ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA

CONSIGLIO NAZIONALE DELLE RICERCHE

A. Germani, C. Manes, P. Palumbo

POLYNOMIAL FILTERING FOR STOCHASTIC NON-GAUSSIAN DESCRIPTOR SYSTEMS

R. 526   Maggio 2000

Alfredo Germani – Dipartimento di Ingegneria Elettrica, Università degli Studi dell'Aquila, 67040 Monteluco (L'Aquila), Italy and Istituto di Analisi dei Sistemi ed Informatica del CNR, Viale Manzoni 30, 00185 Roma, Italy. Email: `germani@iasi.rm.cnr.it`.

Costanzo Manes – Dipartimento di Ingegneria Elettrica, Università degli Studi dell'Aquila, 67040 Monteluco (L'Aquila), Italy and Istituto di Analisi dei Sistemi ed Informatica del CNR, Viale Manzoni 30, 00185 Roma, Italy. Email: `manes@ing.univaq.it`.

Pasquale Palumbo – Istituto di Analisi dei Sistemi ed Informatica del CNR, Viale Manzoni 30, 00185 Roma, Italy. Email: `palumbo@iasi.rm.cnr.it`.

Abstract

The class of stochastic descriptor systems, also named singular systems, has been widely investigated and many important results in the linear filtering theory have been achieved in the framework of Gaussian processes. Nevertheless, such results are far to be optimal when the state and measurement noises are not Gaussian. This paper investigates the estimation problem for stochastic singular systems affected by non-Gaussian noises, and proposes a polynomial filter based on the minimum variance criterion. The polynomial filter improves its performances by increasing its degree. The proposed filtering scheme can be well considered as a proper extension of the filter presented in [27], because its restriction to first order polynomials in the case of Gaussian noises gives exactly the same results of the maximum likelihood estimator developed in [27]. The improvement of the polynomial filter can be highly significative when the noises are strongly asymmetrically distributed, far to be Gaussian. Simulations support theoretical results.

Key words: Descriptor systems, non-Gaussian noise, minimum variance estimate, polynomial filtering

1. Introduction

In many engineering applications the mathematical model that represents the dynamic relationships among the variables describing a system is given by a set of linear equations in the following *descriptor* form (see Luenberger in [17] and [18]):

$$J(k+1)x(k+1) = A(k)x(k) + v_k, \qquad k \geq k_0, \qquad (1.1)$$

where:

i) $x(k) \in \mathbb{R}^n$ is the descriptor vector;

ii) $\{v_k, \quad k \geq k_0\}$ is an input sequence, $v_k \in \mathbb{R}^m$;

iii) $J(k), A(k)$ are matrices in $\mathbb{R}^{m \times n}$;

The implicit formulation of (1.1) contains, as a special case, the explicit standard form when $J(k)$ is the identity matrix; in other particular situations it is useful to retain the implicit formulation although $J(k)$ is a square, non-singular matrix. However there are many important classes of dynamical systems in which $J(k)$ is a singular or even a non square matrix. In these cases the descriptor form (1.1) is unavoidable. In literature a system in the form (1.1) is denoted as a *singular* or *descriptor system*.

Since the first works of Luenberger ([17] and [18]), a growing literature has been developed: main references concerning descriptor systems can be found in Campbell ([3] and [4]), or in the surveys of Lewis ([15]) and Verghese et al. ([32]). Here are briefly reported some of the most significant examples of descriptor systems:

- sometimes the first step to create a dynamical model for a system is the definition of a set of descriptor variables, associated to an equivalent set of suitable quantities, without wondering whether this is a minimal set. If some relations among the variables are purely static, which models the presence of constrains, the recursive equations are represented by an implicit singular form (see Crandall et al. in [5]);

- it can be useful to deduce dynamical relations among the descriptor variables without wondering whether they are causal or not (Willems [33]): the result is a non-causal system, described in a descriptor form. These techniques are fundamental when studying economic systems (the Leontief model, for instance, [16]) or 2-D systems, in which the dynamical equations do not evolve in a temporal domain, so that they are intrinsically non causal (see Luenberger in [17], [18] and Nikoukhah et al. in [23], [24] and [25]);

- a descriptor form can also be used in treating causal systems whose dynamical model is not completely known, modeling the fact that only a reduced set of relationships is available among a wider set of descriptor variables. In this case the singular system is just a part of a large-scale interconnected system (Sing et al., [31]). A system in a descriptor form can be also used to model the presence of unknown inputs (see Darouach et al. in [10]). This model is of particular interest when a failure on a system can be represented by means of an unknown forcing term (see e.g. [20]).

