Necessary and sufficient conditions for uniqueness of the minimum in Prediction Error Identification [1,2]

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

* *Electrical Engineering Department, Universidade Federal do Rio Grande do Sul, Porto Alegre-RS, Brazil.* `bazanela@ece.ufrgs.br`
** *ICTEAM, University of Louvain, B1348 Louvain-la-Neuve, Belgium, and Dept ELEC, Vrije Universiteit Brussel, Pleinlaan 2, B1050 Brussels, Belgium.* `Michel.Gevers@uclouvain.be`
*** *Delft Center for Systems and Control, Delft University of Technology, The Netherlands.* `x.j.a.bombois@tudelft.nl`

Abstract: The contribution of this paper is to establish computable necessary and sufficient conditions on the model structure and on the experiment under which the Prediction Error Identification (PEI) criterion has a unique global minimum. We consider a broad class of rational model structures whose numerator and denominator are affine in the unknown parameter vector; this class encompasses all classical model structures used in system identification. The main results in this paper rely on the standard assumption that the system is in the model set, while some intermediate results are valid even when this assumption does not hold (in particular Theorem 4.2 and Lemma 6.1). This is achieved by first establishing necessary and sufficient conditions on the model structure and on the experiment under which a global minimum is isolated; these conditions must hold locally, at the global minimum. A second contribution is to show that these conditions are equivalent to the nonsingularity of the information matrix at that minimum. For open loop identification and, with some additional constraints also for closed loop identification, the nonsingularity of the information matrix is then also equivalent to the uniqueness of the global minimum.

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

1. INTRODUCTION

The objective of this paper is to seek checkable and nonconservative conditions under which the Prediction Error Identification criterion has a unique global minimum. From a practical point of view, this is an important objective leading to a successful identification procedure. It is well known that the existence of a unique global minimum requires the combination of two ingredients: the identifiability of the model structure, and the informativity of the data: see e.g. Theorem 8.3 in Ljung [1999]. 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. Necessary and sufficient conditions for the existence of a unique global minimum of the PE criterion are available only for simple model structures, such as ARX, ARMA or OE model structures (Åström and Söderström [1974], Söderström [1975]). For more general model structures, sufficient conditions for the uniqueness of the global minimum of the PEI criterion exist but they are either conservative or difficult to enforce a priori Ljung [1999].

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*. Our motivation for deriving results on the minima of the identification criterion that are based on the positivity of the information matrix is enhanced by our recent results that connect richness of the experiment with the nonsingularity of the information matrix Gevers et al. [2009b]. Indeed, these results present the minimal richness conditions that make the information matrix nonsingular, for both open loop and closed loop experimental conditions.

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 positivity of the information matrix at a given parameter is equivalent with "identifiability" at

[1] Work supported by the Brazilian Ministry of Science and Technology through CNPq.
[2] Work supported by the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Science Policy Office (BELSPO), and by the Fonds National de la Recherche Scientifique (FNRS).

that parameter [3]: Rothenberg [1971], Bowden [1973]. In the presence of measured inputs, however, this is no longer the case. Indeed, the analysis in those papers was based on the idea that the measured output data were described by parametrized probability density functions, and the concept of identifiability was related only to the injectivity of the mapping from parameter to density function. Thus, the question of generating informative data by the selection of sufficiently rich input signals was not addressed.

In the engineering literature, the definition of the term "identifiability" has evolved over the years as we shall explain in Section 3, and eventually a clear separation was introduced between the concept of *identifiability of a model structure*, which is a parametrization issue, and *informative data*, which is the issue of applying sufficiently rich signals to the system. Thus, the concept of informative data was introduced, which together with identifiability of the model structures implies the existence of a global minimum of the identification criterion when the system is in the model set. Precise definitions will be given in Section 3.

Given the usefulness of the information matrix as an experiment design tool, as explained above, it was therefore tempting to replace the conditions on identifiability and informativity - that guarantee a unique global minimum - by equivalent conditions on the nonsingularity of the information matrix. Our attempts to prove such equivalence were unsuccessful because, as we show in this paper, the traditional definition of informative experiments is unnecessarily strong. We therefore introduce the new concept of *local informativity*, i.e. informativity at a given parameter value. With this new definition, we first show that a global minimum - say θ^* - of the PE criterion is an isolated minimum if and only if the model structure is locally identifiable at θ^* and the data are locally informative at θ^*. This necessary and sufficient result is valid independently of the chosen model structure, provided that it can describe exactly the real system for some parameter value (that is, provided that the true system belongs to the model set considered).

Our second main result is to show the equivalence between *local identifiability plus local informativity at a given* θ_1 and *positivity of the information matrix at* θ_1. This equivalence is true for model structures that are rational functions whose numerator and denominator are affine functions of the parameter vector and regardless of whether or not the true system belongs to the model set considered. Hence, if the information matrix is positive definite at some global minimum, that minimum is isolated. Using results from Gevers et al. [2009b] we then briefly recall how to create experimental conditions that produce a positive definite information matrix: these conditions are easy to compute and enforce.

Our third main result is to establish conditions under which the global minimum is not only isolated but also unique. This requires some affinity condition on the minimizer set of the PE criterion. We show that in open loop identification this affine property is always guaranteed for the class of model structures considered (i.e whose numerator and denominator are affine polynomials in the parameter vector), while in closed loop some additional conditions must hold.

A practical consequence of our results is that we now have easily checkable conditions on the choice of model structure and on the experiment that produce a nonsingular information matrix guaranteeing a unique global minimum of the identification criterion for a broad class of model structures.

The paper is organized as follows. We formulate the problem and establish the notations in Section 2. In Section 3 we present the definitions of identifiability and informative data at a given parameter value, and we relate them with classical definitions used in system identification and econometrics. The equivalence between these properties and the nonsingularity of the information matrix is established in Section 4 for model structures that are rational functions whose numerator and denominator are affine functions of the parameter vector. It is shown that these sets of equivalent conditions are necessary and sufficient for *isolation* of the global minimum. Verification and enforcement of this nonsingularity condition by means of design choices is studied in Section 5, where analysis tools recently developed are used to show that this can be conveniently performed. The uniqueness of the global minimum is studied in Section 6. For the same class of model structures as above we establish that nonsingularity of the information matrix is necessary and sufficient for uniqueness of the global minimum of the PEI criterion in open loop identification and, under additional conditions, also in closed loop identification. Conclusions are given in Section 7.

2. PROBLEM FORMULATION

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

$$\mathcal{S}: \quad y(t) = G_0(z)u(t) + H_0(z)e(t) \tag{1}$$

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 σ_e^2. Both transfer functions are rational and proper; furthermore, $H_0(z)$ is monic, i.e. $H_0(\infty) = 1$. To be precise, we shall define $\mathcal{S} \triangleq [G_0(z) \quad H_0(z)]$. This system may or may not be under feedback control with a proper rational stabilizing controller $K(z)$:

$$u(t) = K(z)[r(t) - y(t)]. \tag{2}$$

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 $\bar{E}[u(t)e(s)] = 0 \quad \forall s$; when they are generated in closed loop, we assume that $\bar{E}[r(t)e(s)] = 0 \ \ \forall s$. Here $\bar{E}[\cdot]$ is defined as

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

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

In this paper we consider the identification of linear models

$$y(t) = G(z,\theta)u(t) + H(z,\theta)e(t) \tag{3}$$

[3] The different notions of identifiability will be defined precisely in Section 3.

where $G(z,\theta)$ and $H(z,\theta)$ are rational and causal transfer functions. The optimal one-step-ahead predictor is given by

$$\hat{y}(t,\theta) = H^{-1}(z,\theta)G(z,\theta)u(t) + (1 - H^{-1}(z,\theta))y(t)$$
$$\triangleq W_u(z,\theta)u(t) + W_y(z,\theta)y(t) \triangleq W(z,\theta)\xi(t) \quad (4)$$

where

$$W(z,\theta) \triangleq [W_u(z,\theta) \;\; W_y(z,\theta)] \;\text{ and }\; \xi(t) \triangleq \begin{pmatrix} u(t) \\ y(t) \end{pmatrix}. \quad (5)$$

The function $W(z,\theta)$ is an injective map $W(z,\theta) : D_\theta \to \mathcal{Q}$ from some set $D_\theta \subseteq \Re^d$ into the set of predictors, defined by $\mathcal{Q} = \{W(z,\theta), \theta \in D_\theta\}$. For any given $\theta \in D_\theta$, the result of this mapping is called a model and denoted by $M(\theta) \triangleq W(z,\theta)$. The set $\mathcal{M} \triangleq \{M(\theta), \forall\theta \in D_\theta\}$ is called the model set.

