Connecting informative experiments, the information matrix and the minima of a Prediction Error Identification criterion

M. Gevers * A.S. Bazanella ** X. Bombois ***

* CESAME, Université catholique de Louvain, Louvain-la-Neuve,
Belgium. Michel.Gevers@uclouvain.be
** Electrical Engineering Department, Universidade Federal do Rio Grande do Sul, Porto Alegre-RS, Brazil. bazanela@ece.ufrgs.br
*** Delft Center for Systems and Control, Delft University of Technology, The Netherlands. x.j.a.bombois@tudelft.nl

Abstract: This paper establishes, in a Prediction Error Identification (PEI) context, the connections that exist between the identifiability of the model structure, the informativity of the data, the information matrix and the existence of a unique global minimum of the PEI criterion. By introducing the concept of informative data at a particular parameter value, we are able to establish a number of equivalences and connections between these four ingredients of the identification problem, for both open-loop and closed-loop identification.

Keywords: Prediction error methods, identifiability, informative data, information analysis

1. INTRODUCTION

It is well known that the existence of a unique global minimum of a Prediction Error Identification criterion requires the combination of two ingredients: the identifiability of the model structure, and the informativity of the data. The first involves the parametrization only, without any assumption on the true system or on the data, while the second involves both model structure and data. The information matrix combines information about the model structure and the informativity of the data. Given its ease of computation, its significance as the inverse of the parameter covariance matrix, and its use as a key tool for the formulation and solution of experiment design problems, it is therefore of interest to investigate in what way (if any) the positivity of the information matrix can be used as a substitute for the combination of identifiability of the model structure and informativity of the data set. This is the main topic of our paper.

In the context of time-series analysis studied by econometricians, the positivity of the information matrix has been the object of intense interest for a long time, and it was shown that it implies identifiability: Rothenberg [1971], Bowden [1973]. In the presence of measured inputs, however, this is no longer the case. Thus, the concept of informative data was introduced, which together with identifiability implies the existence of a unique global minimum of the identification criterion when the system is in the model set. We show in this paper that the traditional concept of informative data sets is, however, unnecessarily strong. Our attempts to prove equivalence between positive information matrix and identifiability plus informativity were therefore doomed to fail.

In this paper we establish a range of connections and some equivalences between four PEI properties: identifiability of the model structure, informativity of the data, full rank information matrix, and the existence of a unique global minimum of the PEI criterion. We establish these connections for a family of model structures that is larger than - and that encompasses - all traditional linear model structures used in system identification, for both open-loop and closed-loop identification, for the case where the system is in the model set as well as for the case where it is not.

The key to establishing these connections is the introduction of the new concept of informativity at a given parameter value. With this new definition, we establish the equivalence between positive information matrix at some parameter \( \theta_0 \) and identifiability plus informativity at that value \( \theta_0 \). This equivalence holds whether or not the system is in the model set. Armed with these results we then establish that, when the system is in the model set, the positivity of the information matrix at the true value \( \theta_0 \) is a sufficient condition for the existence of a unique minimum of the criterion provided some affinity condition holds, whereas the converse requires additionally that the model structure is identifiable at \( \theta_0 \). When the system is not in the model set, we establish that, for open-loop identification, informativity of the data at the minimum of the criterion is a necessary condition for the existence of a unique global minimum of the criterion. Further connections are under investigation.

The paper is organized as follows. We formulate the problem and establish the notations in Section 2. In Section 3 we present the new definitions of identifiability.
and informative data at a given parameter value. Section 4 establishes conditions under which the set of minimizers of a PEI criterion, in open and in closed loop, is affine in the parameters. Our main results are presented in Section 5 where we establish connections between the four properties mentioned above.

2. PROBLEM FORMULATION

We consider the Prediction Error Identification (PEI) of a linear time-invariant discrete-time single-input single-output “real system”:

\[ S: \quad y(t) = G_0(z)u(t) + H_0(z)e(t) \]  

where \( G_0(z) \) and \( H_0(z) \) are the process transfer functions, \( u(t) \) is the input and \( e(t) \) is white noise with variance \( \sigma^2_e \). Both transfer functions are rational and proper; furthermore, \( H(z) \) is monic, i.e. \( H_0(\infty) = 1 \). To be precise, we shall define \( S \triangleq \{ G_0(z) \cdot H_0(z) \} \). This system may or may not be under feedback control with a proper rational stabilizing controller \( K(z) \):

\[ u(t) = K(z)[r(t) - y(t)]. \]

The signals \( u(t) \) and \( r(t) \) are assumed to be quasistationary [Ljung 1999]. When the data are generated in open loop, we assume that \( E[u(t)e(s)] = 0 \) \( \forall s \); when they are generated in closed loop, we assume that \( E[r(t)e(s)] = 0 \) \( \forall s \). Here \( E[\cdot] \) is defined as

\[ E[f(t)] \triangleq \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{\infty} E[f(t)] \]

with \( E[\cdot] \) denoting expectation: [Ljung 1999].

