

---

# Identifiability and informative experiments in open and closed-loop identification\*

Michel Gevers<sup>1</sup>, Alexandre Sanfelice Bazanella<sup>1,2</sup>, and Ljubiša Mišković<sup>1</sup>

<sup>1</sup> CESAME, Université Catholique de Louvain, Louvain-la-Neuve, Belgium  
{bazanella,gevers,miskovic}@csam.ucl.ac.be

<sup>2</sup> On leave from Electrical Engineering Department, Universidade Federal do Rio Grande do Sul, Porto Alegre-RS, Brazil bazanella@ece.ufrgs.br

*To Giorgio, who to our great surprise has not solved the rather fundamental problems raised in this chapter, we hope to offer some food for thought.*

## 1 Introduction

This chapter takes a new look at the concept of identifiability and of informative experiments for linear time-invariant systems, both in open-loop and in closed-loop identification. Some readers might think that everything has been said and written about these concepts, which were much studied all through the 1970's. We shared the same view . . . until recently. The motivation for our renewed interest into these very fundamental questions is the recent surge of interest in the question of experiment design, itself triggered by the new concept of *least costly identification experiment for robust control* [1, 2, 3, 4]. Briefly speaking, least costly experiment design for robust control refers to achieving a prescribed accuracy at the lowest possible price, which is typically measured in terms of the duration of the identification experiment, the perturbation induced by the excitation signal, or any combination of these. In this context, questions like the following become relevant:

1. what is the smallest amount of external excitation that is required to achieve identifiability (or to achieve a given accuracy level)?
2. assuming that the system operates in closed-loop, is the noise excitation sufficient to guarantee identifiability?

---

\* This chapter presents research results of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office. The first author is also partially supported by the Brazilian Ministry of Education through CAPES.

3. if noise excitation is not sufficient to guarantee identifiability in a closed-loop experiment, then how much additional reference excitation is required?
4. assuming that excitation can be applied at different entry points of a multi-input system operating in closed loop, is it necessary to excite each input to achieve identifiability (or to achieve a given accuracy level)?

Sufficient conditions for identifiability using noise excitation only (question 2) have been given, under different sets of assumptions, in [5, 6, 2]. The key condition for identifiability using noise excitation only is in terms of the complexity of the feedback controller; this complexity condition relates the controllability (or observability) indices of the controller to the controllability (or observability) indices of the plant. Question 4 has been addressed in [7] where it is shown that, when identifiability cannot be achieved using noise excitation only, this does not imply that all reference inputs must be excited.

In attempting to address questions 1 and 3 above, we discovered to our surprise that these questions do not seem to have been addressed (or at least solved) before. As is well-known, besides the choice of an identifiable model structure, the key ingredient to achieve identifiability is the informativity of the experiment. In open-loop identification, and in all closed-loop identification experiments where the noise excitation by itself does not make the experiment informative, the informativity is achieved by applying a sufficiently rich input signal. The degree of richness of a signal is a concept that is precisely defined; a signal is said to be sufficiently rich of degree  $n$  if its spectral density is nonzero in at least  $n$  distinct frequency points in the interval  $[-\pi, \pi]$ . But whereas the scientific literature abounds with sufficient conditions on input signal richness, there appear to be no result on the smallest possible degree of richness that delivers informative data in a given identification setup. In other words, necessary conditions on input richness that will guarantee an informative experiment are strangely lacking.

The purpose of this contribution is to attempt to find the smallest possible degree of richness of the excitation signal that makes an experiment informative with respect to a chosen model structure, both in open-loop and in closed-loop identification. More precisely, we address the following two questions:

- assuming open-loop identification, what is the smallest degree of input signal richness that is necessary to achieve an informative experiment with respect to a chosen model structure?
- assuming closed-loop identification with a controller that is not sufficiently complex to yield identifiability using noise excitation only, what is then the smallest degree of reference signal excitation that is necessary to achieve an informative experiment with respect to a chosen model structure?

In addressing these questions, we shall introduce a new framework that allows one to handle in the same way the range space spanned by station-

ary stochastic vectors and that spanned by vectors of rational transfer functions. We believe this new framework to be a convenient tool to establish results on the transfer of excitation from input signals to regression vectors through linear time-invariant filters. Our analysis and results will be established for single-input single-output (SISO) systems, but the framework we develop lends itself easily to extensions to multi-input multi-output (MIMO) systems.

The chapter is organized as follows. In Section 2 we set up the notations and the key tools of the prediction error identification framework. In Section 3, we recall the basic concepts of identifiability and informative experiments. The body of our results are in Section 4 which focuses on the key role of the information matrix. Our main results concern the derivation of necessary and sufficient conditions on the input signal that make a regressor persistently exciting. This allows us to formulate necessary and sufficient conditions on input signal richness that makes the information matrix have full rank. Finally, in Section 5 we apply these results to some widely utilized model structures.

## 2 The prediction error identification setup

Consider the identification of a linear time-invariant discrete-time single-input-single-output process

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

In (1)  $z$  is the forward-shift operator,  $G_0(z)$  and  $H_0(z)$  are the process transfer functions,  $u(t)$  is the control input and  $e(t)$  is white noise with variance  $\sigma_e^2$ . Both transfer functions,  $G_0(z)$  and  $H_0(z)$ , are rational and causal (proper); furthermore,  $H_0(\infty) = 1$ , that is the impulse response  $h(t)$  of the filter  $H_0(z)$  satisfies  $h(0) = 1$ .

This true system may be under feedback control with a causal rational stabilizing controller  $K(z)$ :

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

The system (1) is identified using a model structure parametrized by a vector  $\theta \in \mathcal{R}^d$ :

$$M(\theta) : y(t) = G(z, \theta)u(t) + H(z, \theta)\varepsilon(t) \quad (3)$$

The set of models  $M(\theta)$ , for all  $\theta$  in some set  $D_\theta \in \mathcal{R}^d$ , defines the model set  $\mathcal{M}$ :  $\mathcal{M} \triangleq \{M(\theta) \mid \theta \in D_\theta\}$ . The true system is said to belong to this model set,  $\mathcal{S} \in \mathcal{M}$ , if there is a  $\theta_0$  such that  $M(\theta_0) = \mathcal{S}$ . In a prediction error identification framework, a model  $[G(z, \theta) \ H(z, \theta)]$  uniquely defines the one-step-ahead predictor of  $y(t)$  given all input/output data up to time  $t$ :

$$\hat{y}(t|t-1, \theta) = W_u(z, \theta)u(t) + W_y(z, \theta)y(t), \quad (4)$$

where  $W_u(z, \theta)$  and  $W_y(z, \theta)$  are stable filters obtained from the model  $[G(z, \theta) \ H(z, \theta)]$  as follows:

$$W_u(z, \theta) = H^{-1}(z, \theta)G(z, \theta), \quad W_y(z, \theta) = [I - H^{-1}(z, \theta)]. \quad (5)$$

Since there is a 1 – 1 correspondance between  $[G(z, \theta), \ H(z, \theta)]$  and  $[W_u(z, \theta), \ W_y(z, \theta)]$ , the model  $M(\theta)$  will in the future refer indistinctly to either one of these equivalent descriptions. For later use, we introduce the following vector notations:

$$W(z, \theta) \triangleq [W_u(z, \theta) \ W_y(z, \theta)], \quad z(t) \triangleq \begin{bmatrix} u(t) \\ y(t) \end{bmatrix} \quad (6)$$

We shall also consider throughout this chapter that the vector process  $z(t)$  is quasistationary [8], so that the spectral density matrix  $\Phi_z(\omega)$  is well defined.

The one-step-ahead prediction error is defined as:

$$\begin{aligned} \varepsilon(t, \theta) &\triangleq y(t) - \hat{y}(t|t-1, \theta) \\ &= W(z, \theta)z(t) = H^{-1}(z, \theta) [y(t) - G(z, \theta)u(t)] \end{aligned} \quad (7)$$

Using a set of input-output data of length  $N$  and a least squares prediction error criterion yields the estimate  $\hat{\theta}_N$  [8]:

$$\hat{\theta}_N = \arg \min_{\theta \in D_\theta} \frac{1}{N} \sum_{t=1}^N \varepsilon^2(t, \theta). \quad (8)$$

Under reasonable conditions [8],  $\hat{\theta}_N \xrightarrow{N \rightarrow \infty} \theta^* \triangleq \arg \min_{\theta \in D_\theta} \bar{V}(\theta)$ , with

$$\bar{V}(\theta) \triangleq E[\varepsilon^2(t, \theta)]. \quad (9)$$

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

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{N \rightarrow \infty} N(0, P_\theta), \quad (10)$$

where

$$P_\theta = [I(\theta)]^{-1} |_{\theta=\theta_0}, \quad (11)$$

$$I(\theta) = \frac{1}{\sigma_e^2} E [\psi(t, \theta)\psi(t, \theta)^T], \quad (12)$$

$$\psi(t, \theta) = -\frac{\partial \varepsilon(t, \theta)}{\partial \theta} = \frac{\partial \hat{y}(t|t-1, \theta)}{\partial \theta} = W(z, \theta)z(t) \quad (13)$$

The matrix  $I(\theta_0)$  is called the information matrix, and will be much coveted in this chapter.