Except when otherwise specified we shall consider the following assumption.

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

$$G(z,\theta_0) = G_0(z) \;\text{ and }\; H(z,\theta_0) = H_0(z).$$

■

Prediction Error Identification (PEI) of θ based on N input-output data consists in finding, among all the models in the pre-specified model class, the one that provides the minimum value for the prediction error criterion, that is, finding the solution of the following optimization

$$\hat{\theta}_N = \arg\min_\theta V_N(\theta)$$
$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}[\hat{y}(t,\theta) - y(t)]^2$$

PEI has the property that under mild conditions the parameter estimate $\hat{\theta}_N$ converges w.p.1, for $N \to \infty$, to a set

$$\Theta^* = \{\theta^* \triangleq \arg\min_{\theta\in D_\theta} V(\theta)\}, \quad (6)$$

with

$$V(\theta) \triangleq \bar{E}[y(t) - \hat{y}(t,\theta)]^2. \quad (7)$$

If $\mathcal{S} \in \mathcal{M}$ and if $\hat{\theta}_N \overset{N\to\infty}{\longrightarrow} \theta_0$, then the parameter error converges to a Gaussian random variable:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \overset{N\to\infty}{\longrightarrow} N(0, P_{\theta_0}), \quad (8)$$

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

$$I(\theta) = \frac{1}{\sigma_e^2}\bar{E}\left[\psi(t,\theta)\psi(t,\theta)^T\right], \quad (9)$$

$$\psi(t,\theta) = \frac{\partial\hat{y}(t|t-1,\theta)}{\partial\theta} = \nabla_\theta W(z,\theta)\xi(t). \quad (10)$$

Here $\nabla_\theta W(e^{j\omega},\theta) \triangleq \frac{\partial W(e^{j\omega},\theta)}{\partial\theta}$ is a $d \times 2$-matrix of stable rational transfer functions. We refer to $I(\theta)$ as the *information matrix at* θ, 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 following problem

$$\min_{\theta\in D_\theta} V(\theta) \quad (11)$$

where $V(\theta)$ is defined in (7). Specifically, we examine conditions on the model structure and on the data that ensure the isolation and uniqueness of the solution of (11). Isolation and uniqueness are defined below.

Definition 1. The point x_0 is called a global minimum of the function $f(x) : \Re^n \to \Re$ if $f(x) \geq f(x_0)\ \forall x \neq x_0$. A global minimum x_0 is unique if $f(x) > f(x_0) \forall x \neq x_0$; it is said to be isolated if $\exists\delta > 0$ such that, for all $x \in ||x - x_0|| \leq \delta,\ x \neq x_0$, we have $f(x) > f(x_0)$. ■

Conditions on the data and the model structure properties that guarantee existence, isolation and uniqueness of the solutions of (11) are the object of this paper. The relationship of these properties with the nonsingularity of $I(\theta)$ is of key relevance, because easily computable necessary and sufficient conditions for that are now available (only sufficient conditions were previously known: see Gevers et al. [2009b]). These conditions are on both the model structure and the richness of the input signal - the parameters of choice for the user. We start by presenting the background concepts and tools that affect the properties of the solutions of (11).

3. IDENTIFIABILITY, INFORMATIVE DATA, AND THE INFORMATION MATRIX

Several concepts of identifiability have been proposed in the scientific literature, and these definitions have evolved over the years. These different concepts are by no means equivalent, and they don't even intend to represent the same idea. The word "identifiable" was at first used to describe the property of an identification problem having a unique parameter value as a solution, like in Rothenberg [1971], Bowden [1973], Ljung [1976], Gustavsson et al. [1977]. This definition was prevalent until the late 1970's and very early 1980's, a period in which a large number of papers were devoted to the identifiability of systems operating in closed loop: see e.g. Gustavsson et al. [1977], Ng et al. [1977], Anderson and Gevers [1982]. An excellent survey of this work of the 1970's can be found in Solo [1986].

In Prediction Error Identification, this definition is equivalent with the objective function $V(\theta)$ having a unique global minimum. But this uniqueness property depends on two conceptually different ingredients which must be set up in an identification procedure: the model structure and the experimental conditions. These are two very distinct user's choices [4]. Thus it was realized that it would be much more enlightening to analyze these two practical ingredients separately, because these are the ingredients that the user of an identification procedure can directly

[4] Even though the choice of experimental conditions is not always totally in the user hand.

manipulate. The first instances of this separation can be traced to Glover and Willems [1974], Ljung [1976]. By 1987, with the publication of the first edition of Ljung [1999], these new definitions became completely prevalent, even though some usages of the old definition can still be found (see e.g. Van den Hof et al. [1992]). Thus the term "identifiable" designates the property, which refers only to the model structure, that two parameter values can not give the same model. Another property, usually (but not always) called "informativity of the experiment", refers to the ability of the experiment and of the resulting data to discriminate between different models. The separation between these two concepts and the introduction of the notion of informativity of the experiment were also crucial to the very important developments in experiment design of the last 25 years.

Here we adopt this more contemporary definition of identifiability, which refers to the injectivity of the mapping from parameter space to the space of transfer function models Ljung [1999]. Let us insist that this is a property of the model structure only, that the model structure is entirely a user's choice, and that it is equally valid for both open-loop and closed-loop identification.

Definition 2. **(Identifiability)** A parametric model structure $M(\theta)$ is locally identifiable at a value θ_1 if $\exists\ \delta > 0$ such that, for all θ in $\|\ \theta - \theta_1\ \| \leq \delta$:

$$W(e^{j\omega}, \theta) = W(e^{j\omega}, \theta_1) \ \text{ at almost all } \ \omega \Rightarrow \theta = \theta_1.$$

The model structure is globally identifiable at θ_1 if the same holds for $\delta \to \infty$. It is called globally identifiable if it is globally identifiable at almost all θ_1. ■

When $W(e^{j\omega}, \theta) = W(e^{j\omega}, \theta_1)$ at almost all ω we will write for short $W(z, \theta) \equiv W(z, \theta_1)$ - thus simplifying the notation. Accordingly, $W(z, \theta) \not\equiv W(z, \theta_1)$ means that this does not happen.

Thus, global identifiability relates to the injectivity of the mapping from θ to the model $M(\theta)$. General results on global identifiability of model structures using algebraic methods were derived in Ljung and Glad [1994].

Most standard model structures (except ARX) are not globally identifiable, but they are globally identifiable at all values θ that do not cause pole-zero cancellations: see Chapter 4 in Ljung [1999]. Less standard model structures are different.

Example 3.1. Consider the model structure

$$y(t) = \frac{ab}{(z-a)(z-b)} u(t) + e(t) \tag{12}$$

It is locally identifiable at any a, b but it is globally identifiable at no a, b: $\theta_1 = [a \quad b]^T$ always gives the same model as $\theta_2 = [b \quad a]^T$. On the other hand, there is no reason to require global identifiability in this case, since θ_1 and θ_2 give the same predictor. ■

We introduce the *identifiability Gramian* $\Gamma(\theta) \in \Re^{d \times d}$:

$$\Gamma(\theta) \triangleq \int_{-\pi}^{\pi} \nabla_\theta W(e^{j\omega}, \theta)\ \nabla_\theta W^H(e^{j\omega}, \theta)\ d\omega \tag{13}$$

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 θ_1; see problem 4G.4 in Ljung [1999]. We state this as a proposition, for the sake of completeness; a proof can be found, for example, in Bazanella et al. [2010].