For the purpose of identifying a model for (1) we consider in this paper model structures of the following form

\[ G(z, \theta) = n^g(z, \theta) = n_0 g_1(z^{-1}) + \theta^T n_1(z^{-1}) \]  
\[ H(z, \theta) = \frac{n^h(z, \theta)}{d^h(z, \theta)} = n_0 h_1(z^{-1}) + \theta^T n_1(z^{-1}) \]

where \( \theta \in D_\theta \subseteq \mathbb{R}^d \); here \( n_0, d_0, n_0, d_0, h_0, d_1, h_1, \) are vectors of polynomials in \( z^{-1} \). They satisfy the constraints \( n_1(\infty) = 0, d_1(\infty) = 0 \), and \( n_0(\infty) = 1, d_0(\infty) = 1 \), so that also \( H(\infty, \theta) = 1 \). This family of model structures encompasses all “classical” model structures (ARMAX, ARX, BJ, OE) but it is much larger. For example, it allows one to consider the classical model structures in which some polynomial coefficients are known, or are zero. As an example, the following generalization of an ARMAX model can be described by the model structure (3)-(4):

\[ y(t) + \theta_1 y(t - 2) - y(t - 3) + \theta_2 y(t - 5) = 2u(t - 1) + \theta_3 u(t - 4) + e(t) - (2\theta_2 + \theta_3)e(t - 4) \]

The classical model structures are obtained from (3)-(4) by setting \( n_0 = 0 \), \( n_0 = d_0 = d_0 = 1 \) and the vectors as zeros and powers of \( z^{-1} \). Thus, for an ARMAX structure:

\[ g_1(z^{-1}) = [z^{-1} \ldots z^{-n}] | 0 \ldots 0 | 0 \ldots 0] \]
\[ n_1(z^{-1}) = [0 \ldots 0 | z^{-1} \ldots z^{-n}] | 0 \ldots 0] \]

The predictions for \( G(z, \theta) \) and \( H(z, \theta) \) defined by (3)-(4) are generated by a model \( M(\theta) : \)

\[ \hat{y}(t, \theta) = H^{-1}(z, \theta)G(z, \theta)u(t) + (1 - H^{-1}(z, \theta))y(t) \]

with the obvious definitions for \( W(z, \theta) \) and \( z \). For given polynomials \( n_0, d_0, n_0, d_0, h_0, d_1, h_1, i = 0, 1 \), the set of all possible models with structure (3)-(4) is called the model set \( M \triangleq \{ M(\theta), \forall \theta \in D_\theta \subseteq \mathbb{R}^d \} \). In some parts of this paper we shall consider the following assumption.

**Assumption 1.** The real system \( S \) belongs to the model set \( M \) (or simply \( S \in M \)), i.e. \( \exists \theta_0 \) such that

\[ G(z, \theta_0) = G_0(z) \quad \text{and} \quad H(z, \theta_0) = H_0(z). \]

Prediction Error Identification of \( \theta \) based on \( N \) input-output data has the property that under mild conditions the parameter estimate \( \hat{\theta}_N \) converges, for \( N \to \infty \), to a value \( \theta^* \triangleq \arg \min_{\theta \in D_\theta} V(\theta) \), with

\[ V(\theta) \triangleq E[y(t) - \hat{y}(t, \theta)]^2. \]

If \( S \in M \) and if \( \theta_N \xrightarrow{N \to \infty} \theta_0 \), then the parameter error converges to a Gaussian random variable:

\[ \sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{N \to \infty} N(0, P_\theta) \]

where \( P_\theta = [I(\theta)]^{-1} |_{\theta = \theta_0} \), with

\[ I(\theta) = \frac{1}{\sigma^2_e} E \left[ \psi(t, \theta) \psi(t, \theta)^T \right], \]

\[ \psi(t, \theta) = \frac{\partial y(t)}{\partial \theta} = \nabla_\theta W(z, \theta)z(t). \]

Here \( W(z, \theta) \triangleq [G(z, \theta) \quad H(z, \theta)] \) and \( \nabla_\theta W(z, \theta) \triangleq \partial W(z, \theta) / \partial \theta \) is a \( d \times 2 \)-matrix of stable rational transfer functions. We refer to \( I(\theta) \) as the information matrix at \( \theta \), although in the literature this term usually refers only to its value at \( \theta = \theta_0 \). The matrix \( I(\theta) \) is positive semi-definite by construction.

In this paper we study the solutions of the problem

\[ \min_{\theta \in D_\theta} V(\theta) \]

where \( V(\theta) \) is defined in (6), and we examine conditions on the model structure properties and on the data that ensure the existence and uniqueness of the solution of (10).

Depending on whether or not Assumption 1 is satisfied, problem (10) may exhibit quite distinct properties. When \( S \in M \), it is reasonable to demand that the identification procedure provides the real system, since this is possible within the chosen model set. We then want \( \theta_0 \) to be the unique parameter value for which Assumption 1 holds and the unique global minimum of problem (10). But when \( S \notin M \), this is impossible, and the objective of an identification is necessarily less ambitious. In this case any global minimum could be an acceptable outcome, and uniqueness of the global minimum might not be a requirement. However, isolation of the global minimum must be required, since the existence of a continuum of global minima leads to numerical difficulties and also
implies that problem (10) may have unbounded solutions \( \theta \).

