \leq n)$. The following statements are equivalent :*
(i) $A^T A$ is invertible.
(ii) $\exists S$ a real nonsingular symmetric $(n \times n)$ matrix such that $A^T S A$ is nonsingular.
(iii) $\forall S$ real nonsingular symmetric $(n \times n)$ matrix, $A^T S A$ is nonsingular.
(iv) $Rank(A) = d$.

Definition 4. *$h : \mathbb{R}^d \to \mathbb{R}^n$ is an immersion at θ_0 if the rank of $\nabla_\theta h(\theta_0)$ is equal to d (see [7] p. 479).*

Theorem 4. *$h : \mathbb{R}^d \to \mathbb{R}^n$ is an immersion at θ_0 if there exists an open set U_{θ_0} containing θ_0 such that the rank of $\nabla_\theta h(\theta)$ is equal to d, whatever θ in U_{θ_0}.*
See [8] for the proof.

Proposition 1. *If $h : \mathbb{R}^d \to \mathbb{R}^n$ is an immersion, than h is locally injective, i.e. it exists an open set U_{θ_0} of $\mathbb{R}^d$ containing θ_0 such that $h : U_{\theta_0} \subset \mathbb{R}^d \to \mathbb{R}^n$ is injective.*
See [8] for the proof.

Remark 4. *If h is a linear mapping i.e. $h(\theta) = A(\theta)$, then $\nabla_\theta h(\theta_1) = \nabla_\theta h(\theta_2) = A^T$ for any pair (θ_1, θ_2). Hence h is locally injective as well as injective anywhere.*

B. The main result

The exploitation of the previous theorems and of the last proposition yields straightforwardly the following

Theorem 5. *Let us consider the measurement equation $X = h(\theta) + \varepsilon$ associated to the noise-free system $X = h(\theta)$. If the support of p_ε is $\mathbb{R}^n$ itself and if the FIM is nonsingular at θ_0 - or, equivalently, if $\nabla_\theta h(\theta_0) \nabla_\theta^T h(\theta_0)$ is nonsingular - then the noise-free system is locally observable at θ_0.*

Proof :

We know that $\nabla_\theta h(\theta) W_\varepsilon \nabla_\theta^T h(\theta)$ is nonsingular. Using Theorem 4, statement (iv), with $A \equiv \nabla_\theta h(\theta_0)$ and $S \equiv W$, we conclude that $Rank(\nabla_\theta h(\theta_0))$ is equal to d ; hence h is an immersion.

Now, thanks to Proposition 1, we know that h is locally injective, and as a consequence, satisfies Definition 3. □

Remark 5.

a) The first pathological case forbids the converse.

b) Remark 4 proves that local observability is equivalent to simple observability for linear system : if h is linear, i.e. $h(\theta_0) = A\theta_0$, then $\nabla_\theta h(\theta_0) \nabla_\theta^T h(\theta_0) = A^T A, \forall \theta_0$. In that case, the system is observable iff $A^T A$ is nonsingular, or equivalently, iff $A^T W_\varepsilon A$ (the FIM) is nonsingular.

VI. Conclusion

The link between the invertibility of the FIM and the observability status has been unambiguously established, for a large class of probability laws in nonlinear regression problems. This theoretical result can help the study of observability.

References

[1] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice Hall, Signal Processing series, 1993.

[2] Y. Bar-Shalom and X. Rong Li, *Estimation and Tracking: Principles, Techniques and Software.* Artech House, Boston, London, 1993.

[3] A. Jazwinski, *Stochastic Processes and Filtering Theory.* Academic Press, 1970.

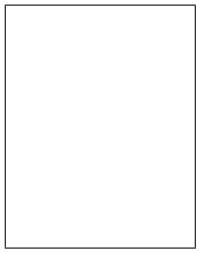
[4] R. Hermann and A. Krener, "Nonlinear controllability and observability," *IEEE Trans. on Automatic Control*, vol. AC-22,5, pp. 728–740, October 1977.

[5] S. Pasupathy and W. Alford, "Range and bearing estimation in passive sonar," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 16, NO 2, pp. 244–249, March 1980.

[6] J. Dennis, J.E., "A user's guide to nonlinear optimization algorithms," *Proceedings of IEEE*, vol. 72, NO 12, pp. 1765–1776, December 1984.

[7] A. Isidori, *Nonlinear Controllability Systems*. Springer, 1989.

[8] L. Schwarz, *Calcul Différentiel et Equations Différentielles*. Hermann, 1992.



Claude Jauffret, born in France on March 29 1957, received the Diplôme d'Etude Approfondies in applied Mathematics from the St Charles University (Marseille, France) in 1981, the Diplôme d'Ingénieur from Ecole Nationale Supérieure d'Informatique et de Mathématiques Appliquées de Grenoble (Grenoble, France) in 1983, the title of "Docteur de l'Université" in 1993 and the "Habilitation à Diriger des Recherches" from the Université de Toulon et du Var (France). From Nov. 1983 to nov. 1988, he worked on passive sonar systems, more precisely on Target Motion Analysis at the GERDSM (France). After a sabbatical year at the University of Connecticut (from nov. 1988 to dec. 1989) during which he worked on tracking problems in cluttered environment, he developped researches in tracking, data fusion and extraction in CERDSM (France). Since sept. 1996, he has been at the Université de Toulon et du Var where he teaches Statistical Signal Processing and Dynamic Linear Systems. His currents researches are about observability, estimation in non-linear systems as they appear in tracking problems