Proposition 3.1. A parametric model structure $M(\theta)$ is locally identifiable at θ_1 if $\Gamma(\theta_1)$ is nonsingular. ■

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 θ_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 $\mathcal{M}$ cannot produce the same prediction errors when driven by the same data. This requires, additionally, that the data set carries enough information to distinguish between different predictor models. The classical definition of informative data with respect to a model structure is as follows Ljung [1999].

Definition 3. **(Informative data - classical)** A quasistationary data set $\xi(t)$ is called informative with respect to a parametric model set $\mathcal{M}$ if, for any two models $W(z, \theta_1)$ and $W(z, \theta_2)$ in that set,

$$\bar{E}[\hat{y}(t, \theta_1) - \hat{y}(t, \theta_2)]^2 = 0 \Longrightarrow W(z, \theta_1) \equiv W(z, \theta_2) \tag{14}$$

■

This classical definition is a global one: (14) must hold between any pair θ_1 and θ_2 in $\mathcal{D}_\theta$. Its relevance is made explicit in the following well known result (Theorem 8.3 in Ljung [1999]).

Proposition 3.2. If $\mathcal{S} \in \mathcal{M}$, that is, $\exists \theta_0 : M(\theta_0) = \mathcal{S}$, the model structure is globally identifiable at θ_0 and the data are informative, then θ_0 is the unique global minimum of $V(\theta)$. ■

So, informativity in this classical sense is, together with global identifiability, sufficient for achieving the desired property: uniqueness of the global minimum of $V(\theta)$. But are these conditions also necessary? The answer is no. The main contribution of this paper is to provide necessary and sufficient conditions for isolation/uniqueness, and the key to this is the following weaker concept of *local* informativity.

Definition 4. **(Informative data - new *local* definition)** A quasistationary data set $\xi(t)$ is called *locally informative at* $\theta_1 \in \mathcal{D}_\theta$ with respect to a parametric model set $\mathcal{M} = \{M(\theta), \theta \in \mathcal{D}_\theta\}$ if $\exists\ \delta > 0$ such that, for all θ in $\|\ \theta - \theta_1\ \| \leq \delta$, we have

$$\bar{E}[\hat{y}(t, \theta) - \hat{y}(t, \theta_1)]^2 = 0 \Longrightarrow \quad W(z, \theta) \equiv W(z, \theta_1). \tag{15}$$

It is called *globally informative at* θ_1 if the same holds for $\delta \to \infty$. It is said to be *totally informative* if it is globally informative at all θ.[5] ■

A similar (but not equivalent) definition of informative data, where the informativity is also made a local attribute at a particular parameter value, appears in Ljung and Glad [1994]. In that paper, two different conditions, one necessary and one sufficient, are given for uniqueness of the solution of the identification problem. These conditions are given in terms of the outcome of Ritt's algorithm (Ritt [1950]), which must be used to decide whether the model structure is identifiable and the data are informative.

Checking informativity is not an easy task in general, and it is often more interesting to work with the information matrix, on which most results of classical information theory and modern experiment design are based. So, we now turn to the information matrix. Combining (9) and (10) and using Parseval's relationship yields:

$$I(\theta) = \frac{1}{2\pi\sigma_e^2} \int_{-\pi}^{\pi} \nabla_\theta W(e^{j\omega}, \theta) \Phi_\xi(\omega) \nabla_\theta W^H(e^{j\omega}, \theta) d\omega \quad (16)$$

where $\Phi_\xi(\omega)$ is the power spectrum of the data $\xi(t)$ generated by an identification experiment. Comparing (16) with (13) shows how the information matrix combines information about the identifiability of the model structure and about the informativity of the data (through $\Phi_\xi(\omega)$). We note that $I(\theta) > 0$ only if $\Gamma(\theta) > 0$, but noninformative data $\xi(t)$ will yield a $\Phi_\xi(\omega)$ that causes the rank of $I(\theta)$ to be lower than the rank of $\Gamma(\theta)$: see Gevers et al. [2009b].

When the system is driven only by noise, it follows from (4), (9) and (10) that the information matrix becomes

$$I(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \nabla_\theta W_y(e^{j\omega}, \theta) H_0(e^{j\omega}) H_0^H(e^{j\omega}) \nabla_\theta W_y^H(e^{j\omega}, \theta) d\omega,$$

while the identifiability Gramian is written as:

$$\Gamma(\theta) \triangleq \int_{-\pi}^{\pi} \nabla_\theta W_y(e^{j\omega}, \theta)\ \nabla_\theta W_y^H(e^{j\omega}, \theta)\ d\omega.$$

The ranks of $I(\theta)$ and $\Gamma(\theta)$ are then identical, which partially 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, so that informativity of the experiment is not an explicit issue: Rothenberg [1971], Bowden [1973] and others. The other part of the explanation is that in those earlier works identifiability was given a different meaning, in which two different parameter values should result in different probability density functions for the predictors. This is in contrast to later definitions in which different parameter values are required to provide predictors that are different in the mean. However, under the influence of Deistler, Hannan and others (see e.g. Deistler [1989]), statisticians and econometricians later also turned their attention to the injectivity of the map from θ to $M(\theta)$, like in our current definition of identifiability.

[5] Total informativity here corresponds to the classical informativity as defined in Definition 3.

4. THE RELATIONSHIPS

In the following we provide results on the relationships between the local concepts of identifiability and informativity, the global minima of the identification criterion, and the positivity of the information matrix. We start with a preparatory Lemma.

Lemma 4.1. Consider the function

$$f(\theta) \triangleq \bar{E}[\hat{y}(t,\theta) - \hat{y}(t,\theta_1)]^2 \quad (17)$$

for some given θ_1. Then θ_1 is an isolated global minimum of $f(\theta)$ if and only if $M(\cdot)$ is locally identifiable at θ_1 and the data are locally informative at θ_1.

Proof i) IF. Assume first that the model structure is locally identifiable and the data are locally informative at θ_1, so that Definitions 2 and 4 apply for some $\delta > 0$. Let $\{\mathcal{B}_\epsilon(\theta)\}$ denote a ball of radius ϵ around θ. To prove that θ_1 is an isolated minimum of $f(\theta)$ we show that $f(\theta) \neq 0\ \forall\theta \in \{\mathcal{B}_\delta(\theta_1) \setminus \theta_1\}$ with δ as defined above. We show this by contradiction. Assume that there exists a $\theta_2 \in \mathcal{B}_\delta(\theta_1)$ such that $f(\theta_2) = 0$. Since $f(\theta_2) = 0$ can also be written as $f(\theta_2) = \bar{E}[(W(z,\theta_2) - W(z,\theta_1))\xi(t)]^2 = 0$, it follows by local informativity that $W(z,\theta_2) \equiv W(z,\theta_1)$; by local identifiability this implies $\theta_2 = \theta_1$.
ii) ONLY IF. We now prove the converse. Assume that θ_1 is an isolated minimum of $f(\theta)$. Then there exists $\delta > 0$ such that $\bar{E}[(W(z,\theta) - W(z,\theta_1))\xi(t)]^2 \neq 0$ for all $\theta \in \{\mathcal{B}_\delta(\theta_1) \setminus \theta_1\}$. Equivalently, if for some $\theta_2 \in \mathcal{B}_\delta(\theta_1)$ we have $\bar{E}[(W(z,\theta_2) - W(z,\theta_1))\xi(t)]^2 = 0$, it must imply $\theta_2 = \theta_1$. Since $\theta_2 = \theta_1$ implies $W(z,\theta_2) \equiv W(z,\theta_1)$, the latter is equivalent with local informativity and local identifiability at θ_1. ■

Theorem 4.1. Assume that $\mathcal{S} \in \mathcal{M}$, that is, $\exists\theta_0 : M(\theta_0) = \mathcal{S}$. Then θ_0 is an isolated global minimum of $V(\theta)$ if and only if $M(\theta)$ is locally identifiable at θ_0 and the experiment is locally informative at θ_0.
Proof The criterion $V(\theta)$ can be rewritten as

$$\begin{aligned} V(\theta) &= \bar{E}[y(t) - \hat{y}(t,\theta)]^2 \\ &= \bar{E}[y(t) - \hat{y}(t,\theta_0)]^2 + \bar{E}[\hat{y}(t,\theta_0) - \hat{y}(t,\theta)]^2 \\ &\quad + 2\bar{E}[y(t) - \hat{y}(t,\theta_0)][\hat{y}(t,\theta_0) - \hat{y}(t,\theta)] \\ &= V(\theta_0) + f(\theta) + 2\bar{E}[y(t) - \hat{y}(t,\theta_0)][\hat{y}(t,\theta_0) - \hat{y}(t,\theta)] \end{aligned} \quad (18)$$

When $\mathcal{S} \in \mathcal{M}$, $y(t) - \hat{y}(t,\theta_0)$ is white noise, so the last term is zero (see e.g. Ljung [1999]). It then follows that $f(\theta) > 0$ if and only if $V(\theta) > V(\theta_0)$. The result then follows from Lemma 4.1. ■

We have now established necessary and sufficient conditions for isolation of the global minimum (or minima) of the prediction error criterion when the system is in the model set: identifiability and informativity, both local, at the global minimum. A successful identification should fulfill such conditions, but these are not easy properties to verify per se.