Conditions on the data and the model structure that guarantee existence, isolation and uniqueness of the solutions of (10) are the object of this paper. We separately analyze four different situations: with or without Assumption 1, in open and in closed loop. We start by presenting some background concepts and tools that affect the properties of the solutions of (10).

3. IDENTIFIABILITY, INFORMATIVE DATA, AND THE INFORMATION MATRIX

Several formal concepts of identifiability have been proposed in the scientific literature, and these definitions have evolved over the years. Here we adopt the uniqueness-oriented definition proposed in Ljung [1976], which refers to the injectivity of the mapping from parameter space to the space of transfer function models.

Definition 1. (Identifiability) A parametric model structure \( M(\theta) \) is locally identifiable at a value \( \theta \) if \( \exists \delta > 0 \) such that, for all \( \theta \) in \( ||\theta - \theta_1|| \leq \delta \):

\[ W(e^{j\omega}, \theta) = W(e^{j\omega}, \theta_1) \]

at almost all \( \omega \Rightarrow \theta = \theta_1 \).

The model structure is globally identifiable at \( \theta_1 \) if the same holds for \( \delta \rightarrow \infty \). It is called globally identifiable if it is globally identifiable at almost all \( \theta_1 \).

Thus, global identifiability relates to the injectivity of the mapping from \( \theta \) to the model \( M(\theta) \). Most commonly used model structures (except ARX) are not globally identifiable, but they are globally identifiable at all values \( \theta \) that do not cause pole-zero cancellations; see Chapter 4 in Ljung [1999]. We introduce the identifiability Gramian \( \Gamma(\theta) \in \mathbb{R}^{d \times d} \):

\[ \Gamma(\theta) \triangleq \int_{\Delta} \nabla_{\theta} W(e^{j\omega}, \theta) \nabla_{\theta} W^H(e^{j\omega}, \theta) d\omega \]  

where for any \( M(e^{j\omega}) \), the notation \( M^H(e^{j\omega}) \) denotes \( M^T(e^{-j\omega}) \). The relevance of this matrix (and the name “identifiability Gramian”) stems from the fact that the positive definiteness of \( \Gamma(\theta_1) \) is a sufficient condition for local identifiability at \( \theta_1 \); see proof 4G.4 in Ljung [1999].

Proposition 3.1. A parametric model structure \( M(\theta) \) is locally identifiable at \( \theta_1 \) if \( \Gamma(\theta) \) is nonsingular at \( \theta_1 \).

Proof. For \( \theta \) close to \( \theta_1 \), we can write

\[ W(e^{j\omega}, \theta) = W(e^{j\omega}, \theta_1) + (\theta - \theta_1)^T \nabla_{\theta} W(e^{j\omega}, \theta_1) + \sigma(|\theta - \theta_1|) \]

where \( \lim_{\theta \to \theta_1} \sigma(|\theta - \theta_1|) = 0 \). Therefore,

\[ \int_{-\pi}^{\pi} ||W(e^{j\omega}, \theta) - W(e^{j\omega}, \theta_1)||^2 d\omega \]

\[ = (\theta - \theta_1)^T \Gamma(\theta_1)(\theta - \theta_1) + \rho(|\theta - \theta_1|^2) \]

where \( \lim_{\theta \to \theta_1} \rho(|\theta - \theta_1|^2) = 0 \). The result follows from the definition of local identifiability.

Identifiability (local, or global) is a property of the parametrization of the model \( M(\theta) \). It tells us that if the model structure is globally identifiable at some \( \theta_1 \), then there is no other parameter value \( \theta \neq \theta_1 \) that yields the exact same predictor \( W(z, \theta) \) as the predictor \( W(z, \theta_1) \) of \( M(\theta_1) \). However, it does not guarantee that two different models in the model set \( M \) cannot produce the same prediction errors when driven by the same data. This requires, additionally, that the data set is informative enough to distinguish between different predictors. The classical definition of informative data with respect to a model structure is as follows; see e.g. Ljung [1999].

Definition 2. (Informative data - classical) A quasistationary data set \( z(t) \) is called informative with respect to a parametric model set \( \{ M(\theta), \theta \in \Theta \} \) if, for any two models \( W(z, \theta_1) \) and \( W(z, \theta_2) \) in that set,

\[ E[(W(z, \theta_1) - W(z, \theta_2)z(t))^2] = 0 \]  

\[ \iff W(e^{j\omega}, \theta_1) = W(e^{j\omega}, \theta_2) \text{ at almost all } \omega. \]

This classical definition of informative data with respect to a parametric model set is a global one: (13) must hold between any pair \( \theta_1 \) and \( \theta_2 \) in \( \Theta \). If, in addition, the model structure is globally identifiable at \( \theta_1 \), say, then (13) implies that \( \theta_2 = \theta_1 \), i.e. there can be no \( \theta_2 \neq \theta_1 \) for which

\[ E[(W(z, \theta_1) - W(z, \theta_2)z(t))^2] = 0 \]  

Definition 2 was introduced to guarantee uniqueness of the global minimum of \( V(\theta) \) for identifiable model structures. One contribution of this paper will be to show that this requirement is unnecessarily strong; we introduce the weaker concept of informative data at a given parameter.