### 3 Identifiability, informative data, and all that jazz

Several concepts of identifiability have been proposed in the scientific literature, and these definitions have evolved over the years. They can be broadly classified into *consistency-oriented definitions*, which focus on whether the parameter estimate  $\hat{\theta}_N$  converges to the ‘true’ parameter  $\theta_0$  in some stochastic sense, and *uniqueness-oriented definitions*, which deal with the question of whether the model structure is such that the identification criterion has a unique global minimum. Here we adopt a uniqueness-oriented definition proposed in [8].

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

$$[W_u(z, \theta) \quad W_y(z, \theta)] = [W_u(z, \theta_1) \quad W_y(z, \theta_1)] \quad \forall \omega \iff \theta = \theta_1.$$

*The model structure is globally identifiable at  $\theta_1$  if the same holds for  $\delta \rightarrow \infty$ . Finally, a model structure is globally identifiable if it is globally identifiable at almost all  $\theta_1$ . ■*

We now introduce the matrix  $\Gamma(\theta) \in \mathcal{R}^{d \times d}$ :

$$\Gamma(\theta) \triangleq \int_{-\pi}^{\pi} \nabla_{\theta} W(e^{j\omega}, \theta) \nabla_{\theta} W^H(e^{j\omega}, \theta) d\omega \quad (14)$$

where  $\nabla_{\theta} W(e^{j\omega}, \theta) \triangleq \frac{\partial W(e^{j\omega}, \theta)}{\partial \theta}$ , and for any  $M(e^{j\omega})$ , the notation  $M^H(e^{j\omega})$  denotes  $M^T(e^{-j\omega})$ . The following result is then an alternative definition for local identifiability of a model structure; see problem 4G.4 in [8].

**Proposition 1.** *A parametric model structure  $M(\theta)$  is locally identifiable at  $\theta_1$  if  $\Gamma(\theta)$  is nonsingular at  $\theta_1$ . ■*

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 [8]. For the existence of a unique global minimum of  $\bar{V}(\theta)$ , it is required that the model structure is globally identifiable at  $\theta_0$ . We illustrate the loss of rank of  $\Gamma(\theta)$  at a pole-zero cancellation with the following example.

*Example 1.* Consider the OE (Output-Error) model structure:  $y(t) = \frac{B(z^{-1})}{F(z^{-1})}u(t) + e(t)$ , where  $B(z^{-1}) = b_1z^{-1} + b_2z^{-2}$  and  $F(z^{-1}) = 1 + f_1z^{-1} + f_2z^{-2}$ , with  $\theta = (b_1 \ b_2 \ f_1 \ f_2)^T$ . Then

$$\nabla_{\theta} W(e^{j\omega}, \theta) = \frac{1}{F^2} \begin{pmatrix} Fz^{-1} \\ Fz^{-2} \\ -Bz^{-1} \\ -Bz^{-2} \end{pmatrix} = \frac{1}{F^2} \underbrace{\begin{pmatrix} 1 & f_1 & f_2 & 0 \\ 0 & 1 & f_1 & f_2 \\ 0 & -b_1 & -b_2 & 0 \\ 0 & 0 & -b_1 & -b_2 \end{pmatrix}}_{S_{BF}} \begin{pmatrix} z^{-1} \\ z^{-2} \\ z^{-3} \\ z^{-4} \end{pmatrix} \quad (15)$$

The matrix  $S_{BF}$  is called a Sylvester matrix [9]: it is nonsingular if and only if the polynomials  $B$  and  $F$  have no common factor. Thus,  $\Gamma(\theta)$  is nonsingular at all values  $\theta$  except those that cause a pole-zero cancellation in  $\frac{B}{F}$ .

The definition of identifiability (local, or global) is a property of the parametrization of the model  $[G(z, \theta), H(z, \theta)]$  or, equivalently,  $[W_u(z, \theta), W_y(z, \theta)]$ . 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 as  $M(\theta_1)$ . However, it does not tell us that if the true system is in the model set for some parameter value  $\theta_0$ , then  $\theta_0$  will be the unique global minimum of the identification criterion. This requires, additionally, that the data set is informative enough to distinguish between different predictors, which leads us to the definition of informative data with respect to a model structure.

**Definition 2. (Informative data)** [8] *A quasistationary data set  $z(t)$  is called informative with respect to a parametric model set  $\{M(\theta), \theta \in \mathcal{D}_{\theta}\}$  if, for any two models  $W(z, \theta_1)$  and  $W(z, \theta_2)$  in that set,*

$$E\{[W(z, \theta_1) - W(z, \theta_2)]z(t)\}^2 = 0 \quad (16)$$

*implies*

$$W(e^{j\omega}, \theta_1) = W(e^{j\omega}, \theta_2) \quad \forall \omega \quad (17)$$

■

By Parseval's theorem, we can rewrite:

$$E\{[W(z, \theta_1) - W(z, \theta_2)]z(t)\}^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |W(e^{j\omega}, \theta_1) - W(e^{j\omega}, \theta_2)|^2 \Phi_z(\omega) d\omega \quad (18)$$

It is easy to see that an experiment that yields  $\Phi_z(\omega) > 0$  for almost all  $\omega$  is informative for all model structures, but such condition is of course unnecessarily strong.

The definition of informative data is with respect to a given model set, not with respect to the true system, which may or may not belong to the model set. In an identification experiment, one typically first selects a globally identifiable model structure; this is a user's choice. Experimental conditions must then be selected that make the data informative with respect to that structure; this is again a user's choice. However, the data are generated by

the true system, in open or in closed loop. Thus, the conditions that make a data set  $z(t)$  informative with respect to some model structure depend on the true system and on the possible feedback configuration. The information matrix (12) combines, as we shall see, information about the identifiability of the model structure and about the informativity of the experiments. In addition, as we have seen in (11), its inverse characterizes the precision with which we can estimate the model parameters from data. We thus rewrite the information matrix in a way that will make the connections with model structure and data much more transparent.

Combining (12) and (13) yields:

$$I(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \nabla_{\theta} W(e^{j\omega}, \theta) \Phi_z(\omega) \nabla_{\theta} W^H(e^{j\omega}, \theta) d\omega \quad (19)$$

where  $\Phi_z(\omega)$  is the power spectrum of the data  $z(t)$  generated by an identification experiment. Comparing this expression with  $\Gamma(\theta)$  in (14), we have the following result.

**Proposition 2.** *Consider an identification experiment that generates data with spectrum  $\Phi_z(\omega)$  and assume that a model structure  $W(z, \theta)$  is used. Then the information matrix is nonsingular at  $\theta_1$  if the following two conditions hold:*

- (i) *the model structure is locally identifiable at  $\theta_1$ ;*
- (ii)  *$\Phi_z(\omega) > 0$  for almost all  $\omega$ .* ■

We introduce the following definition.

**Definition 3. (Regularity)** *We say that the information matrix  $I(\theta)$  is regular at  $\theta_1$  if  $I(\theta_1) \succ 0$ .* ■

While the identifiability of the model structure at  $\theta_1$  is a necessary condition for the regularity of the information matrix, the positivity of the joint spectrum  $\Phi_z(\omega) > 0$  at almost all  $\omega$  is again unnecessarily strong. A major contribution of this chapter will be to describe the weakest possible richness conditions on the input signal  $u(t)$  (in open-loop identification) or  $r(t)$  (in closed-loop identification) that make the information matrix full rank for a given model structure. This turns out to be a remarkably difficult problem. We shall examine this problem in the situation where the system is in the model set. Thus, we make the following assumption.

**Assumption 1** *The true system (1) belongs to the model set  $\mathcal{M}$ , that is  $M(\theta_0) = \mathcal{S}$  for some  $\theta_0 \in \mathcal{D}_{\theta}$ .*

Under Assumption 1 we have the following classical result [8].