Nonsingular information matrix
One objective in this paper is to check relevant properties

such as local identifiability and local informativity by analyzing the rank of the information matrix. It turns out that under quite general conditions nonsingularity of the information matrix is equivalent with local identifiability and local informativity.

From now on, the following class of model structures will be considered, with numerator and denominator polynomials which are affine in the parameter vector:

$$\mathcal{A} \triangleq \begin{cases} G(z,\theta) = \dfrac{n^g(z,\theta)}{d^g(z,\theta)} = \dfrac{ng_0(z^{-1}) + \theta^T ng_1(z^{-1})}{dg_0(z^{-1}) + \theta^T dg_1(z^{-1})} \\ H(z,\theta) = \dfrac{n^h(z,\theta)}{d^h(z,\theta)} = \dfrac{nh_0(z^{-1}) + \theta^T nh_1(z^{-1})}{dh_0(z^{-1}) + \theta^T dh_1(z^{-1})} \end{cases} \quad (19)$$

where $\theta \in D_\theta \subseteq \Re^d$; here ng_0, dg_0, nh_0, dh_0 represent polynomials in z^{-1}, while ng_1, dg_1, nh_1, dh_1, are d-vectors of polynomials in z^{-1}. These polynomials satisfy the constraints $nh_1(\infty) = 0$, $dh_1(\infty) = 0$, and $nh_0(\infty) = 1$, $dh_0(\infty) = 1$, so that $H(\infty, \theta) = 1$. The class of model structures of the form (19) for all possible degrees of the polynomials will be denoted class $\mathcal{A}$ from now on.

This class of model structures encompasses all "classical" model structures (ARMAX, ARX, BJ, OE). The "classical" model structures are obtained from (19) by setting $ng_0 = 0$, $nh_0 = dg_0 = dh_0 = 1$ and the vectors as zeros and powers of z^{-1}. For an ARMAX model structure, for example:

$$dg_1(z^{-1}) = \left[z^{-1} \ldots z^{-n_a} \mid 0 \ldots 0 \mid 0 \ldots 0\right]^T$$
$$ng_1(z^{-1}) = \left[0 \ldots 0 \mid z^{-1} \ldots z^{-n_b} \mid 0 \ldots 0\right]^T$$
$$nh_1(z^{-1}) = \left[0 \ldots 0 \mid 0 \ldots 0 \mid z^{-1} \ldots z^{-n_c}\right]^T$$

with $dh_1(z^{-1}) = dg_1(z^{-1})$ and $n_a + n_b + n_c = d$.

But the class of rational model structures (19) is much larger than just the classical model structures. It allows, for example, to naturally include prior knowledge of the system by considering model structures in which some polynomial coefficients are known, or are zero, or are repeated. The predictor in Example 5.1 (in the next Section) is an example: it "looks like" an output error model where two parameters are known is known (the z^{-3} coefficient in the numerator is equal to one and the z^{-1} coefficient in the denominator is equal to zero) and two parameters are known to have the same value, thus becoming the single parameter a in the model set (24).

Theorem 4.2. Let the model structure belong to class $\mathcal{A}$. Then at any given θ_1, $I(\theta_1) > 0$ if and only if $M(\theta)$ is locally identifiable at θ_1 and the data are locally informative at θ_1.

Proof Nonsingularity of the information matrix at θ_1 is equivalent with:

$$\bar{E}[\alpha^T \psi(t, \theta_1)]^2 = 0 \Longleftrightarrow \alpha = 0 \quad (20)$$

where $\alpha \in \Re^d$ and $\psi(t, \theta_1) = \frac{\partial \hat{y}(t,\theta)}{\partial \theta} \mid_{\theta_1}$. On the other hand, local informativity and local identifiability at θ_1 are, by definition, described by

$$\bar{E}[\hat{y}(t,\theta) - \hat{y}(t,\theta_1)]^2 = 0 \Longleftrightarrow \theta = \theta_1 \quad (21)$$

We show in the Appendix, through simple yet lengthy calculations, that for the model class $\mathcal{A}$ conditions (20) and (21) are equivalent. ■

An important observation is that Theorem 4.2 does not require that the system belongs to the model set; this result holds even when Assumption 1 is violated.

The following Theorem, which follows directly from the equivalences in Theorems 4.1 and 4.2, summarizes the findings presented so far.

Theorem 4.3. Assume that $\mathcal{S} \in \mathcal{M}$, that is, $\exists \theta_0 : M(\theta_0) = \mathcal{S}$ and that the model structure belongs to the class $\mathcal{A}$. Then the following three properties are equivalent:

(1) θ_0 is an isolated global minimum of $V(\theta)$;
(2) $I(\theta_0) > 0$;
(3) At θ_0 the model structure is locally identifiable and the data are locally informative.

■

Thus the user can choose his/her model structure and experimental conditions in such a way as to guarantee the nonsingularity of the information matrix, which in turn will guarantee isolation of the global minimum. The next Section shows how to do it and gives an illustrative example.

5. ACHIEVING A NONSINGULAR INFORMATION MATRIX

In Gevers et al. [2009b] necessary and sufficient conditions on the experiment were obtained that guarantee a full rank information matrix. These conditions on the signal richness were specialized to ARMAX and Box-Jenkins model structures for both open-loop and closed-loop identification. In closed-loop identification, a simplifying assumption was made on the absence of pole-zero cancellations between the closed-loop poles and the zeroes of the noise model. For closed loop identification of ARMAX models Shardt and Huang [2011] derived sufficient conditions that do not require this simplifying assumption.

In order to keep the presentation brief, in this section only the particular case of open loop identification for the class of models $H(z, \theta) = 1$ is considered. Other cases are covered in Gevers et al. [2009b] and can be treated similarly. Then the information matrix is given by

$$I(\theta) = \bar{E}[\psi(t,\theta)\psi(t,\theta)^T].$$

where

$$\psi(t,\theta) = V_u(z,\theta)u(t) = \nabla_\theta G(z,\theta)u(t), \quad (22)$$

Following the same procedure as in Gevers et al. [2009b], the vector $V_u(z, \theta)$ can be uniquely decomposed as

$$V_u(z,\theta) = \frac{z^{-m}}{d(z^{-1},\theta)} R(\theta) \mathcal{B}_{0,k-1}(z^{-1}) \quad (23)$$

where $d(z^{-1}) = 1 + d_1 z^{-1} + \ldots + d_p z^{-p}$, with $d_p \neq 0$, $R(\theta) \in \Re^{d \times k}$ is a matrix of real coefficients, m is a possible common delay in all elements of the numerator of $V_u(z^{-1})$ and

$$\mathcal{B}_{0,k-1}(z^{-1}) = \begin{bmatrix} 1 \\ z^{-1} \\ \ldots \\ z^{-(k-1)} \end{bmatrix}$$

With this formulation, it is clear that $V_u(z, \theta)$ has full row rank at some θ_1 only if $R(\theta_1)$ has full row rank, which requires that $k \geq d$. If the model structure is locally identifiable, the input $u(t)$ must still be rich enough to make the matrix $I(\theta)$ nonsingular at $\theta = \theta_1$. The richness of a signal is defined as follows.