Definition 3. (Informative data - new) A quasistationary data set \( z(t) \) is called locally informative at \( \theta_1 \in \Theta \) with respect to a parametric model set \( \{ M(\theta), \theta \in \Theta \} \) if

\[ \exists \delta > 0 \text{ such that, for all } \theta \text{ in } ||\theta - \theta_1|| \leq \delta, \text{ we have} \]

\[ E[(W(z, \theta) - W(z, \theta_1)z(t))^2] = 0 \]  

\[ \iff W(e^{j\omega}, \theta) = W(e^{j\omega}, \theta_1) \text{ at almost all } \omega. \]

It is called globally informative at \( \theta_1 \) if the same holds for \( \delta \rightarrow \infty \).

These definitions of informative data are with respect to a parametric model set, not with respect to the real system. In a PEI experiment, one first selects an identifiable model structure, which defines a model set; this is a user’s choice. The choice of experimental conditions, often also a user’s choice, must make the data informative (in the classical, global, or local sense) with respect to that model set. Since the data are generated by the real system, in open or in closed loop, the conditions that make the data informative with respect to some model set depend on the real system and the possible feedback configuration.

We now turn to the information matrix. Combining (8) and (9) and using Parseval’s relationship yields:

\[ I(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \nabla_{\theta} W(e^{j\omega}, \theta) \Phi_z(\omega) \nabla_{\theta} W^H(e^{j\omega}, \theta) d\omega \]

where \( \Phi_z(\omega) \) is the power spectrum of the data \( z(t) \) generated by an identification experiment. Comparing (15) with (11) shows how the information matrix combines information about the identifiability of the model structure and about the informativity of the data (through \( \Phi_z(\omega) \)). We note that \( I(\theta) > 0 \) only if \( \Gamma(\theta) > 0 \), but noninformative data \( z(t) \) will yield a \( \Phi_z(\omega) \) that causes the rank of \( I(\theta) \)
to be lower than the rank of \( \Gamma(\theta) \); see Gevers et al. [2009]. When the system is driven only by noise, the identifiability Gramian can be written as:

\[
\Gamma(\theta) \triangleq \int_{-\pi}^{\pi} \nabla_{\theta} W_{e}(e^{j\omega}, \theta) \, \nabla_{\theta} W_{e}^{\ast}(e^{j\omega}, \theta) \, d\omega
\]

where \( \nabla_{\theta} W_{e}(e^{j\omega}, \theta) \) is a \( d \)-vector of rational transfer functions (see Gevers et al. [2009]). The information matrix is then \( I(\theta) = \sigma_{2}^{2} \Gamma(\theta) \). This explains why the nonsingularity of \( I(\theta) \) has often been taken as a criterion for identifiability in the statistical literature where time-series without measured inputs are prevalent: Rothenberg [1971], Bowden [1973] and others. However, under the influence of Deistler, Hannan and others (see e.g., Deistler [1989]), statisticians and econometricians later also viewed identifiability as the injectivity of the map from \( \theta \) to \( M(\theta) \).

4. THE MINIMIZERS OF \( V(\theta) \)

In Section 3 we have defined identifiability, informative data and the rank of the information matrix. In this section we examine the properties of the minimizers of \( V(\theta) \). When the input spectrum has a continuous support, the asymptotic criterion generically has a unique minimum at a value \( \theta^* \) and the analysis is simple. Here we focus on the more difficult situation where the spectrum of \( u \) (or \( r \)) is discrete, and we examine conditions under which the set of minimizers of \( V(\theta) \) may be affine in \( \theta \).

Define the set of minimizers of the cost function \( V(\theta) \) as

\[
\Theta^* \triangleq \{ \arg \min_{\theta} V(\theta) \} \subset D_{\theta}.
\]

The question of isolation and uniqueness of the global minimum can be studied by characterizing the structure and cardinality of this set. We perform this analysis separately for open-loop and closed-loop identification. In order to characterize the set of minimizers, we use the following result.

**Lemma 4.1.** Consider a transfer function of the form (3) with \( \theta \in \mathbb{R}^{k} \), and a set of distinct frequencies \( \Omega = \{\omega_{1}, \omega_{2}, \ldots, \omega_{q}\} \). Define \( \Theta^* \triangleq \{ \theta : G(e^{j\omega}, \theta) = G_{0}(e^{j\omega}) \; \forall \omega \in \Omega \} \). The set \( \Theta^* \) is affine.