**Proposition 3.** *Consider a model structure that obeys Assumption 1, let this model structure be globally identifiable at  $\theta_0$ , and let the data be informative with respect to this model structure. Then  $\theta_0$  is the unique global minimum of  $\bar{V}(\theta)$  defined by (9), and in addition  $I(\theta_0) > 0$ .* ■

## 4 Analysis of the information matrix

Convergence of an identification algorithm to the exact  $\theta_0$  when  $\mathcal{S} \in \mathcal{M}$  rests on the satisfaction of two different conditions:

- the use of a model structure that is identifiable, at least at the global minimum  $\theta_0$  of the asymptotic criterion  $\bar{V}(\theta)$ ;
- the application of experiments that are informative with respect to the model structure used.

These two conditions depend essentially on the used model structure. They depend on the true system only via the generation of the data. Indeed, the data  $z(t)$  must be informative w.r.t. the model structure, but they are generated by the true system. As noted in Proposition 2, the information matrix combines information on the model structure and information on the data generated by the experiment.  $I(\theta)$  can be regular only at values of  $\theta$  that are (at least) locally identifiable, i.e. where  $\Gamma(\theta) > 0$ . At those values, the regularity of the information matrix depends additionally on the informativity of the data set, i.e. on  $\Phi_z(\omega)$ .

Thus the focus of our attention, from now on, will be to seek conditions under which the information matrix  $I(\theta)$  is regular at all values of  $\theta$  at which  $\Gamma(\theta) > 0$ , and in particular at the true  $\theta_0$ , assuming that the system is globally identifiable at  $\theta_0$ . To simplify all expressions, we shall assume that  $\sigma_e = 1$ . The information matrix is then defined as  $I(\theta) = E[\psi(t, \theta)\psi(t, \theta)]$  where  $\psi(t, \theta) = W(z, \theta)z(t)$  is the gradient of the predictor, which we shall call the *pseudoregression vector*: see (13). We first examine the expressions of this gradient.

### 4.1 Expressions of the pseudoregression vector

The pseudoregression vector can be written:

$$\psi(t, \theta) = [\nabla_{\theta} W_u(z, \theta) \quad \nabla_{\theta} W_y(z, \theta)] \begin{bmatrix} u(t) \\ y(t) \end{bmatrix} = \nabla_{\theta} W(z, \theta)z(t) \quad (20)$$

We rewrite this gradient in terms of the external excitation signals,  $u$  and  $e$  in the case of open-loop data,  $r$  and  $e$  in the case of closed-loop data. To improve readability, we delete the explicit dependence on the variables  $z$  and  $\theta$  whenever it creates no confusion.

#### Open-loop identification setup

In open-loop identification, the data are generated as

$$\begin{bmatrix} u(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ G_0 & H_0 \end{bmatrix} \begin{bmatrix} u(t) \\ e(t) \end{bmatrix}$$

The pseudoregressor is then expressed in terms of the external signals as



$$\psi(t, \theta) = [\nabla_\theta W_u + G_0 \nabla_\theta W_y \quad H_0 \nabla_\theta W_y] \begin{bmatrix} u(t) \\ e(t) \end{bmatrix} \quad (21)$$

$$\triangleq V_{uol}(z, \theta)u(t) + V_{eol}(z, \theta)e(t) \quad (22)$$

### Closed-loop identification setup

In closed-loop identification, the data are generated as

$$\begin{bmatrix} u(t) \\ y(t) \end{bmatrix} = S \begin{bmatrix} K & -KH_0 \\ KG_0 & H_0 \end{bmatrix} \begin{bmatrix} r(t) \\ e(t) \end{bmatrix}$$

where  $K = K(z)$  is the controller, and  $S = S(z) = \frac{1}{1+K(z)G_0(z)}$  is the sensitivity function. The pseudoregressor is then expressed in terms of the external signals as

$$\psi(t, \theta) = [SK(\nabla_\theta W_u + G_0 \nabla_\theta W_y) \quad SH_0(\nabla_\theta W_y - K \nabla_\theta W_u)] \begin{bmatrix} r(t) \\ e(t) \end{bmatrix} \quad (23)$$

$$\triangleq V_{rcf}(z, \theta)r(t) + V_{ecf}(z, \theta)e(t) \quad (24)$$

## 4.2 The range and kernel of rank-one vector processes

We observe that in both cases the pseudoregressor  $\psi(t, \theta)$  that “feeds” the information matrix is made up of filtered versions of quasistationary scalar signals, where the filters are  $d$ -vectors of rational transfer functions. In order to study the rank of the matrix  $I(\theta)$  that results from taking the expectation of these rank-one processes, we introduce the following definitions.

**Definition 4.** Let  $V(z) : \mathcal{C} \mapsto \mathcal{K}^d(z)$  be a  $d$ -vector of proper stable rational functions. The left-kernel of  $V(z)$ , denoted  $\text{Ker}\{V(z)\}$ , is the set spanned by all real-valued vectors  $\alpha \in \mathbb{R}^d$  such that  $\alpha^T V(z) = 0 \forall z \in \mathcal{C}$ . Its dimension is called the nullity and annotated  $\nu_V$ . The rank of  $V(z)$  is defined as  $\rho_V = d - \nu_V$ , and  $V(z)$  is said to have full rank if  $\rho_V = d$ . ■

**Definition 5.** Let  $\psi(t) : \mathbb{R} \mapsto \mathbb{R}^d$  be a  $d$ -vector of quasi-stationary processes. The left-kernel of  $\psi(t)$ , denoted  $\text{Ker}\{\psi(t)\}$ , is the set spanned by all real-valued vectors  $\alpha \in \mathbb{R}^p$  such that  $E[\alpha^T \psi(t)]^2 = 0$ , or alternatively  $\alpha^T \Phi_\psi(\omega) \alpha = 0 \forall \omega$  where  $\Phi_\psi(\omega)$  is the spectral density matrix of  $\psi(t)$ . Its dimension is called the nullity and annotated  $\nu_\psi$ . The rank of  $\psi(t)$  is defined as  $\rho_\psi = d - \nu_\psi$ , and  $\psi(t)$  is said to have full rank if  $\rho_\psi = d$ . ■

**Observation.** A  $d$ -vector  $V(z)$  of proper stable rational functions has full rank if  $V(z)$  is output reachable. A  $d \times m$  transfer function matrix  $H(z) = \sum_{k=0}^{\infty} H_k z^{-k}$  is called *output reachable* if  $\nexists \alpha \in \mathbb{R}^d$  such that  $\alpha^T H(z) = 0 \forall z \in \mathcal{C}$  or, equivalently,  $\alpha^T H_k = 0 \forall k$ : see e.g. [10].

With these definitions under our belt, we are now ready to analyze the rank of the information matrix  $I(\theta)$  as a function of the signals  $u$  and  $e$  (in an open-loop setup), or  $r$  and  $e$  (in a closed-loop setup). The following result follows immediately from the definitions.

**Lemma 1.** *The rank of the information matrix  $I(\theta_1)$  at some value  $\theta_1$  is the rank of  $\psi(t, \theta_1)$ . In particular, the information matrix is regular at  $\theta_1$  if and only if  $\text{Ker}\{\psi(t, \theta_1)\} = \{\mathbf{0}\}$ ; equivalent statements are  $\nu_\psi(\theta_1) = 0$  and  $\rho_\psi(\theta_1) = d$ . ■*

The analysis of the rank of  $I(\theta)$  thus reduces to the analysis of the rank of  $\psi(t, \theta)$  which itself is composed of the sum of two vector filters of scalar stationary stochastic processes: see (22) and (24). For the white noise driven terms, the analysis is very simple: we have the following theorem.

**Theorem 1.** *Let  $\psi_e(t, \theta) = V_e(z, \theta)e(t)$ , where  $V_e(z, \theta)$  is a  $d$ -vector of stable proper rational filters and  $e(t)$  is white noise. Then  $\text{Ker}\{\psi_e(t, \theta)\} = \text{Ker}\{V_e(z, \theta)\}$ , and hence  $\rho_{\psi_e} = \rho_{V_e}$ .*

*Proof.* The result follows immediately by observing that, for any  $\alpha \in \mathcal{R}^d$ :

$$\begin{aligned} 0 &= \alpha^T E[\psi_e(t, \theta)\psi_e^T(t, \theta)]\alpha \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \alpha^T V_e(e^{j\omega}, \theta) V_e^H(e^{j\omega}, \theta) \alpha \, d\omega \end{aligned} \quad \blacksquare$$

This proof shows the coherence and the usefulness of our apparently disconnected definitions of kernels for vectors of stationary stochastic processes and for vectors of proper stable transfer functions.

### 4.3 Regularity conditions for $I(\theta)$ : a first analysis

We now exploit the definitions we have just introduced to produce some first conditions on the regularity of the information matrix.