Definition 5. A quasistationary scalar signal $u(t)$ is called sufficiently rich of order p (denoted SRp) if the frequency support of its spectrum has at least p nonzero components. It is called sufficiently rich of order exactly p (denoted SREp) if its frequency support has exactly p nonzero components. ∎

Consider now any θ at which $R(\theta)$ has full row rank. It then follows from Theorems 5.1 and 5.2 of Gevers et al. [2009b], adapted to the present case, that the necessary and sufficient richness conditions on $u(t)$ for the information matrix to be nonsingular are as follows:

(1) if $u(t)$ is not SRd then the information matrix is singular;
(2) if $u(t)$ is SRk then the information matrix is nonsingular;
(3) if $u(t)$ is SRd but not SRk, then the information matrix is nonsingular for almost all such $u(t)$; singularity of the information matrix may occur at specific values of θ that depend on the frequencies in the support of $u(t)$.

The following example illustrates these concepts and calculations. It also serves as an example where total informativity can not be achieved, but local informativity is achieved at almost every θ.

Example 5.1. Consider the following predictor:

$$\hat{y}(t,\theta) = \frac{az^{-1} + bz^{-2} + z^{-3}}{1 + az^{-2}} u(t) \tag{24}$$

with $\theta = [a \quad b]^T$. The pseudoregressor $\psi(t,\theta) = \frac{\partial \hat{y}(t,\theta)}{\partial \theta}$ and its unique decomposition in the form (23) are as follows:

$$\psi(t,\theta) = \begin{bmatrix} \frac{z^{-1}}{1+az^{-2}} - \frac{z^{-2}(az^{-1}+bz^{-2}+z^{-3})}{(1+az^{-2})^2} \\ \\ \frac{z^{-2}}{1+az^{-2}} \end{bmatrix} u(t)$$

$$= \frac{z^{-1}}{(1+az^{-2})^2} R(\theta) \mathcal{B}_{0,4}(z^{-1})$$

with

$$R(\theta) = \begin{bmatrix} 1 & 0 & 0 & -b & -1 \\ 0 & 1 & 0 & a & 0 \end{bmatrix}$$

The matrix $R(\theta)$ has rank 2 for all θ. Moreover, this is a 2×5 matrix, and thus the conditions 1, 2 and 3 above are:

- if $u(t)$ is not SR2 then the information matrix is singular;
- if $u(t)$ is SR5 then the information matrix is nonsingular;
- if $u(t)$ is SRE2, 3 or 4, then the rank of the information matrix depends on θ and on the frequencies in the support of $u(t)$.

To further detail the last condition, consider, for example, the SRE2 signal $u(t) = \sin(\omega t)$, and define

$$F(e^{j\omega}) \triangleq \begin{pmatrix} 1 & e^{j\omega} & e^{2j\omega} & e^{3j\omega} & e^{4j\omega} \\ 1 & e^{-j\omega} & e^{-2j\omega} & e^{-3j\omega} & e^{-4j\omega} \end{pmatrix}^T .$$

The information matrix is given by Gevers et al. [2009b]

$$I(\theta) = R(\theta)F(e^{j\omega})F^H(e^{j\omega})R^T(\theta)$$

and the square root matrix $R(\theta)F(e^{j\omega})$ can be calculated as

$$R(\theta)F(e^{j\omega}) = \begin{bmatrix} 1 - be^{j3\omega} - e^{j4\omega} & 1 - be^{-j3\omega} - e^{-j4\omega} \\ e^{j\omega} + ae^{j3\omega} & e^{-j\omega} + ae^{-j3\omega} \end{bmatrix}$$

The singularity of the information matrix can be determined by looking at the determinant of its square root, which is given by:

$$\det(RF) = -2j[a(\sin\omega + \sin 3\omega) + b\sin 2\omega + \sin\omega + \sin 3\omega]$$

For any fixed frequency ω, there will be an affine set in the parameter space defined by

$$U = \{ \begin{bmatrix} a \\ b \end{bmatrix} \in \Re^2 : \\ a(\sin\omega + \sin 3\omega) + b(\sin 2\omega) + (\sin\omega + \sin 3\omega) = 0 \}$$

at which the information matrix will be singular.

The information matrix is nonsingular for all $\theta \notin U$. So, according to Theorem 4.3, the global minimum of $V(\theta)$ is isolated as long as it is not in this set U. Whether the global minimum lies in this set or not depends on the real system, of course, but nothing of what has been said before depends on the real system. Since the model structure is identifiable for all θ, we know that the singularity of the information matrix for $\theta \in U$ is due to lack of informativity provided by the experiment proposed, which consists of applying a single sinusoid to the system's input. Note that each frequency ω defines a different set U. ∎

We thus have **computable necessary and sufficient** conditions for isolation of the global minimum. They are computable because the user can select an identifiable model structure and, having selected his/her model structure, he/she knows from Gevers et al. [2009b] exactly how rich a signal must be to make the information matrix nonsingular. These conditions are valid for general model structures of the class $\mathcal{A}$, which includes all classical model structures (ARMAX, BJ), and often do not even depend on the real system (like in the example above). Recall that, for the sake of brevity, in this Section we have treated only a particular case: open-loop identification with $H(z,\theta) = 1$. Different identification scenarios can all be treated in the same way, using the tools developed in Gevers et al. [2009b] for achieving a nonsingular information matrix and thus obtaining isolation of the global minimum, with conclusions quite similar to the ones above.

In the next Section we will see under which conditions the isolation of the global minimum also implies its uniqueness.

6. UNIQUENESS

In Section 4 we have given conditions for isolation of the global minimum of the PE criterion $V(\theta)$. But how about uniqueness? Assuming that there exists an isolated global minimum, can there exist other global minima? In this section we establish conditions under which the existence of different isolated global minima is ruled out; then isolation becomes equivalent with uniqueness under these conditions.

Define the set of minimizers Θ^* of the cost function $V(\theta)$ as in (6). We show that, with the class $\mathcal{A}$ of model structures defined by (19), in most cases this set has an affine structure, which rules out the existence of isolated global minima. The affinity comes from the following basic result (Garatti et al. [2004], Gevers et al. [2009a]).

Lemma 6.1. Consider a given transfer function $G_0(z)$, a class of transfer functions of the form (19) with $\theta \in \Re^d$, and a set of distinct frequencies $\Omega = \{\omega_1, \omega_2, \ldots, \omega_q\}$. Define $\Theta_G^+ \triangleq \{\theta : G(e^{j\omega}, \theta) = G_0(e^{j\omega}) \quad \forall \omega \in \Omega\}$. Then the set Θ_G^+ is affine.

Proof: The set Θ_G^+ is, by definition, the solution set of the q equations:

$$\frac{ng_0(e^{j\omega_i}) + \theta^T ng_1(e^{j\omega_i})}{dg_0(e^{j\omega_i}) + \theta^T dg_1(e^{j\omega_i})} = G_0(e^{j\omega_i}), i = 1, ..., q. \quad (25)$$

Define $G_0(e^{j\omega_i}) \triangleq \frac{gn_0(e^{j\omega_i})}{gd_0(e^{j\omega_i})}$, and

$$a_i(e^{j\omega_i}) \triangleq \left[gd_0(e^{j\omega_i}) ng_1(e^{j\omega_i}) - gn_0(e^{j\omega_i}) dg_1(e^{j\omega_i})\right]^T \quad (26)$$

Note that $a_i(e^{j\omega_i})$ is a row d-vector of polynomials. Now rewrite (25) as

$$a_i(e^{j\omega_i})^T \theta = b_i, \quad i = 1, ..., q, \quad (27)$$

where we have defined $b_i = gn_0(e^{j\omega_i}) dg_0(e^{j\omega_i}) - gd_0(e^{j\omega_i}) ng_0(e^{j\omega_i})$. Rewriting (27) for the q frequencies in Ω, and stacking these q equations, we have

$$A\theta = b \quad (28)$$

where A 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. ■

We have considered a discrete support for clarity of presentation, but Lemma 6.1 is valid for continuous supports Ω as well, as shown in Garatti et al. [2004]. The set Θ_G^+, being the solution of a system of linear equations, may well be empty. This will be so generically when $q > d$, because then the system contains more equations than variables. However, if $\mathcal{S} \in \mathcal{M}$, then the set Θ_G^+ will never be empty.