This work deals with the filtering problem for descriptor discrete-time systems, whose state and measurement equations are affected by non-Gaussian noises:

$$\begin{cases} J(k+1)x(k+1) = A(k)x(k) + B(k)u(k) + f_k, \\ y(k) = C(k)x(k) + g_k, \end{cases} \qquad k \geq k_0, \qquad (1.2)$$

where $u(k)$ is the control input, $y(k)$ is the measured output and f_k, g_k are respectively the state and observation noise sequences.

4.

The estimation problem concerning the descriptor vector of a system modeled in an implicit form has been widely investigated, and many important contributes have been given in literature. Early papers are due to Dai (see [6] and [7]) and are based on the following two steps:
- conditions are given that allow suitable transformation that changes the singular system into a non singular one, structurally non causal;
- the state of the new system is estimated by a Kalman filter.

The result is a linear filter, which recursively estimates the state. It provides the optimal solution, in the sense that in the presence of Gaussian noises, it gives the minimum variance estimate. The drawback of the proposed approach consists in its rather restrictive hypotheses that restricts its application to a small class of syngular systems. For instance, the algorithm does not apply to rectangular systems, and therefore it can not be applied to the important case of unknown inputs and of failure systems.

First steps in proposing filtering algorithms for rectangular systems have been made by Darouach et al. (see [9] and [10]): based on weaker hypotheses, they estimate the descriptor vector by minimizing a suitable functional. It can be classified as a least square approach; noises are not explicitly supposed Gaussian, although only second order noise moments are used to achieve the descriptor vector estimate. Here only an *estimability* condition is required (see [10] for more details): matrix $\begin{bmatrix} J^T & C^T \end{bmatrix}^T$ is assumed to be of full column rank.

Important results can be found in the works of Adams, Levy, Nikoukhaha and Willsky (see [23], [24], [25], [26] [27] and [28]); in [27], in particular, it is proposed a solution for the filtering problem for time-varying, linear, discrete-time, descriptor systems affected by Gaussian noises by using a maximum likelihood approach. Throughout the paper it is called the NLW filter, an acronym coming from the initials of the authors. Its main features are that:
- usual hypotheses concerning regularity or well-posedness have not to be considered, so that the NLW filter can be applied to a really wide class of time-varying singular systems;
- there are no restrictions concerning dimensions of matrices, in the sense that they can also change size with time k;
- the covariance matrix of the output noise may also be singular;
- it is consistent with respect to explicit standard systems, in the sense that if $J(k)$ is a fixed size identity matrix for any k, the NLW filter coincides with the Kalman filter, the optimal linear estimate filter for linear systems in explicit form;

A recent contribution [28] further extends the class of estimable descriptor systems, structurally involving future contributes (outputs, inputs and future dynamics). Also in this case the estimate is performed according to the maximum likelihood criterion and noises are assumed Gaussian.

This paper aims to treat the descriptor filtering problem in the presence of non-Gaussian noises. Motivations take place from previous works concerning explicit linear systems affected by non-Gaussian noises (see [11], [12] and [13]). A growing literature has shown an increasing interest in estimation problems related to non-Gaussian systems, particularly in the field of digital communications systems (see for instance [21] [22], [29] and Yaz in [34] and [35]).

Filtering a discrete-time linear system in explicit form, affected by Gaussian noises, is a problem solved by the well known Kalman filter. It is a linear and recursive algorithm that provides the optimal state estimate in the sense that it ensures the minimum error variance among all the Borel transformation of measured output. Moreover, in the presence of non-Gaussian noises, it still guarantees the best linear estimate (see [14] for more details).