**Proof:** The set \( \Theta^* \) is, by definition, the solution set of the system of linear equations

\[
\begin{align*}
-g_{0}(e^{j\omega_{i}}) + \theta^{T} g_{1}(e^{j\omega_{i}}) = 0, & \quad i = 1, \ldots, q.
\end{align*}
\]

Define \( G_{0}(e^{j\omega_{i}}) \triangleq g_{0}(e^{j\omega_{i}}) \) and

\[
\begin{align*}
a_{i}(e^{j\omega_{i}}) \triangleq \begin{bmatrix} g_{0}(e^{j\omega_{i}}) & g_{1}(e^{j\omega_{i}}) \end{bmatrix}^{T} = g_{0}(e^{j\omega_{i}}) d_{0}(e^{j\omega_{i}}) + g_{1}(e^{j\omega_{i}}) d_{1}(e^{j\omega_{i}}).
\end{align*}
\]

Note that \( a_{i}(e^{j\omega_{i}}) \) is a row \( d \)-vector of polynomials. Now (16) can be rewritten as

\[
\begin{align*}
A_{i} e^{j\omega_{i}} \theta = b_{i}, & \quad i = 1, \ldots, q,
\end{align*}
\]

where \( A_{i} \) is a \( q \times d \) matrix whose rows are \( a_{1}, a_{2}, \ldots, a_{q} \) and \( b \) is a \( q \)-vector whose elements are the \( b_{i} \)’s. The set of all solutions of a system of linear equations is an affine subspace, which completes the proof.

4.1 Open-loop identification

Assumption 1 (\( S \in M \)) states that there exists a parameter \( \theta_{0} \) such that \( G(z, \theta_{0}) = G_{0}(z) \) and \( H(z, \theta_{0}) = H_{0}(z) \). Let us now work under a weaker assumption.

**Assumption 2.** There exists \( \theta^{*} \in \mathbb{R}^{d} \) such that for some finite set \( \Omega = \{\omega_{1}, \omega_{2}, \ldots, \omega_{q}\} \) of distinct frequencies:

\[
\begin{align*}
(1) & \quad G(e^{j\omega}, \theta^{*}) = G_{0}(e^{j\omega}) \forall \omega \in \Omega \\
(2) & \quad H(e^{j\omega}, \theta^{*}) = H_{0}(e^{j\omega}) \forall \omega \in [-\pi, \pi).
\end{align*}
\]

Part (2) is equivalent with part (2) of Assumption 1, but part (1) is weaker than part (1) of Assumption 1: it requires that the model fits the real system only at a finite number \( q \) of frequencies.

**Theorem 4.1.** Consider the bilinear model structures of the form (3)-(4), and suppose that Assumption 2 holds for some \( \theta^{*} \) with \( \Omega = \{\omega_{1}, \omega_{2}, \ldots, \omega_{q}\} \). If the identification is performed in open loop, and if \( \Omega_{u} \) (the support of \( u \)) is chosen as \( \Omega_{u} = \Omega^{*} \), then the minimizer set \( \Theta^{*} \) is affine.

**Proof:** In open-loop identification the criterion \( V(\theta) \) can be written as

\[
V(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{G(e^{j\omega}, \theta) - G_{0}(e^{j\omega})}{H(e^{j\omega}, \theta)} \right|^{2} \Phi_{u}(\omega) d\omega
\]

where \( \Phi_{u}(\omega) \) is nonzero only at the frequencies contained in \( \Omega^{*} \), it follows that the first term of \( V(\theta) \) is zero; the second term is equal to \( \sigma_{2}^{2} \) since \( H(e^{j\omega}, \theta^{*}) = H_{0}(e^{j\omega}) \) for all \( \omega \). Since \( \sigma_{2}^{2} \) is a lower bound for \( V(\theta) \), it follows that \( \theta^{*} \) is a minimizer of \( V(\theta) \). Since \( \theta^{*} \) obeys Assumption 2, it follows that \( \Theta^{*} \) is affine.

**Theorem 4.1** is an extension to the model structures (3)-(4) of a result of Garatti et al. [2004] where affinity was proved for the classical model structures and for the situation \( S \in M \). Observe that the minimizer set \( \Theta^{*} \) may contain one or an infinity of points, depending on the relation between the number \( q \) of support points in \( \Omega^{*} \) and the number of free parameters \( \theta_{0} \) that appear in \( G(e^{j\omega}, \theta) \).

4.2 Closed-loop identification

For closed-loop identification, our results are presently limited to the situation where \( S \in M \).

**Theorem 4.2.** Consider closed-loop identification with the model structure (3)-(4) and with a controller \( K(z) = \frac{n^{c}(z)}{\pi(z)} \). Assume that \( S \in M \) and that the reference \( r \) has a discrete spectrum with support \( \Omega_{r} \). Then the minimizer set \( \Theta^{*} \) of \( V(\theta) \) is affine in \( \theta \) if the following equation is affine in \( \theta \):

\[
\begin{align*}
n^{b}(z, \theta) d^{b}(z, \theta) = d^{b}(z, \theta) d^{b}(z, \theta) + n^{c}(z) n^{c}(z) \end{align*}
\]

where \( n^{b}, d^{b} \) are defined in (3)-(4).