6.1 Open-loop identification

Let now the identification be performed in open loop and let Ω be the support of the input $u(t)$, with spectrum $\Phi_u(\omega)$. In open-loop identification, the criterion $V(\theta)$ defined in (7) can be written as

$$V(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|\, G(e^{j\omega}, \theta) - G_0(e^{j\omega}) \,|^2}{|\, H(e^{j\omega}, \theta) \,|^2} \Phi_u(\omega) d\omega$$

$$+ \sigma_e^2 \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|\, H_0(e^{j\omega}) \,|^2}{|\, H(e^{j\omega}, \theta) \,|^2} d\omega \quad (29)$$

Also, define the sets

$$\Theta_H^+ \triangleq \{\theta : H(e^{j\omega}, \theta) = H_0(e^{j\omega}) \quad \forall \omega \in (-\pi, \pi]\}$$
$$\Theta^+ \triangleq \Theta_G^+ \cap \Theta_H^+$$

where Θ_G^+ has been defined in Lemma 6.1. From Lemma 6.1, both sets Θ_G^+ and Θ_H^+ are affine, hence so is Θ^+.

The set Θ^+ is of interest because, when $\mathcal{S} \in \mathcal{M}$, $\Theta^+ = \Theta^*$, where Θ^* was defined in (6). Indeed it follows from (29) that the minimum value of $V(\theta)$ is achieved at a given θ_1 if and only if $\theta_1 \in \Theta^+$, since this yields $V(\theta_1) = \sigma_e^2$, which is the smallest possible value of the cost function $V(\theta)$. Affinity of the set Θ^* is then established, thus ruling out the possibility of distinct isolated global minima: when the minimizers' set is affine, isolation of a global minimum is equivalent to its uniqueness. We formalize this reasoning in the following Theorem.

Theorem 6.1. Let the model structure belong to the class $\mathcal{A}$ and assume that $\mathcal{S} \in \mathcal{M}$, that is, $\exists \theta_0 : M(\theta_0) = \mathcal{S}$. Suppose that the identification is performed in open loop. Then the following three conditions are equivalent:

(1) θ_0 is the unique global minimum of $V(\theta)$;
(2) $I(\theta_0) > 0$;
(3) At θ_0 the model structure is locally identifiable and the data are locally informative.

■

Proof: Under the conditions of the Theorem the minimizers' set is affine, thus isolation of a global minimum is equivalent with its uniqueness. Equivalence between (1) and (3) then follows from Theorem 4.1, while equivalence with (2) follows from Theorem 4.2. ■

Previously known necessary and sufficient conditions for uniqueness of the global minimum of the PEI criterion $V(\theta)$ covered much more restricted classes of problems. Conditions for ARX models are standard textbook material, whereas conditions for ARMA and OE model structures in open-loop are given in Åström and Söderström [1974], Söderström [1975] respectively.

6.2 Closed-loop identification

The same reasoning applies to closed-loop identification, that is: isolation+affinity implies uniqueness of the global minimum. The sole difference to the open-loop case lies in the conditions that guarantee affinity of the minimizer set, which in the closed-loop case is not a generic property. Conditions for affinity are given in the following Lemma.

Lemma 6.2. Consider closed-loop identification with a model structure in the form $\mathcal{A}$, the control law (2) with a rational controller $K(z) = \frac{n^c(z)}{d^c(z)}$, where $n^c(z)$ and $d^c(z)$ are fixed polynomials. Assume that $\mathcal{S} \in \mathcal{M}$. Then the minimizer set Θ^* of $V(\theta)$ is affine in θ if the following equation is affine in θ:

$$n^h(z,\theta)d^g(z,\theta) = d^h(z,\theta)[d^c(z)d^g(z,\theta) + n^c(z)n^g(z,\theta)] \quad (30)$$

where n^g, d^g, n^h, d^h are defined in (19).

Proof: The proof is given in the Appendix. ■

The following corollary enumerates a number of cases where the minimizer set Θ^* is affine in θ.

Corollary 6.1. Under the conditions of Lemma 6.2, the set Θ^* defined in (6) is affine in θ if either one of the following conditions is satisfied:

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

Proof: It is easy to see that when any of these conditions is satisfied, equation (30) becomes affine in θ. ■

The family of model structures satisfying the conditions of Corollary 6.1 encompasses (but is not limited to) the "classical" model structures ARX, ARMAX and OE. Interestingly enough, a Box-Jenkins model structure does not satisfy these conditions. This was first observed in Garatti et al. [2004], where an example has been given where the closed-loop PEI criterion for a Box-Jenkins model possesses two isolated global minima.

Theorem 6.2. Let the model structure belong to the class $\mathcal{A}$ and assume that $\mathcal{S} \in \mathcal{M}$, that is, $\exists\theta_0 : M(\theta_0) = \mathcal{S}$. Suppose that the identification is performed in closed loop. Suppose further that the model structure satisfies at least one of the four conditions of Corollary 6.1. Then the following three conditions are equivalent:

(1) θ_0 is the unique global minimum of $V(\theta)$;
(2) $I(\theta_0) > 0$;
(3) at θ_0 the model structure is locally identifiable and the data are locally informative.

Proof: The same as in Theorem 6.1. ■

7. CONCLUSIONS

In this paper we have given necessary and sufficient conditions for the isolation of the global minima for a class of model structures which encompasses all "traditional" model structures - ARMAX, BJ and the like. These necessary and sufficient conditions are given in terms of the model structure and the richness of the input signal, which are exactly the choices that the user must make in setting up his/her identification. None of these conditions require any knowledge about the real system.

Moreover, our new necessary and sufficient conditions can be verified and enforced efficiently because they are equivalent with the nonsingularity of the information matrix. Obtaining these necessary and sufficient conditions presented in this paper has required abandoning the traditional concept of informative data, and replacing it by the new concept of local informativity at a parameter value. Previously known conditions were conservative and/or restricted to much narrower families of model structures.

It has been shown that for open-loop identification it is impossible to have distinct isolated global minima, which amounts to an equivalence between isolation and uniqueness of the global minimum. We have thus shown that nonsingularity of the information matrix (a local property) becomes a necessary and sufficient condition for uniqueness of the global minimum (a global property) in open-loop identification for a wide class of model structures. The existence of distinct isolated global minima can also be ruled out in closed-loop identification for a wide variety of model structures in the family considered, but not all. For these model structures (which include ARMAX but not Box-Jenkins) nonsingularity of the information matrix is also a necessary and sufficient condition for uniqueness of the global minimum in closed-loop identification.

Appendix

Proof of Theorem 4.2 To complete the proof, we must show that for the class $\mathcal{A}$ of model structures conditions (20) and (21) are equivalent. We will show this separately for open-loop and for closed-loop identification.