**Theorem 2.** *With the notations introduced in (22) and (24), the information matrix  $I(\theta)$  is regular*

- *in open-loop identification if and only if*

$$\begin{aligned} \text{Ker}\{V_{uol}(z, \theta)u(t) + V_{eol}(z, \theta)e(t)\} &= \text{Ker}\{V_{uol}(z, \theta)u(t)\} \cap \text{Ker}\{V_{eol}(z, \theta)\} \\ &= \{\mathbf{0}\} \end{aligned} \quad (25)$$

- *in closed-loop identification if and only if*

$$\begin{aligned} \text{Ker}\{V_{rcl}(z, \theta)r(t) + V_{ecl}(z, \theta)e(t)\} &= \text{Ker}\{V_{rcl}(z, \theta)r(t)\} \cap \text{Ker}\{V_{ecl}(z, \theta)\} \\ &= \{\mathbf{0}\} \end{aligned} \quad (26)$$

*Proof.* Consider the case of open-loop identification. It follows from (22) and the independence of the signals  $u$  and  $e$  that

$$\alpha^T E[\psi(t, \theta)\psi^T(t, \theta)]\alpha = E[\alpha^T V_{uol}(z, \theta)u(t)]^2 + E[\alpha^T V_{eol}(z, \theta)e(t)]^2. \quad (27)$$

Therefore  $\alpha \in \text{Ker}\{\psi(t, \theta)\}$  if and only if  $\alpha$  belongs to the left-kernels of both  $V_{uol}(z, \theta)u(t)$  and  $V_{eol}(z, \theta)e(t)$ , and hence to their intersection. Next, it follows from Theorem 1 that  $\text{Ker}\{V_{eol}(z, \theta)e(t)\} = \text{Ker}\{V_e(z, \theta)\}$ . The proof is identical for the closed-loop case. ■

Observe that the conditions (25) and (26) use the two distinct but compatible notions of kernel, defined in Definitions 4 and 5, respectively, in the same statement. These conditions show how the regularity of  $I(\theta)$  depends on both the model structure through  $V_{uol}(z, \theta)$  and  $V_{eol}(z, \theta)$  (respectively,  $V_{rci}(z, \theta)$  and  $V_{eci}(z, \theta)$ ) and the excitation signal  $u(t)$  (respectively  $r(t)$ ). We now elaborate on these conditions, separately for the open-loop and for the closed-loop identification setup.

### Open-loop identification

In open-loop identification, the filters  $V_{uol}(z, \theta)$  and  $V_{eol}(z, \theta)$  are given by (21) and (22). Simple calculations show that they are expressed in terms of the model transfer functions  $G(z, \theta)$  and  $H(z, \theta)$  as follows<sup>3</sup>.

$$\begin{aligned} V_{uol}(z, \theta) &= \nabla_{\theta} W_u + G_0 \nabla_{\theta} W_y = \frac{1}{H^2(\theta)} [H(\theta) \nabla_{\theta} G(\theta) + (G_0 - G(\theta)) \nabla_{\theta} H(\theta)] \\ V_{eol}(z, \theta) &= H_0 \nabla_{\theta} W_y = \frac{H_0}{H^2(\theta)} \nabla_{\theta} H(\theta) \end{aligned}$$

We then have the following result.

**Theorem 3.** *Let  $\mathcal{N}_H$  denote the left-kernel of  $\nabla_{\theta} H(z, \theta)$ . Then  $I(\theta)$  is regular either if  $\mathcal{N}_H = \{\mathbf{0}\}$  or if for each non-zero  $d$ -vector  $\alpha \in \mathcal{N}_H$  we have*

$$E[\alpha^T \nabla_{\theta} G(z, \theta) u(t)]^2 \neq 0. \quad (28)$$

*Proof.* First note that the set of vectors  $\{\alpha \in \mathcal{N}_H \subseteq \mathcal{R}^d\}$  spans  $\text{Ker}\{\nabla_{\theta} W_y\} = \text{Ker}\{V_{eol}(z, \theta)\}$ . Therefore, by Theorem 2,  $I(\theta) > 0$  if and only if either  $\mathcal{N}_H = \{\mathbf{0}\}$  or, for each nonzero  $\alpha \in \mathcal{N}_H$ , we have  $E[\alpha^T (\nabla_{\theta} W_u + G_0 \nabla_{\theta} W_y) u(t)]^2 \neq 0$ . Since  $\alpha^T \nabla_{\theta} H(z, \theta) = \mathbf{0}$ , this is equivalent with  $E[\alpha^T \nabla_{\theta} G(z, \theta) u(t)]^2 \neq 0$ . ■

### Closed-loop identification

In closed-loop identification, the filters  $V_{rci}(z, \theta)$  and  $V_{eci}(z, \theta)$  are given by (23) and (24). They are expressed in terms of the model transfer functions  $G(z, \theta)$  and  $H(z, \theta)$  as follows.

$$\begin{aligned} V_{rci}(z, \theta) &= KS(\nabla_{\theta} W_u + G_0 \nabla_{\theta} W_y) \\ &= KS\left\{\frac{1}{H^2(\theta)} [H(\theta) \nabla_{\theta} G(\theta) + (G_0 - G(\theta)) \nabla_{\theta} H(\theta)]\right\} \\ V_{eci}(z, \theta) &= H_0 S(\nabla_{\theta} W_y - K \nabla_{\theta} W_u) \\ &= \frac{H_0 S}{H^2(\theta)} \{ \nabla_{\theta} H(\theta) - K [H(\theta) \nabla_{\theta} G(\theta) - G(\theta) \nabla_{\theta} H(\theta)] \} \end{aligned}$$

For the closed-loop identification setup we have the following result.

<sup>3</sup> We omit the argument  $z$  here for reasons of brevity.

**Theorem 4.** Let  $\mathcal{N}_{V_{ecl}}$  denote the left-kernel of  $V_{ecl}(z, \theta)$ . Then  $I(\theta)$  is regular either if  $\mathcal{N}_{V_{ecl}} = \{\mathbf{0}\}$  or if for each non-zero  $d$ -vector  $\alpha \in \mathcal{N}_{V_{ecl}}$  we have

$$E[\alpha^T \nabla_{\theta} W_y(z, \theta) r(t)]^2 = E[\alpha^T K(z) \nabla_{\theta} W_u(z, \theta) r(t)]^2 \neq 0. \quad (29)$$

*Proof.* First note that for each  $\alpha \in \mathcal{N}_{V_{ecl}} \subseteq \mathcal{R}^d$  we have  $\alpha^T \nabla_{\theta} W_y(z, \theta) = \alpha^T K(z) \nabla_{\theta} W_u(z, \theta)$ . By Theorem 2,  $I(\theta) > 0$  if and only if either  $\mathcal{N}_{V_{ecl}} = \{\mathbf{0}\}$  or if, for each non-zero  $\alpha \in \mathcal{N}_{V_{ecl}}$  we have  $E[\alpha^T K S(\nabla_{\theta} W_u + G_0 \nabla_{\theta} W_y) r(t)]^2 \neq 0$ . Now observe that  $\alpha^T K S(\nabla_{\theta} W_u + G_0 \nabla_{\theta} W_y) = \alpha^T S(1 + KG_0) \nabla_{\theta} W_y = \alpha^T \nabla_{\theta} W_y = \alpha^T K \nabla_{\theta} W_u$ . This proves the result. ■

#### 4.4 Rich and exciting signals

In Theorem 1 we have seen that a regressor  $\psi(t)$  obtained by filtering a white noise signal  $e(t)$  through a vector filter  $V(z)$  has the same left-kernel as  $V(z)$ , i.e. white noise causes no drop of rank. The same is actually true for any input signal that has a continuous spectrum. For the parts of  $\psi(t, \theta)$  driven by the controlled signals  $u(t)$  or  $r(t)$  (see (22) and (24)), we want to consider input signals ( $u(t)$  or  $r(t)$ ) that have discrete spectra, such as multisines. In order to analyze the rank properties of regressors obtained by filtering such signals with discrete spectra, we need to introduce the concept of richness of a signal. We first define a persistently exciting regression vector.

**Definition 6.** A quasistationary vector signal  $\psi(t)$  is called persistently exciting (denoted PE) if  $E[\psi(t)\psi^T(t)] > 0$ . ■

Whether a quasistationary vector signal  $\psi(t)$  obtained as a filtered version (by a vector  $V(z)$  of transfer functions) of a quasistationary scalar signal  $u(t)$  is PE or not depends not only on whether  $\text{Ker}\{V(z)\} = \{\mathbf{0}\}$  but also on the degree of richness of the input  $u(t)$ . The richness of a scalar signal is defined as follows.