**Proof:** The proof is given in the Appendix.

2 By free parameters in \( G(e^{j\omega}, \theta) \) we mean those components \( \theta_{k} \) of \( \theta \) that are not determined by part (2) of Assumption 2.
The following corollary enumerates a number of cases where the set $\Theta^*$ is affine in $\theta$.

**Corollary 4.1.** Let $G(z, \theta)$ and $H(z, \theta)$ be rational functions of the form (3)-(4). Then the set $\Theta^*$ is affine in $\theta$ if $n^g(z, \theta)$, $d^h(z, \theta)$, and $n^h(z, \theta)$ are affine parametrizations, and either one of the following conditions is satisfied:

1. $n^h(z, \theta) = d^h(z, \theta)$
2. $d^h(z, \theta) = d^h(z, \theta)$
3. $d^h(z, \theta) = [d^h(z) d^h(z, \theta) + n^e(z) n^\theta(z, \theta)]$
4. $d^h(z, \theta) = \frac{n^e(z)}{1 - d^e(z)} n^\theta(z, \theta)$

**Proof:** It is easy to see that when any of these conditions is satisfied, equation (21) becomes affine in $\theta$. □

The family of model structures satisfying the conditions of Corollary 4.1 encompasses (but is not limited to) the “classical” model structures ARX, ARMAX and OE. It also encompasses BJ model structures, but only in open loop; this is consistent with the observations made in Garatti et al. [2004].

### 5. MAIN RESULTS

We are now ready to establish the connections between the following four properties, established at a parameter $\theta^*$.

**Properties**

1. The information matrix is positive definite: $I(\theta^*) > 0$
2. The model structure is locally identifiable at $\theta^*$: $\Gamma(\theta^*) > 0$
3. The data are locally informative at $\theta^*$
4. $\theta^*$ is the unique minimum of $V(\theta^*)$

We first establish the following result, which is valid whether or not the system is in the model set.

**Theorem 5.1.** Let $\theta_1$ be any value in $D_{\theta}$. Then the following two statements are equivalent:

(i) $I(\theta_1) > 0$
(ii) $\Gamma(\theta_1) > 0$ and the data are locally informative at $\theta_1$.

**Proof:** The proof is given in the Appendix.

**Comment.** The result applies to open-loop and closed-loop identification. We observe that $I(\theta^*) > 0$ is equivalent with “$\Gamma(\theta^*) > 0$ AND the data are locally informative at $\theta^*$.” Since $I(\theta^*) > 0$ necessarily implies $\Gamma(\theta^*) > 0$, adding the requirement of local informativity at $\theta^*$ is precisely what is needed to ensure the equivalence.

#### 5.1 System in the model set: $S \in \mathcal{M}$

**Theorem 5.2.** Assume that the model structure is such that $S = \mathcal{M}(\theta_0)$ for some $\theta_0$. Assume further that the minimizer set $\Theta^*$ of $V(\theta)$ is affine in $\theta$.\(^3\) Consider the four properties above but with $\theta^*$ replaced by $\theta_0$. Then the following implications hold.

A. $(1) \iff \{(2) + (3)\}$
B. $(1)$ (or equivalently $\{(2) + (3)\}$) $\implies (4)$
C. $\{(4) + (2)\} \implies (1)$

**Proof:** see the Appendix.

\(^3\) In Section 4 we have given a number of situations under which this holds.

#### 5.2 System not in the model set: $S \notin \mathcal{M}$

When $S \notin \mathcal{M}$, $\theta^* \triangleq \arg \min_\theta V(\theta)$ is a function of $\Phi(\omega)$ (in open loop) or $\Phi(\omega)$ in closed loop. Thus, the minimum (or minima) of $V(\theta)$ depend on the data set. This is a major difficulty in the analysis of this situation. For open-loop identification, we have the following result.

**Theorem 5.3.** Consider the affine parametrization (3)-(4). Then $V(\theta)$ has a unique global minimum at $\theta^*$ only if the data are globally informative at $\theta^*$.

**Proof:** see the Appendix.

### 6. CONCLUSION

We have replaced the traditional definition of informative data by the new concept of informative data at a parameter value. With this weaker concept, we have established the equivalence between a full rank information matrix at a parameter value and identifiability plus informativity at the same parameter. When the system is in the model set we have established further connections between existence of a unique global minimum of the identification criterion, full rank information matrix, and identifiability and informativity at the true parameter value, provided the minimizer set of the criterion is affine. In Section 4 we have shown a range of cases where this affine property holds. Finally we have shown that the global informativity at the minimum of the cost function is a necessary condition for the existence of a unique global minimum of the criterion.