From a geometrical point of view, the minimum variance state estimator is the projection

of the state onto the space of all Borel transformations of the output. In this framework the Kalman filter implements such projection when the noise is Gaussian. In the non Gaussian case the Kalman filter it only operates the projection of the state onto the space of linear transformations of the output.

A natural improvement over the Kalman filter performances is obtained by considering projections onto classes of output transformations larger than the linear ones. This technique has been successfully presented in [11], [12] and [13] where polynomial output transformations are considered. The developed filter provides the minimum variance estimate among all the fixed degree polynomial transformations of the measurements. Moreover, when the polynomial degree is unitary, it coincides with the Kalman filter (best linear estimate).

An approach based on suboptimal estimates involving polynomial filtering methodologies is not directly realizable for non-Gaussian singular systems. The main reason lies on the fact that almost all the contributes available in literature on the filtering of singular systems follow a maximum likelihood approach, that is not easily extendable for probability distribution that are not Gaussian.

The novelty of this paper is to introduce a minimum variance methodology in treating filtering problems for descriptor systems affected by non-Gaussian noises. In particular, by defining a suitable projection for the state estimate, it will be possible to construct polynomial filters which guarantee the improvement of the error variance when the order of the polynomial increases.

The paper is organized as follows: in section two the basic features concerning projections in estimation problems are considered; in section three the structure of the descriptor system and its fundamental properties concerning estimability are studied; in section four the minimum variance solution of the filtering problem is introduced and the polynomial estimation algorithm is proposed; in section five, as a particular case, the linear algorithm is studied, and its coincidence with the maximum likelihood linear one is pointed out. Finally, in section six, simulation results are reported that show high performances of the proposed filter.

2. Estimation as a projection

It is well known that the minimum variance estimate of a partially observed random variable can be considered as a projection onto the Hilbert space of the Borel functions of the observations. In this section suboptimal estimates are characterized as projections onto suitable $\mathcal{L}^2$ spaces, generalizing this procedure also to sequences of random variables.

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ a sub σ-algebra of $\mathcal{F}$ and $\mathcal{L}^2(\mathcal{G}, n)$ the Hilbert space of the n-dimensional, $\mathcal{G}$-measurable, random variables, with finite second order moments:

$$\mathcal{L}^2(\mathcal{G}, n) = \left\{ X : \Omega \mapsto I\!\!R^n, \ \mathcal{G}\text{-measurable}, \ \int_\Omega \|X(\omega)\|^2 dP(\omega) < +\infty \right\}.$$

If $\mathcal{G}$ is a σ-algebra generated by a random variable $Y : \Omega \mapsto I\!\!R^m$, the previous space is also written as $\mathcal{L}^2(Y, n)$. Such space can also be charachterized as the Hilbert space of the n-dimensional random variables, with finite second order moments, given by Borel functions of Y:

$$X(\omega) = f\big(Y(\omega)\big), \qquad f : I\!\!R^m \mapsto I\!\!R^n \qquad \text{Borel function.}$$

The minimum variance estimate of a random variable $X \in \mathcal{L}^2(\mathcal{F}, n)$ as a function of the observation vector $Y \in \mathcal{L}^2(\mathcal{F}, m)$ is given by the conditional expectation: $I\!\!E\big[X|Y\big]$. This

6.

operation is mathematically equivalent to project the random variable X onto $\mathcal{L}^2(Y, n)$, and is denoted by $\Pi[X|\mathcal{L}^2(Y, n)]$ or, in a more compact form, by $\Pi[X|Y]$:

$$\widehat{X} = \mathbb{E}[X|Y] = \Pi[X|\mathcal{L}^2(Y, n)] = \Pi[X|Y]. \tag{2.1}$$

If X and Y are jointly Gaussian, the minimum variance estimate is given by an affine transformation of the measurements:

$$\widehat{X} = m_X + \Psi_{XY} \Psi_Y^\dagger (Y - m_Y), \tag{2.2}$$

with $m_X = \mathbb{E}[X]$, $m_Y = \mathbb{E}[Y]$ mean vectors and $\Psi_{XY} = \mathbb{E}[(X - m_X)(Y - m_Y)^T]$, $\Psi_Y = \mathbb{E}[(Y - m_Y)(Y - m_Y)^T]$ covariance matrices. The symbol $\dagger$ denotes the pseudoinversion of a matrix, according to the Moore-Penrose definition (see [2]).