Open-loop identification

Start with local informativity and identifiability - condition (21), reproduced below:

$$F(\theta,\theta_1) \triangleq \bar{E}[\hat{y}(t,\theta) - \hat{y}(t,\theta_1)]^2 = 0 \Longleftrightarrow \theta = \theta_1. \quad (31)$$

The bracketed expression is given by:

$$\hat{y}(t,\theta) - \hat{y}(t,\theta_1) = \left[\frac{G(z,\theta) - G_0(z)}{H(z,\theta)} - \frac{G(z,\theta_1) - G_0(z)}{H(z,\theta_1)}\right] u(t) + \left[\frac{H_0(z)}{H(z,\theta_1)} - \frac{H_0(z)}{H(z,\theta)}\right] e(t) \quad (32)$$

Because $u(t)$ and $e(t)$ are independent we can write:

$$F(\theta,\theta_1) = F_u(\theta,\theta_1) + F_e(\theta,\theta_1) \quad (33)$$

with the obvious definitions for $F_u(\theta,\theta_1)$ and $F_e(\theta,\theta_1)$. Because $F_u(\theta,\theta_1)$ and $F_e(\theta,\theta_1)$ are both nonnegative, $F(\theta,\theta_1) = 0$ if and only if $F_u(\theta,\theta_1) = F_e(\theta,\theta_1) = 0$. Since $e(t)$ is white noise, $F_e(\theta,\theta_1) = 0$ if and only if $H(z,\theta_1) = H(z,\theta)$ which, when substituted in the first term, yields

$$F_u(\theta,\theta_1) = \bar{E}\{[H^{-1}(z,\theta_1)[G(z,\theta) - G(z,\theta_1)]u(t)\}^2.$$

This expression can in turn be rewritten as

$$F_u(\theta,\theta_1) = \bar{E}\{\frac{d^h(z,\theta_1)}{n^h(z,\theta_1)d^g(z,\theta_1)d^g(z,\theta)} [n^g(z,\theta)d^g(z,\theta_1) - d^g(z,\theta)n^g(z,\theta_1)]u(t)\}^2 \quad (34)$$

because $G(\theta) = \frac{n^g(z,\theta)}{d^g(z,\theta)}$ and $H(\theta) = \frac{n^h(z,\theta)}{d^h(z,\theta)}$.

Now use the model structure (19) and define $\alpha = \theta - \theta_1$. Then

$$n^g(z,\theta) = n_{g0}(z^{-1}) + \theta^T n_{g1}(z^{-1})$$

$$= n_{g0}(z^{-1}) + (\theta_1 + \alpha)^T n_{g1}(z^{-1})$$
$$= n^g(z, \theta_1) + \alpha^T n_{g1}(z^{-1})$$

and similarly for $d^g(z,\theta)$, so that the bracketed term in (34) becomes

$$\begin{aligned} & n^g(z,\theta)d^g(z,\theta_1) - d^g(z,\theta)n^g(z,\theta_1) \\ & = [n^g(z,\theta_1) + \alpha^T n_{g1}(z^{-1})]d^g(z,\theta_1) \\ & -[d^g(z,\theta_1) + \alpha^T d_{g1}(z^{-1})]n^g(z,\theta_1) \\ & = \alpha^T[n_{g1}(z^{-1})d^g(z,\theta_1) - d_{g1}(z^{-1})n^g(z,\theta_1)] \end{aligned}$$

Finally, substituting this expression in (34) shows that the local identifiability and local informativity at θ_1 are defined by the following equivalence

$$\begin{aligned} & \bar{E}\{\alpha^T \frac{d^h(z,\theta_1)}{n^h(z,\theta_1)d^g(z,\theta_1)d^g(z,\theta)} \\ & [n_{g1}(z^{-1})d^g(z,\theta_1) - d_{g1}(z^{-1})n^g(z,\theta_1)]u(t)\}^2 = 0 \\ & \Longleftrightarrow \alpha = 0. \end{aligned} \tag{35}$$

Now turn to the information matrix. According to Gevers et al. [2009b], $I(\theta_1)$ is nonsingular if the pseudoregressor $\psi(t,\theta_1)$ is full-rank, that is if condition (20) is valid. It thus remains to show that (35) is equivalent with (20). For open-loop identification the pseudoregressor $\psi(t,\theta)$ can be written as (see Gevers et al. [2009b])

$$\psi(t,\theta) = V_u(z,\theta)u(t) + V_e(z,\theta)e(t)$$

with

$$\begin{aligned} V_u(z,\theta) &= \frac{1}{H^2(z,\theta)}[H(z,\theta)\nabla_\theta G(z,\theta) \\ & \quad -(G(z,\theta) - G_0(z))\nabla_\theta H(z,\theta)] \\ V_e(z,\theta) &= \frac{H_0(z)}{H^2(z,\theta)}\nabla_\theta H(z,\theta), \end{aligned}$$

where for any $W(z,\theta)$ we denote $\nabla_\theta W(z,\theta) \triangleq \frac{\partial W(z,\theta)}{\partial \theta}$.

Because $u(t)$ and $e(t)$ are independent, $\bar{E}[\alpha^T\psi(t,\theta_1)]^2 = 0$ is equivalent with $\bar{E}[\alpha^T V_e(z,\theta_1)e(t)]^2 = 0$ and $\bar{E}[\alpha^T V_u(z,\theta_1)u(t)]^2 = 0$. Since $e(t)$ is white noise, $\bar{E}[\alpha^T V_e(z,\theta_1)e(t)]^2 = 0$ is equivalent with $\alpha^T\nabla_\theta H(z,\theta)|_{\theta_1} = 0$ which, substituted in the other term, yields

$$\bar{E}[\alpha^T V_u(z,\theta_1)u(t)]^2 = \bar{E}[\alpha^T \frac{1}{H(z,\theta_1)}\nabla_\theta G(z,\theta)|_{\theta_1}u(t)]^2 = 0.$$

But

$$\nabla_\theta G(z,\theta) = \frac{1}{d^g(z,\theta)^2}[d^g(z,\theta)\nabla_\theta n^g(z,\theta) - n^g(z,\theta)\nabla_\theta d^g(z,\theta)]$$

and

$$\nabla_\theta n^g(z,\theta) = n_{g1}(z^{-1}) \qquad \nabla_\theta d^g(z,\theta) = d_{g1}(z^{-1})$$

Thus, condition (20) on the nonsingularity of the information matrix at θ_1 is equivalent with:

$$\begin{aligned} & \bar{E}[\alpha^T V_u(z,\theta_1)u(t)]^2 \\ & = \bar{E}\{\alpha^T \frac{d^h(z,\theta_1)}{n^h(z,\theta_1)d^g(z,\theta_1)^2} \\ & [d^g(z,\theta_1)n_{g1}(z,\theta_1) - n^g(z,\theta_1)d_{g1}(z,\theta_1)]u(t)\}^2 = 0 \\ & \Longleftrightarrow \alpha = 0 \end{aligned} \tag{36}$$

Comparing (35) and (36) we observe that the conditions on the model structure and on the data for local identifiability and local informativity are identical to those for nonsingularity of the information matrix, i.e. conditions (35) and (36) are equivalent. Actually the sole difference between (35) and (36) is the term $d^g(z,\theta_1)$ appearing in the denominator of (36) in lieu of $d^g(z,\theta)$ in (35), which has no bearing in making the expression equal to zero. ■

Closed-loop identification

In closed-loop, the predictor can be written as

$$\begin{aligned} \hat{y}(t,\theta) &= KS\left[\frac{G(\theta) - G_0}{H(\theta)} + G_0\right]r(t) \\ & +H_0 S\left[1 - \frac{1 + KG(\theta)}{H(\theta)}\right]e(t), \end{aligned} \tag{37}$$

where K is the controller and $S = \frac{1}{1+KG_0}$. We start again with condition (31). For closed-loop identification the bracketed expression becomes

$$\begin{aligned} \hat{y}(t,\theta) - \hat{y}(t,\theta_1) &= KS\left[\frac{G(\theta)-G_0}{H(\theta)} - \frac{G(\theta_1)-G_0}{H(\theta_1)}\right]r(t) \\ & +H_0 S\left[\frac{1+KG(\theta_1)}{H(\theta_1)} - \frac{1+KG(\theta)}{H(\theta)}\right]e(t) \end{aligned} \tag{38}$$

It follows from the independence of r and e and the whiteness of e that $\bar{E}\{\hat{y}(t,\theta) - \hat{y}(t,\theta_1)\}^2 = 0$ implies that the transfer function multiplying $e(t)$ in (38) is identically zero. This is equivalent with

$$H(\theta) - H(\theta_1) + K[H(\theta)G(\theta_1) - H(\theta_1)G(\theta)] \equiv 0. \tag{39}$$