**Definition 7.** A quasistationary scalar signal  $u(t)$  is sufficiently rich of order  $n$  (denoted SRn) if the following regressor is PE:

$$\phi_{1,n}(t) \triangleq \begin{bmatrix} u(t-1) \\ u(t-2) \\ \vdots \\ u(t-n) \end{bmatrix} = \begin{bmatrix} z^{-1} \\ z^{-2} \\ \vdots \\ z^{-n} \end{bmatrix} u(t) \quad (30)$$

The vector  $\phi_{1,n}(t)$  serves as a basis for all regression vectors that are obtained as (vector)-filtered versions of a scalar signal  $u(t)$ . For future use, we introduce the notation:

$$\mathcal{B}_{k,n}(z) \triangleq [z^{-k} \ z^{-k-1} \ \dots \ z^{-n}]^T, \text{ for } k \leq n. \quad (31)$$

Observe that, by our assumption of quasistationarity,  $u(t)$  is SRn if  $\mathcal{B}_{k+1,k+n}(z)u(t)$  is PE for any  $k$ . We denote by  $\mathcal{U}_n$  the set of all SRn processes.

**Definition 8.** A scalar signal  $u(t)$  is sufficiently rich of order exactly  $n$  (denoted *SRE $n$* ) if  $\phi_{1,n}(t)$  is PE, but  $\phi_{1,n+1}(t)$  is not. ■

This definition is equivalent with many other classically used definitions, except that nowadays the most common terminology is to say that a signal is PE of order  $n$  rather than SR of order  $n$ . At the risk of being considered old-fashioned, we prefer the term *sufficiently rich* because sufficient intuitively reflects the notion of *degree of richness* while persistent does not. The following are commonly used definitions that are equivalent with Definitions 7 and 8.

**Proposition 4.** A scalar quasistationary signal  $u(t)$  is *SR $n$*  if

- its spectral density is nonzero in at least  $n$  frequency points in the interval  $(-\pi, \pi]$ .
  - it cannot be filtered to zero by a FIR filter of degree  $n$ :  $\alpha_1 z^{-1} + \dots + \alpha_n z^{-n}$ .
- A scalar signal  $u(t)$  is *SRE $n$*  if its spectral density is nonzero in exactly  $n$  frequency points in the interval  $(-\pi, \pi]$ . ■

The equivalence comes by observing that

$$\alpha^T E[\phi_{1,n}(t)\phi_{1,n}^T(t)]\alpha = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\alpha_1 e^{-j\omega} + \dots + \alpha_n e^{-jn\omega}|^2 \Phi_u(\omega) d\omega.$$

The question of interest here is how the richness of a scalar signal  $u(t)$  transfers into the persistence of excitation of a regression vector  $\psi(t)$  when this regression vector is obtained as a (vector)-filter of  $u(t)$ , i.e.  $\psi(t) = V(z)u(t)$ , where the components of  $V(z)$  are stable proper transfer functions. More precisely, we would like to determine the smallest possible degree of richness of  $u(t)$  that will make  $\psi(t)$  PE. To help us in solving this problem, we have . . . . . not much. As it happens, the only available results, as far as we know, are sufficiency results. We briefly recall here the main available results.

**Proposition 5.** [6] Let  $u(t)$  be *SR $n$*  and let  $y(t) = G(z)u(t)$ ; then  $y(t)$  is *SR $n$*  if the filter  $G(z)$  has no zeroes on the unit circle, i.e. if  $G(e^{j\omega}) \neq 0$  for all  $\omega$ .

*Proof.* Since  $\Phi_y(\omega) = |G(e^{j\omega})|^2 \Phi_u(\omega)$ , it follows immediately that if  $G(e^{j\omega})$  is nowhere zero on the unit circle, then the frequency points where  $\Phi_y(\omega) \neq 0$  and where  $\Phi_u(\omega) \neq 0$  are identical. ■

**Proposition 6.** [11] Let  $\psi(t) \in \mathcal{R}^d$  be a vector that has the state-space model

$$\psi(t+1) = A\psi(t) + Bu(t) \tag{32}$$

with  $A \in \mathcal{R}^{d \times d}$ . Then  $\psi(t)$  is PE if  $u(t)$  is *SR $d$*  and the pair  $[A, B]$  is completely reachable. ■

**Proposition 7.** [10] Let  $\psi(t) \in \mathcal{R}^d$  be the output of a vector  $V(z)$  of proper stable filters driven by  $u(t)$ :

$$\psi(t) = V(z)u(t), \quad (33)$$

and let  $\delta_V \triangleq$  McMillan degree of  $V(z)$ . Then  $\psi(t)$  is PE if the following two conditions hold:

- the system (33) is output reachable, i.e.  $\nexists \alpha \in \mathcal{R}^d, \alpha \neq 0$ , such that  $\alpha^T V(z) = 0 \forall z$ ;
- $u(t)$  is SRn with  $n \geq \delta_V + 1$ . ■

We now attempt to relate the left-kernel of  $\psi(t)$ , the left-kernel of  $V(z)$  and the richness of  $u(t)$ . We first state the trivial lemma.

**Lemma 2. The trivial lemma.** Let  $\psi(t) = V(z)u(t)$  with  $\psi(t) \in \mathcal{R}^d, u(t)$  quasistationary, and all components of  $V(z)$  proper and stable. Then

$$\text{Ker}\{V(z)\} \subseteq \text{Ker}\{\psi(t)\}. \quad (34)$$

■

The question we address now can be stated as follows.

*What are the necessary and sufficient conditions on the richness of  $u(t)$  such that  $\text{Ker}\{\psi(t)\} = \text{Ker}\{V(z)\}$  when  $\psi(t) = V(z)u(t)$ ?*

Let the least common denominator of all elements of  $V(z)$  have degree  $\delta_V$ ; one can then always write

$$V(z) = \frac{N(z^{-1})}{d(z^{-1})} = \frac{z^{-m}}{d(z^{-1})} R\mathcal{B}_{0,k-1}(z) \quad (35)$$

where  $d(z^{-1}) = 1 + d_1 z^{-1} + \dots + d_p z^{-p}$ , with  $d_p \neq 0$  and  $p + m = \delta_V$ , where  $R \in \mathcal{R}^{d \times k}$  is the matrix of real coefficients of the expansion of the numerator matrix  $N(z^{-1})$  into powers of  $z^{-1}$ , and  $m$  is a possible common delay in all elements of  $N(z^{-1})$ . A necessary condition for  $V(z)$  to be output reachable (i.e.  $\text{Ker}\{V(z)\} = \mathbf{0}$ ) is that  $k \geq d$ . As we shall see, in many cases of interest for the transfer of excitation from the signal  $u$ , or  $r$ , to the pseudo-regression vector  $\psi$ , it so happens that  $R$  is square, i.e.  $k = d$ . Thus, we first handle this important (and much easier) special case.

**Theorem 5.** Let  $\psi(t) = V(z)u(t)$  with  $\psi(t) \in \mathcal{R}^d, u(t)$  quasistationary,  $V(z)$  proper and stable, and let  $V(z)$  be decomposed as in (35) with  $d = k$ . Then  $\rho_\psi = d$  if and only if  $\rho_V = d$  and  $u(t)$  is SRd. (Stated otherwise:  $\psi(t)$  has full rank if and only if  $V(z)$  has full rank and  $u(t)$  is SRd).

*Proof.* Using the decomposition (35) with  $k = d$ , we can write

$$E[\alpha^T \psi(t)]^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\alpha^T R \mathcal{B}_{1,d}(e^{j\omega})|^2 \frac{\Phi_u(\omega)}{|d(e^{j\omega})|^2} d\omega, \quad (36)$$

where we have used the fact that  $|e^{-mj\omega} \alpha^T R \mathcal{B}_{0,d-1}(e^{j\omega})|^2 = |\alpha^T R \mathcal{B}_{1,d}(e^{j\omega})|^2$ . If  $\rho_V = d$ , and since  $k = d$ , it follows that  $R$  is nonsingular. Therefore  $\alpha^T R \neq 0$  for all nonzero  $\alpha$ , and  $\alpha^T R$  spans the space of all vectors in  $\mathcal{R}^d$ . If in addition  $u(t)$  is SRd, then by Proposition 4 the integral on the right hand side is nonzero for all  $\alpha \neq 0$ . Conversely, if  $\rho_V < d$  then there exists  $\alpha \neq 0$  such that  $\alpha^T R = 0$ , and if  $u(t)$  is SREn with  $n < d$ , then there exists  $\alpha \neq 0$  such that  $|\alpha^T R \mathcal{B}_{1,d}(e^{j\omega})|^2 \Phi_u(\omega) = 0$  for all  $\omega$ . ■

We observe that the richness condition on  $u$  in this last result is weaker than that given by Proposition 7, since in the decomposition (35) we have  $k \leq \delta_V$ , where  $\delta_V$  is the McMillan degree of  $V(z)$ .