### Appendix

**Proof of Theorem 4.2:** This theorem is an extension to the model structure (3)-(4) of Theorem 5 of Garatti et al. [2004]. Following the same derivations as in the proof of their Theorem 5, it is easy to show that the set of minimizers is described by $\Theta^* = \Theta^\theta \cap \Theta^s$ with

$$\Theta^\theta \triangleq \{ \theta : G(e^{j\omega}, \theta) = G_0(e^{j\omega}) \ \forall \omega \in \Omega_r \}$$

and

$$\Theta^s \triangleq \{ \theta : H_0(e^{j\omega})[1 + K(e^{j\omega})G(e^{j\omega}, \theta)] = H(e^{j\omega}, \theta) \ [1 + K(e^{j\omega})G_0(e^{j\omega})] \ \forall \omega \}.$$ 

$\Theta^*$ is the set of solutions of (omitting dependence on $\omega$):

$$n^h(\theta_0)[d^h d^h(\theta) + n^e n^\theta(\theta)] d^h(\theta) d^h(\theta) = 0$$

which must hold at all $\omega \in [-\pi, \pi]$. In general the solution set is an ellipsoid, which is not affine since $\phi(\omega)$ is quadratic in $\theta$, and hence different minimizers can be isolated from each other. We note that the factors on both sides of (22) do not depend on $\theta$ have no influence on the affine (in $\theta$) character of this equation. Thus, the solution of (22) is affine in $\theta$ if and only if (21) is affine in $\theta$. Since the intersection of affine subspaces is an affine subspace, the intersection $\Theta^\theta \cap \Theta^s$ is then also affine. □

**Proof of Theorem 5.1:** Consider first that $I(\theta_1) > 0$ at some $\theta_1$. Then, necessarily $\Gamma(\theta_1) > 0$ since $\alpha^T I(\theta_1) = 0$ implies $\alpha^T I(\theta_1) = 0$ for any $\alpha \in \mathbb{R}^d$. Now for $\theta$ close to $\theta_1$ we have

$$E \left[ \left| \hat{y}(t, \theta) - \hat{y}(t, \theta_1) \right|^2 \right] = (\theta - \theta_1)^T I(\theta_1) (\theta - \theta_1) + \rho(\|\theta - \theta_1\|^2)$$

(23)
where \( \lim_{\theta \to \theta_1} \frac{\rho(|\theta-\theta_0|^2)}{\rho(|\theta-\theta_1|^2)} = 0 \). It follows from \( I(\theta_1) > 0 \) that \( \bar{E}[(\hat{y}(t, \theta) - \hat{y}(t, \theta_1))^2] = 0 \) implies \( \theta = \theta_1 \), i.e. the data are locally informative.

We prove the converse by contradiction. It has been shown in Gevers et al. [2009] that the regressor \( \psi(t, \theta) \) in (8)-(9) can always be written as

\[
\psi(t, \theta) = V_w(z, \theta)w(t) + V_e(z, \theta)e(t)
\]

where \( w(t) \) is either \( u(t) \) (in open loop) or \( r(t) \) (in closed loop), while \( V_w(z, \theta) \) and \( V_e(z, \theta) \) are corresponding d-vectors of stable rational transfer functions, and where \( e(t) \) is white noise independent of \( w(z) \). Suppose \( \exists z \in \mathbb{R}^d \) with \( \alpha \neq 0 \) such that \( \alpha^T I(\theta_1) = 0 \). It then follows from Gevers et al. [2009] that either \( \alpha^T V_w(z, \theta_1) \equiv 0 \) and \( \alpha^T V_e(z, \theta_1) \equiv 0 \), which is equivalent with \( \alpha^T I(\theta_1) = 0 \), or \( \bar{E}[\alpha^T V_w(z, \theta_1)w(t)]^2 = 0 \). Using the results of Gevers et al. [2009] again, the last expression implies that the degree of richness of \( w(t) \) is strictly smaller than the degree of the numerator polynomial of \( \alpha^T V_w(z, \theta_1) \); and since the degree of this numerator polynomial is smaller than the degree of the predictor at \( \theta_1 \), it follows that \( w(t) \) is not informative at \( \theta_1 \). Thus \( \alpha^T I(\theta_1) = 0 \) implies that either \( \alpha^T I(\theta_1) = 0 \) or that the data are not informative at \( \theta_1 \).

\[\text{Proof of Theorem 5.2:} \] Equivalence A has been proved in Theorem 5.1. Let now \( \theta \neq \theta_0 \) be any value in \( D_0 \). Then

\[
V(\theta) - V(\theta_0) = \bar{E}[(\varepsilon(t, \theta) - \varepsilon(t, \theta_0))(\varepsilon(t, \theta_0))] + \bar{E}[(\varepsilon(t, \theta) - \varepsilon(t, \theta_0))^2].
\]

(25)

Since \( M(\theta_0) = S \), we have \( \varepsilon(t, \theta_0) = \varepsilon(t) \). Also, \( \varepsilon(t, \theta) - \varepsilon(t, \theta_0) = \hat{y}(t, \theta) - \hat{y}(t, \theta_0) \). Therefore the first term in (25) is zero because \( e(t) \) is independent of past data. Thus:

\[
V(\theta) - V(\theta_0) = \bar{E}[(\hat{y}(t, \theta) - \hat{y}(t, \theta_0))^2] \geq 0.
\]