$\bar{E}\{\hat{y}(t,\theta) - \hat{y}(t,\theta_1)\}^2 = 0$ also implies

$$\begin{aligned} & \bar{E}\{\frac{KS}{H(\theta_1)H(\theta)}[(G(\theta) - G_0)H(\theta_1) \\ & \qquad -(G(\theta_1) - G_0)H(\theta)]\,r(t)\}^2 = 0. \end{aligned}$$

Inserting (39) into this last expression yields, after some simple manipulations, that local infomativity and local identifiability is equivalent with the following condition:

$$\bar{E}\{\frac{1}{H(\theta_1)H(\theta)}[H(\theta) - H(\theta_1)]r(t)\}^2 = 0 \Longleftrightarrow \theta = \theta_1 \tag{40}$$

We now use the model structure (19) and define $\alpha = \theta - \theta_1$. Then, using the same manipulations as were done in open-loop identification, we find that condition (40) is equivalent with the following equivalence:

$$\begin{aligned} & \bar{E}\{\frac{\alpha^T}{n^h(z,\theta_1)n^h(z,\theta)} \\ & [n_{h1}(z^{-1})d^h(z,\theta_1) - d_{h1}(z^{-1})n^h(z,\theta_1)]r(t)\}^2 = 0 \\ & \Longleftrightarrow \alpha = 0 \end{aligned} \tag{41}$$

We now turn to the information matrix. In closed-loop the pseudoregressor $\psi(t,\theta)$ can be written (see Gevers et al. [2009b]):

$$\psi(t,\theta) = V_r(z,\theta)r(t) + V_e(z,\theta)e(t) \tag{42}$$

with

$$V_r(z,\theta) = KS\{\frac{1}{H^2(\theta)}\left[H(\theta)\nabla_\theta G(\theta) + (G_0 - G(\theta))\nabla_\theta H(\theta)\right]\}$$

$$V_e(z,\theta) = \frac{H_0 S}{H^2(\theta)}\{\nabla_\theta H(\theta) - K\left[H(\theta)\nabla_\theta G(\theta) - G(\theta)\nabla_\theta H(\theta)\right]\}$$

Because $r(t)$ and $e(t)$ are independent, $\bar{E}[\alpha^T\psi(t,\theta_1)]^2 = 0$ is equivalent with $\bar{E}[\alpha^T V_e(z,\theta_1)e(t)]^2 = 0$ and $\bar{E}[\alpha^T V_r(z,\theta_1)r(t)]^2 = 0$. Consider first the last expression; it holds if and only if

$$\alpha^T\{\nabla_\theta H(\theta) - K\left[H(\theta)\nabla_\theta G(\theta) - G(\theta)\nabla_\theta H(\theta)\right]\} \equiv 0.$$

Substituting this into $\bar{E}[\alpha^T V_r(z,\theta_1)r(t)]^2 = 0$ yields

$$\bar{E}\{\frac{\alpha^T}{H^2(\theta_1)}[\nabla_\theta H(\theta_1)]r(t)\}^2 = 0 \tag{43}$$

With the model structure (19) this last condition is equivalent with

$$\bar{E}\{\frac{\alpha^T}{n^h(z,\theta_1)^2}[n_{h1}(z^{-1})d^h(z,\theta_1) - d_{h1}(z^{-1})n^h(z,\theta_1)]r(t)\}^2 = 0$$

Thus, in closed loop the information matrix is nonsingular at θ_1 if and only if the following equivalence holds:

$$\bar{E}\{\frac{\alpha^T}{n^h(z,\theta_1)^2}[n_{h1}(z^{-1})d^h(z,\theta_1) - d_{h1}(z^{-1})n^h(z,\theta_1)]r(t)\}^2 = 0 \Longleftrightarrow \alpha = 0 \tag{44}$$

We observe that the condition (44) on the model structure and on $r(t)$ is equivalent with condition (41), which concludes the proof. ■

Proof of Theorem 6.2: This theorem is an extension to the model structure (19) 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 $\Theta^* = \Theta^g \cap \Theta^s$ with

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

where Ω_r is the support of the external excitation signal r, and

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

Θ^s is the set of solutions of (omitting dependence on ω):

$$n^h(\theta_0)[d^c d^g(\theta) + n^c n^g(\theta)]d^h(\theta)d^c d^g(\theta_0) = n^h(\theta)[d^c d^g(\theta_0) + n^c n^g(\theta_0)]d^h(\theta_0)d^c d^g(\theta), \tag{45}$$

which must hold at all $\omega \in [-\pi,\pi)$. In general the solution set is an ellipsoid, which is not affine since (45) is quadratic in θ, and hence different minimizers can be isolated from each other. We note that the factors on both sides of (45) that do not depend on θ have no influence on the affine (in θ) character of this equation. Thus, the solution of (45) is affine in θ if and only if (30) is affine in θ. Since the intersection of affine subspaces is an affine subpsace, the intersection $\Theta^g \cap \Theta^s$ is then also affine. ■

REFERENCES

B. D. O. Anderson and M. Gevers. Identifiability of linear stochastic systems operating under linear feedback. *Automatica, 18*, 2:195–213, 1982.

K. J. Åström and T. Söderström. Uniqueness of the Maximum Likelihood estimates of the parameters of an arma model. *IEEE Transactions on Automatic Control*, 19(6):769–773, December 1974.

A.S. Bazanella, M. Gevers, and L. Mišković. Closed-loop identification of MIMO systems: a new look at identifiability and experiment design. *European Journal of Control*, 16:228–239, May 2010.

R. Bowden. The theory of parametric identification. *Econometrica*, 41(6):1069–1074, November 1973.

M. Deistler. General structure and parametrization of ARMA and state-space systems and its relation to statistical problems. *Handbook of Statistics*, pages 257–277, 1989.

S. Garatti, M.C. Campi, and S. Bittanti. Assessing the quality of identified models through the asymptotic theory - when is the result reliable? *Automatica*, 40 (8):1319–1332, 2004.

M. Gevers, A.S. Bazanella, and X. Bombois. Connecting informative experiments, the information matrix and the minima of a Prediction Error Identification criterion. In *Proc. 15th IFAC Symposium on System Identification (SYSID2009)*, pages 675–680, Saint-Malo, France, 2009a.

M. Gevers, A.S. Bazanella, X. Bombois, and L. Mišković. Identification and the information matrix: how to get just sufficiently rich? *IEEE Transactions on Automatic Control*, 54(12):2828–2840, December 2009b.

K. Glover and J.C. Willems. Parametrizations of linear dynamical systems: canonical forms and identifiability. *IEEE Transactions on Automatic Control*, 19(6):640–646, 1974.

I. Gustavsson, L. Ljung, and T. Söderström. Identification of processes in closed loop - identifiability and accuracy aspects. *Automatica*, 13:59–75, 1977.

L. Ljung. On consistency and identifiability. *Mathematical Programming Study*, 5:169–190, 1976.

L. Ljung. *System Identification: Theory for the User, 2nd Edition.* Prentice-Hall, Englewood Cliffs, NJ, 1999.

L. Ljung and T. Glad. On global identifiability for arbitrary model parametrizations. *Automatica*, 30(2): 265–276, 1994.

T. S. Ng, G. C. Goodwin, and B. D. O. Anderson. Identifiability of MIMO linear dynamic systems operating in closed loop. *Automatica*, 13:477–485, 1977.

J. Ritt. *Differential Algebra*. American Mathematical Society, Providence, R.I., 1950.

T. Rothenberg. Identification in parametric models. *Econometrica*, 39:577–591, May 1971.

Y. A. W. Shardt and B. Huang. Closed-loop identification condition for armax models using routine operating data. *Automatica*, 47(7):1534–1537, 2011. URL `www.scopus.com`.

T. Söderström. On the uniqueness of Maximum Likelihood identification. *Automatica*, 11(2):193–197, March 1975.

V. Solo. Topics in advanced time series analysis. In Guido de Pino and Rolando Rebolledo, editors, *Lectures in Probability and Statistics*, volume 1215, pages 165–328, Berlin New York, 1986. Springer-Verlag.

P.M.J. Van den Hof, D.K. de Vries, and P. Schoen. Delay structure conditions for identifiability of closed loop systems. *Automatica*, 28(5):1047–1050, May 1992.