*Example 2.* Let

$$\begin{aligned} V(z) &= \frac{1}{z^3(z+0.5)} \begin{bmatrix} z+1 \\ z^2 \\ 1 \end{bmatrix} = \frac{z^{-2}}{1+0.5z^{-1}} \begin{bmatrix} z^{-1} + z^{-2} \\ 1 \\ z^{-2} \end{bmatrix} \\ &= \frac{z^{-2}}{1+0.5z^{-1}} R \mathcal{B}_{0,2} \quad \text{with } R = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

The McMillan degree of  $V(z)$  is  $\delta_V = 4$ , but  $\psi(t) \triangleq V(z)u(t)$  will be PE if  $u(t)$  is SR3, since  $R$  has full rank.

In the more general situation where  $k > d$ , it is clear that  $u(t) = \text{SR}k$  is a sufficient condition to guarantee that  $\text{Ker}\{\psi(t)\} = \text{Ker}\{V(z)\}$ . However, we shall now show that  $\text{Ker}\{\psi(t)\} = \text{Ker}\{V(z)\}$  for almost all  $u(t) \in \mathcal{U}_N$  if and only if  $N \geq \rho_V$ ; recall that  $\mathcal{U}_N$  is the set of  $u(t)$  that are SRN, and  $\rho_V$  is the rank of  $V(z)$ . To show this, we need a preliminary lemma.

**Lemma 3.** *Let  $\psi(t) = V(z)u(t)$  with  $\psi(t) \in \mathcal{R}^d$ ,  $u(t)$  quasistationary,  $V(z)$  proper and stable, and let  $V(z)$  be decomposed as in (35) with  $\rho_V = c$ . Let the rows of  $Q \in \mathcal{R}^{c \times k}$  be a basis for the rowspace of  $R$ , and define the  $c$ -vectors  $W(z) = \frac{z^{-m}}{d(z^{-1})} Q \mathcal{B}_{0,k-1}(z)$  and  $\phi(t) = W(z)u(t)$ . Then, for any  $u(t)$ ,  $\text{Ker}\{\psi(t)\} = \text{Ker}\{V(z)\}$  if and only if  $\text{Ker}\{\phi(t)\} = \text{Ker}\{W(z)\} = \mathbf{0}$ .*

*Proof.* Since the rows of  $Q$  form a basis for the rowspace of  $R$  we can write

$$R = T \begin{bmatrix} Q \\ 0 \end{bmatrix} \quad (37)$$

for some nonsingular matrix  $T \in \mathcal{R}^{d \times d}$ . Then for any  $\alpha \in \mathcal{R}^d$  we have:

$$\alpha^T R = \alpha^T T \begin{bmatrix} Q \\ 0 \end{bmatrix} = \beta^T Q \quad (38)$$

where  $\beta$  is uniquely defined by  $\alpha^T T \triangleq (\beta^T \ \gamma^T)$  with  $\beta \in \mathcal{R}^c$  and  $\gamma \in \mathcal{R}^{d-c}$ . It follows from (38) that

$$\alpha^T \psi(t) = \frac{z^{-m}}{d(z^{-1})} \alpha^T R \mathcal{B}_{0,k-1}(z) u(t) = \frac{z^{-m}}{d(z^{-1})} \beta^T Q \mathcal{B}_{0,k-1}(z) u(t) = \beta^T \phi(t)$$

Therefore the following four statements are all equivalent:

- $Ker\{\psi(t)\} = Ker\{V(z)\}$
- $E[\alpha^T \psi(t)]^2 = 0$  if and only if  $\alpha^T R \in Ker\{\mathcal{B}_{0,k-1}(z)u(t)\}$
- $E[\beta^T \phi(t)]^2 = 0$  if and only if  $\beta^T Q \in Ker\{\mathcal{B}_{0,k-1}(z)u(t)\}$
- $Ker\{\phi(t)\} = Ker\{W(z)\}$

Finally, since  $Q$  has full rank,  $Ker\{W(z)\} = \mathbf{0}$ . ■

**Theorem 6.** *Let  $\psi(t) = V(z)u(t)$  with  $\psi(t) \in \mathcal{R}^d$ ,  $u(t)$  quasistationary,  $V(z)$  proper and stable, and let  $V(z)$  be decomposed as in (35). Then  $Ker\{\psi(t)\} = Ker\{V(z)\}$  for almost all  $u(t) \in \mathcal{U}_N$  if and only if  $N \geq \rho_V$ .*

*Proof.* If  $\rho_V < d$ , we can replace  $\psi(t) = V(z)u(t)$  by  $\phi(t) = W(z)u(t)$  with  $W(z)$  defined from  $V(z)$  as in Lemma 3 above, where  $W(z)$  has full rank. Thus, using Lemma 3, we can assume without loss of generality that  $\rho_V = d$ .

Using Parseval's Theorem and (35) we can write

$$\begin{aligned} E[\alpha^T \psi(t)]^2 &= E[\alpha^T \frac{z^{-m}}{d(z^{-1})} R \mathcal{B}_{0,k-1}(z) u(t)]^2 \\ &= \alpha^T R \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) \frac{1}{|d(e^{-j\omega})|^2} \mathcal{B}_{0,k-1}(e^{-j\omega}) \mathcal{B}_{0,k-1}^H(e^{-j\omega}) d\omega \right) R^T \alpha \end{aligned}$$

Let  $u(t)$  be SRN with finite  $N$ . Its spectrum can then be written as  $\Phi_u(\omega) = \sum_{i=1}^N \lambda_i \Phi_u(\omega_i)$  with  $\omega_i \neq \omega_j$ ,  $i \neq j$ . Define its support as the vector  $\mathbf{z} = [e^{j\omega_1} \ e^{j\omega_2} \ \dots \ e^{j\omega_N}] \in \Omega_N$ , where  $\Omega_N \subset \mathcal{C}^N$  is the set of all supports  $\mathbf{z}$  which result in an SRN signal, that is those  $\mathbf{z}$  such that  $\omega_i \neq \omega_j \ \forall i \neq j$ .  $\Omega_N$  is an  $N$ -dimensional subset of  $\mathcal{C}^N$  which defines the class of signals  $u(t)$  that we consider. Then we can write

$$E[\alpha^T \psi(t)]^2 = \alpha^T R \left( \frac{1}{2\pi} \sum_{i=1}^N \lambda'_i \mathcal{B}_{0,k-1}(e^{-j\omega_i}) \mathcal{B}_{0,k-1}^H(e^{-j\omega_i}) \right) R^T \alpha$$

where  $\lambda'_i = \frac{\lambda_i \Phi_u(\omega_i)}{|d(e^{-j\omega_i})|^2}$ . Hence

$$E[\alpha^T \psi(t)]^2 = \alpha^T R Z(\mathbf{z}) \Lambda Z^H(\mathbf{z}) R^T \alpha \quad (39)$$

with



$$Z(\mathbf{z}) = [\mathcal{B}_{0,k-1}(e^{-j\omega_1}) \mathcal{B}_{0,k-1}(e^{-j\omega_2}) \dots \mathcal{B}_{0,k-1}(e^{-j\omega_N})]$$

and  $A = \text{diag}\{\lambda'_1, \lambda'_2, \dots, \lambda'_N\}$ ; note that  $\rho(Z(\mathbf{z})) = N$  whenever  $N \leq k$ .

But  $\psi(t)$  is full-rank if and only if  $P(\mathbf{z}) \triangleq RZ(\mathbf{z})AZ^H(\mathbf{z})R^T$  has rank equal to  $d$ , which is equivalent to  $\det(P(\mathbf{z})) \neq 0$ . Suppose that  $N < d$ ; then  $\rho(Z(\mathbf{z})) = N < d$  which, noting that  $\rho(P(\mathbf{z})) \leq \rho(Z(\mathbf{z}))$ , implies  $\det(P(\mathbf{z})) = 0$ , thus proving the necessity of  $N \geq d$ .

For  $N \geq d$ , the determinant  $\det(P(\mathbf{z}))$  is a nontrivial polynomial in the vector variable  $\mathbf{z}$  and  $\psi(t)$  loses rank exactly at the roots of this polynomial. Since the roots of a polynomial define a set of measure zero in the space of its variable,  $\psi(t)$  is full-rank for almost all  $\mathbf{z} \in \Omega_N$ . ■

Our Theorem above completely characterizes the required signal richness of  $u(t)$  that keeps the range of the regressor vector  $V(z)u(t)$  identical to the range of  $V(z)$ . Yet, it is worth specifying what happens for different levels of excitation, which is given in the following theorem.