(26)

Let \( B(\theta_0) \equiv \{ \theta \text{ s.t. } |\theta - \theta_0| < \delta \} \) for some positive \( \delta \). For \( \theta \in B(\theta_0) \) we have

\[
\hat{y}(t, \theta) - \hat{y}(t, \theta_0) = (\theta - \theta_0)^T \psi(t, \theta_0) + \sigma(|\theta - \theta_0|)
\]

where \( \lim_{\theta_0 \to \theta_0} \frac{\sigma(|\theta_0 - \theta_0|)}{\sigma(|\theta - \theta_0|)} = 0 \). Combining with (25), it follows that for \( \theta \in B(\theta_0) \):

\[
V(\theta) - V(\theta_0) = (\theta - \theta_0)^T I(\theta_0)(\theta - \theta_0) + \rho(|\theta - \theta_0|^2)
\]

(27)

where \( \lim_{\theta_0 \to \theta_0} \frac{\rho(|\theta_0 - \theta_0|)}{\rho(|\theta - \theta_0|^2)} = 0 \). Thus, (1) implies \( V(\theta) > V(\theta_0) \) for \( \theta \in B(\theta_0) \) and, since the minimizer set \( \Theta^* \) is affine, this implies (4) which establishes B. Finally, assume that (4) holds. It then follows from (26) that \( \bar{E}[(\hat{y}(t, \theta) - \hat{y}(t, \theta_0))^2] = 0 \) implies \( \theta = \theta_0 \). Therefore the data are globally (and hence also locally) informative at \( \theta_0 \), i.e. (3) holds at \( \theta_0 \). Since \( \{(3) + (2)\} \) is equivalent to (1), it follows that \( \{(4) + (2)\} \implies (1) \), and thus C is proved.

\[\text{Proof of Theorem 5.3:} \] By contradiction. Suppose that the data are not globally informative at \( \theta^* \). Then there exists \( \theta_1 \neq \theta^* \) such that

\[
\bar{E}[(\hat{y}(t, \theta_1) - \hat{y}(t, \theta^*))^2] = 0.
\]

(28)

We omit henceforth the dependence on \( z \) and \( \omega \) whenever it is not essential. Some easy calculations lead to

\[
\bar{y}(\theta_1) - \bar{y}(\theta^*) = \{H^{-1}(\theta_1)[G(\theta_1) - G_0] - H^{-1}(\theta^*)[G(\theta^*) - G_0]\}u + [H^{-1}(\theta^*) - H^{-1}(\theta_1)]H_0 e
\]

Using Parseval's theorem and the independence of \( u(t) \) and \( e(t) \), it follows that \( \bar{E}[(\hat{y}(t, \theta_1) - \hat{y}(t, \theta^*))^2] \) is the sum of two integrals whose integrands are nonnegative for all \( \omega \). (28) then implies that each integrand must be identically zero. Therefore

\[
H(e^{i\omega}, \theta_1) = H(e^{i\omega}, \theta^*) \text{ for all } \omega.
\]

(30)

Inserting (30) into the second integrand, it also follows that

\[
\int_{-\pi}^{\pi} |G(e^{i\omega}, \theta_1)S_\omega - G(e^{i\omega}, \theta^*)S_\omega|^2 d\omega = 0
\]

where \( S_\omega \) is a spectral factor of \( \Phi_u \), i.e. \( \Phi_u = S_\omega S_\omega^* \). This in turn implies that:

\[
G(\epsilon^{i\omega}, \theta_1)S_\omega(\omega) = G(\epsilon^{i\omega}, \theta^*)S_\omega(\omega) \text{ for all } \omega.
\]

(31)

Now consider the cost evaluated at \( \theta_1 \):

\[
V(\theta_1) = \bar{E}[(\bar{y}(t, \theta_1) - y(t))^2] = \sigma_\epsilon^2 \int_{-\pi}^{\pi} \left| \frac{H(\epsilon^{i\omega}, \theta_1)}{H(\epsilon^{i\omega}, \theta_1)^2} \right|^2 d\omega + \int_{-\pi}^{\pi} \left| \frac{G(\epsilon^{i\omega}, \theta_1)S_\omega - G_0(\epsilon^{i\omega})S_\omega}{|H(\epsilon^{i\omega}, \theta_1)|^2} \right|^2 d\omega.
\]

(32)

The minimum cost, evaluated at \( \theta^* \), is identical to (32) with \( \theta_1 \) replaced by \( \theta^* \). It then follows from (30) and (31) that \( V(\theta_1) = V(\theta^*) \), i.e. \( \theta_1 \) is also a global minimum, which contradicts the hypothesis of the theorem.

\[\text{Acknowledgements} \]

The authors wish to thank Lennart Ljung and Roland Hildebrand for useful discussions on earlier versions of this paper.

\[\text{REFERENCES} \]