**Theorem 7.** *Let  $\psi(t) = V(z)u(t)$  with  $\psi(t) \in \mathcal{R}^d$ ,  $u(t)$  quasistationary,  $V(z)$  proper and stable, and let  $V(z)$  be decomposed as in (35) with  $\text{rank}(V) = \rho_V$ .*

- *If  $u(t)$  is **not** SR of order  $\rho_V$ , then  $\text{Ker}(\psi(t)) \subset \text{Ker}(V(z))$ .*
- *If  $u(t)$  is SRk then  $\text{Ker}(\psi(t)) = \text{Ker}(V(z))$ .*

*Proof.* This result follows immediately from Sylvester's inequality:

$$\rho(R) + \rho(Z) - k \leq \rho(RZ) \leq \min(\rho(R), \rho(Z))$$

which yields  $\rho(RZ) < \rho(R)$  for  $N < \rho_V$  and  $\rho(RZ) \geq \rho(R)$  for  $N \geq k$ . ■

The following example illustrates the results of our theorems.

*Example 3.* Consider the regressor  $\psi(t) = V(z)u(t)$ , with

$$V(z) = R [1 \ z^{-1} \ z^{-2} \ z^{-3} \ z^{-4}]^T \quad \text{where } R = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Consider first  $u(t) = \lambda_1 + \lambda_2 \sin(\omega t)$ , which is SRE3. For such signal,  $RZ$  is a  $3 \times 3$  matrix, whose determinant is  $\det(RZ) = -2j[3 \sin(\omega) - 2 \sin(2\omega) - \sin(3\omega) + \sin(4\omega)]$ . Its roots in  $(-\pi, \pi]$  are at  $-\frac{\pi}{3}$ ,  $0$ ,  $\frac{\pi}{3}$  and  $\pi$ , but  $\omega = 0$  and  $\omega = \pi$  do not keep  $u(t) \in \mathcal{U}_3$ . Thus,  $\psi(t)$  will have rank 3 for all  $u(t) \in \mathcal{U}_3$  except for  $u(t) = \lambda_1 + \lambda_2 \sin(\frac{\pi}{3}t)$ , i.e. for  $\omega = \frac{\pi}{3}$ .

Now let  $u(t) = \lambda_1 \sin(\omega_1 t) + \lambda_2 \sin(\omega_2 t)$  which is SRE4. From Theorem 7 we know that the richness of this signal is in between the ‘‘necessary’’ richness (SR3) and the sufficient richness (SR5). We have

$$RZ = \begin{bmatrix} e^{j\omega_1} + e^{j4\omega_1} & e^{j\omega_2} + e^{j4\omega_2} & e^{-j\omega_1} + e^{-j4\omega_1} & e^{-j\omega_2} + e^{-j4\omega_2} \\ e^{j2\omega_1} & e^{j2\omega_2} & e^{-j2\omega_1} & e^{-j2\omega_2} \\ 1 + e^{j4\omega_1} & 1 + e^{j4\omega_2} & 1 + e^{-j4\omega_1} & 1 + e^{-j4\omega_2} \end{bmatrix}$$

It is rather easy to see that  $RZ$  will have full rank for all values of  $\omega_1$  and  $\omega_2$ ,  $\omega_1 \neq \omega_2$  except those for which  $\omega_1 + \omega_2 = \pi$ .

## 5 Regularity of $I(\theta)$ for ARMAX and BJ model structures

We now combine the results of Theorem 3 with those on the transfer of sufficiently rich input signals to regression vectors in order to produce necessary and sufficient richness conditions on the input signal that guarantee regularity of the information matrix at all  $\theta$  at which the model structure is identifiable, i.e.  $I(\theta) > 0$ . We do this for ARMAX and Box-Jenkins (BJ) model structures in an open-loop identification setup.

### ARMAX model structure

Consider the ARMAX model structure

$$A(z^{-1})y(t) = B(z^{-1})u(t) + C(z^{-1})e(t) \quad (40)$$

where  $A(z^{-1}) = 1 + a_1z^{-1} + \dots + a_{n_a}z^{-n_a}$ ,  $B(z^{-1}) = b_1z^{-1} + \dots + b_{n_b}z^{-n_b}$ , and  $C(z^{-1}) = 1 + c_1z^{-1} + \dots + c_{n_c}z^{-n_c}$ . We have the following result.

**Theorem 8.** *For the ARMAX model structure (40), the information matrix  $I(\theta)$  is regular at a  $\theta$  at which the model structure is identifiable if and only if  $u(t)$  is SRk, where  $k = n_b + n_u(\theta)$  and  $n_u(\theta)$  is the number of common roots of the polynomials  $A(z^{-1})$  and  $C(z^{-1})$  at that  $\theta$ .  $I(\theta)$  is regular at all  $\theta$  at which the model structure is identifiable if and only if  $u(t)$  is SRk with  $k = n_b + \min\{n_a, n_c\}$ .*

*Proof.* We first comment that for ARMAX model structures, common roots between the polynomials  $A$  and  $B$ , as well as between  $A$  and  $C$ , must be considered, because they can generically occur. However, the three polynomials  $A$ ,  $B$  and  $C$  must be coprime at any identifiable  $\theta$ . For the ARMAX model structure, we have:

$$\nabla_{\theta}G(z, \theta) = \frac{1}{A^2} \begin{pmatrix} -Bz^{-1} \\ \vdots \\ -Bz^{-n_a} \\ \hline Az^{-1} \\ \vdots \\ Az^{-n_b} \\ \hline 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \nabla_{\theta}H(z, \theta) = \frac{1}{A^2} \begin{pmatrix} -Cz^{-1} \\ \vdots \\ -Cz^{-n_a} \\ \hline 0 \\ \vdots \\ 0 \\ \hline Az^{-1} \\ \vdots \\ Az^{-n_c} \end{pmatrix} \quad (41)$$

Let  $\alpha^T = (\alpha_A^T \mid \alpha_B^T \mid \alpha_C^T)$  denote any vector in the left-kernel of  $\nabla_{\theta}H(z, \theta)$ , and let  $\gamma_A(z^{-1}) \triangleq \alpha_A^T \mathcal{B}_{1, n_a}$ ,  $\gamma_B(z^{-1}) \triangleq \alpha_B^T \mathcal{B}_{1, n_b}$ , and  $\gamma_C(z^{-1}) \triangleq \alpha_C^T \mathcal{B}_{1, n_c}$ . Then

$$\alpha^T \nabla_\theta H(z, \theta) = 0 \Leftrightarrow z\gamma_A(z^{-1})C(z^{-1}) = z\gamma_C(z^{-1})A(z^{-1}) \quad (42)$$

At all values of  $\theta$  at which the polynomials  $A$  and  $C$  are coprime, it follows from the theory of Diophantine equations (see e.g. [12]) that  $\alpha_A = \mathbf{0}$  and  $\alpha_C = \mathbf{0}$ , because  $\deg(z\gamma_A(z^{-1})) < \deg(A(z^{-1}))$  and  $\deg(z\gamma_C(z^{-1})) < \deg(C(z^{-1}))$ . Consider now a  $\theta$  at which there are common factors between  $A$  and  $C$  and let  $U(z^{-1})$  denote the Greatest Common Divisor (GCD) of  $A$  and  $C$ , with  $\deg(U(z^{-1})) = n_u$ . Then  $A = A_1U$  and  $C = C_1U$  for some coprime polynomials  $A_1$  and  $C_1$ . Then (42) is equivalent with  $z\gamma_A(z^{-1})C_1(z^{-1}) = z\gamma_C(z^{-1})A_1(z^{-1})$  where  $\deg(z\gamma_A) = n_a - 1$  and  $\deg(z\gamma_C) = n_c - 1$ . The set of all solutions of this equation is described by

$$z\gamma_A = \alpha_A^T \mathcal{B}_{0, n_a - 1} = A_1 T, \quad z\gamma_C = \alpha_C^T \mathcal{B}_{0, n_c - 1} = C_1 T \quad (43)$$

where  $T(z^{-1})$  is an arbitrary polynomial of degree  $n_u - 1$ . The left-kernel of  $\nabla_\theta H(z, \theta)$  is thus defined by those vectors  $\alpha^T = (\alpha_A^T \mid \alpha_B^T \mid \alpha_C^T)$  such that  $\alpha_A$  and  $\alpha_C$  are solution of (43), while  $\alpha_B$  is arbitrary. As stated earlier, we consider values of  $\theta$  at which  $I(\theta) > 0$ . At these values of  $\theta$ ,  $\alpha^T \nabla_\theta G(z, \theta) \neq 0$  for all vectors  $\alpha$  defined above and, by Theorem 3,  $I(\theta) > 0$  if  $u(t)$  is such that  $E[\alpha^T \nabla_\theta G(z, \theta) u(t)]^2 \neq 0$  for all such  $\alpha$ . For such  $\alpha$ , we have:

$$\begin{aligned} \alpha^T \nabla_\theta G(z, \theta) u(t) &= \frac{1}{A^2} [-\alpha_A^T \mathcal{B}_{1, n_a} B + \alpha_B^T \mathcal{B}_{1, n_b} A] u(t) \\ &= \frac{1}{A^2} [z^{-1} A_1 T B + \alpha_B^T \mathcal{B}_{1, n_b} A_1 U] u(t) \\ &= \frac{1}{AU} [z^{-1} T B + \alpha_B^T \mathcal{B}_{1, n_b} U] u(t) \end{aligned} \quad (44)$$

where the coefficients of the polynomial  $T$ , of degree  $n_u - 1$ , as well as the coefficients of  $\alpha_B$  are completely free. Therefore  $E[\alpha^T \nabla_\theta G(z, \theta) u(t)]^2 \neq 0$  if and only if the following pseudoregressor has full rank:

$$\psi(t) = \frac{1}{AU} \begin{pmatrix} -Bz^{-1} \\ \vdots \\ -Bz^{-n_u} \\ Uz^{-1} \\ \vdots \\ Uz^{-n_b} \end{pmatrix} u(t) = \frac{1}{AU} R \mathcal{B}_{1, n_b + n_u} u(t) \quad (45)$$

with  $R \in \mathcal{R}^{(n_b + n_u) \times (n_b + n_u)}$ . Since  $A, B, C$  are coprime at all  $\theta$ , and  $U$  is the common factor of  $A$  and  $C$ , it follows that  $B$  and  $U$  are coprime, and hence  $R$  in (45) is nonsingular. Therefore, by Theorem 5,  $\psi(t)$  in (45) is PE (and hence  $I(\theta) > 0$ ) if and only if  $u(t)$  is sufficiently rich of degree  $n_b + n_u(\theta)$ , where  $n_u(\theta)$  represents the number of common roots between  $A$  and  $C$ . Since the maximum number of such common roots is  $\min\{n_a, n_c\}$ ,  $I(\theta)$  is regular at all identifiable  $\theta$  if and only if  $u(t)$  is SRk with  $k = n_b + \min\{n_a, n_c\}$ . ■

**BJ model structure**

Consider now the BJ model structure:

$$y(t) = \frac{B(z^{-1})}{F(z^{-1})}u(t) + \frac{C(z^{-1})}{D(z^{-1})}e(t) \quad (46)$$

where  $B(z^{-1})$  and  $C(z^{-1})$  are as above, with  $F(z^{-1}) = 1 + f_1z^{-1} + \dots + f_{n_f}z^{-n_f}$  and  $D(z^{-1}) = 1 + d_1z^{-1} + \dots + d_{n_d}z^{-n_d}$ .

**Theorem 9.** *For the BJ model structure (46), the information matrix  $I(\theta)$  is regular at a  $\theta$  at which the model structure is identifiable if and only if  $u(t)$  is SRk, where  $k = n_b + n_f$ .*

*Proof.* The gradient vectors  $V_{uol}(z, \theta)$  and  $V_{eol}(z, \theta)$  defined in (22) are now partitioned into 4 blocks corresponding, successively, to the parameters of the polynomials  $B, F, C$ , and  $D$ . It is easy to see that the left-kernel of  $V_{eol}(z, \theta)$  (i.e. of  $\nabla_{\theta}H(z, \theta)$ ) is spanned by the set of vectors  $\alpha^T = (\alpha_B^T \mid \alpha_F^T \mid 0 \dots 0 \mid 0 \dots 0)$ . Therefore, by Theorem 3,  $I(\theta) > 0$  if and only if the following pseudoregressor is PE:

$$\begin{aligned} \psi_{B,F}(t) &\triangleq \frac{1}{F^2} \begin{pmatrix} Fz^{-1} \\ \vdots \\ Fz^{-n_b} \\ -Bz^{-1} \\ \vdots \\ -Bz^{-n_f} \end{pmatrix} u(t) & (47) \\ &= \frac{1}{F^2} \underbrace{\begin{pmatrix} 1 & f_1 & \dots & f_{n_f} & 0 & \dots & 0 \\ 0 & 1 & f_1 & \dots & f_{n_f} & \dots & 0 \\ 0 & 0 & \dots & \ddots & & \ddots & 0 \\ 0 & \dots & \dots & 1 & f_1 & \dots & f_{n_f} \\ 0 & -b_1 & \dots & -b_{n_b} & 0 & \dots & 0 \\ 0 & 0 & -b_1 & \dots & -b_{n_b} & 0 & \dots \\ 0 & 0 & \dots & \ddots & & \ddots & 0 \\ 0 & \dots & \dots & 0 & -b_1 & \dots & -b_{n_b} \end{pmatrix}}_{S_{BF}} \mathcal{B}_{1, n_b + n_f}(z) u(t) & (48) \end{aligned}$$

$S_{BF}$  is a Sylvester matrix, with dimensions  $(n_b + n_f) \times (n_b + n_f)$ . It is nonsingular for all values of  $\theta$  at which the polynomials  $B$  and  $F$  are coprime. Applying Theorem 5 again, we conclude that  $I(\theta) > 0$  at all values of  $\theta$  at which the polynomials are coprime if and only if  $u(t)$  is SRk, where  $k = n_b + n_f$ . ■

Just in the same vein, one can apply the results of Theorem 4 to the identification of closed-loop systems using ARMAX or BJ model structures. The results for these two closed-loop setups, which are quite illuminating, will be presented at Giorgio Picci's next 65th birthday celebration.

## 6 Conclusions

The information matrix plays a fundamental role in system identification, given that it combines information about the identifiability of the model structure and about the informativity of the data set. We have illustrated these connections, and we have provided conditions on the richness of the input signals that make the information matrix full rank at all values of the parameter space where the model structure is identifiable. Our objective has been to find the smallest possible degree of richness of the input signal that delivers a non-singular information matrix. In deriving these conditions, we have presented some new results on the degree of richness required to produce a persistently exciting regressor.

### Acknowledgements

We wish to thank the organizers of the Giorgio Picci workshop for giving us the opportunity to contribute this chapter. We also thank Roland Hildebrand and Luc Haine for some useful hints for the proof of Theorem 6.

### References

1. X. Bombois, G. Scorletti, M. Gevers, R. Hildebrand, and P. Van den Hof, "Cheapest open-loop identification for control," in *CD-ROM Proc. 33rd IEEE Conf on Decision and Control*, The Bahamas, December 2004, pp. 382–387.
2. X. Bombois, G. Scorletti, M. Gevers, P. Van den Hof, and R. Hildebrand, "Least costly identification experiment for control," *Automatica*, vol. 42, no. 10, pp. 1651–1662, October 2006.
3. H. Jansson and H. Hjalmarsson, "Optimal experiment design in closed loop," in *16th IFAC World Congress on Automatic Control, paper 04528*, July 2005.
4. M. Gevers, L. Mišković, D. Bonvin, and A. Karimi, "Identification of multi-input systems: variance analysis and input design issues," *Automatica*, vol. 42, no. 4, pp. 559–572, April 2006.
5. T. S. Ng, G. C. Goodwin, and B. D. O. Anderson, "Identifiability of MIMO linear dynamic systems operating in closed loop," *Automatica*, vol. 13, pp. 477–485, 1977.
6. T. Söderström and P. Stoica, *System Identification*. Hemel Hempstead, Hertfordshire: Prentice-Hall International, 1989.
7. A. Bazanella, M. Gevers, and L. Mišković, "Closed-loop identification of MIMO systems: a new look at identifiability and experiment design," in *To appear, European Control Conference*, Kos, Greece, July 2007.
8. L. Ljung, *System Identification: Theory for the User, 2nd Edition*. Englewood Cliffs, NJ: Prentice-Hall, 1999.
9. T. Kailath, *Linear Systems*. Englewood Cliffs, New Jersey: Prentice-Hall, 1980.
10. I. Mareels and M. Gevers, "Persistence of excitation criteria for linear, multi-variable, time-varying systems," *Mathematics of Control, Signals and Systems*, vol. 1, no. 3, pp. 203–226, 1988.
11. E. W. Bai and S. S. Sastry, "Persistence of excitation, sufficient richness and parameter convergence in discrete time adaptive control," *Systems and Control Letters*, vol. 6, pp. 153–163, 1985.
12. V. Kučera, *Discrete linear control: the polynomial approach*. John Wiley, 1979.