Expression (2.2) can also be considered as a linear transformation of the $(m+1)$-dimensional *extended measurements* vector $Y_e = \begin{bmatrix} 1 & Y^T \end{bmatrix}^T$.

Unfortunately, as it is well known, in general in the non-Gaussian case a nice formula, like expression (2.2), for the computation of the conditional expectation (2.1) can not be obtained. It is particularly useful, consequently, to look for estimates with simpler mathematical structure and with the lowest possible error variance. These are *suboptimal estimates*, i.e. estimates that are optimal in a given subclass of Borel functions of the observations. In this paper suboptimal estimates are defined and computed as the projections of the random variable X onto suitable subspaces of $\mathcal{L}^2(Y, n)$.

The simplest suboptimal estimate is the *optimal affine estimator*, in the sense that the minimum variance estimate of X is taken among all affine transformations of Y. This operation corresponds to project X onto the subspace of all the linear transformations of the extended measurements vector Y_e:

$$\widetilde{X}_1 = \Pi[X|L(Y_e, n)], \qquad L(Y_e, n) = \left\{ Z : \Omega \mapsto \mathbb{R}^n, \quad Z = AY_e, \quad A \in \mathbb{R}^{n \times (1+m)} \right\},$$

that can be computed by (2.2). In general, denoting with $\widetilde{X}$ the projection of X onto a subspace of $\mathcal{L}^2(Y, n)$, the variance related to the error $X - \widetilde{X}$ is not smaller than the one related to $X - \widehat{X}$: this is the meaning of *suboptimal estimate*.

Assume that the observation vector Y has finite 2μ-th order moments, for a given $\mu \in \mathbb{N}$, that is:

$$Y \in \mathcal{L}^{2\mu}(\mathcal{F}, m) = \left\{ Y : \Omega \mapsto \mathbb{R}^m, \ \mathcal{F}\text{-measurable}, \ \int_\Omega \|Y(\omega)\|^{2\mu} dP(\omega) < +\infty \right\},$$

The polynomial extension of degree μ of the linear suboptimal estimate (the best *μ-th degree polynomial estimate* of X) is obtained as the projection of X onto the subspace of all the μ-th degree polynomial transformations of the measurements Y: $\widetilde{X}_\mu = \Pi[X|\mathcal{P}_\mu(Y, n)]$, with

$$\mathcal{P}_\mu(Y, n) = \left\{ Z : \Omega \mapsto \mathbb{R}^n, \quad Z = \sum_{i=0}^\mu A_i Y^{[i]}, \quad A_i \in \mathbb{R}^{n \times m^i}, \quad Y \in \mathcal{L}^{2\mu}(\mathcal{F}, m) \right\},$$

where $Y^{[i]}$ stands for the i-th Kronecker power (see appendix for more details concerning Kronecker algebra). $\mathcal{P}_\mu(Y, n)$ can also be characterized as the space of all the linear transformations

of a μ-th degree extended measurements vector Y^μ,

$$Y^\mu = \begin{pmatrix} 1 \\ Y \\ Y^{[2]} \\ \vdots \\ Y^{[\mu]} \end{pmatrix} \in I\!\!R^{1+m+\cdots+m^\mu},$$

(note that $Y^1 = Y_e$) so that:

$$\mathcal{P}_\mu(Y,n) = L(Y^\mu,n) = \left\{ Z : \Omega \mapsto I\!\!R^n, \quad Z = AY^\mu, \quad A \in I\!\!R^{n \times (1+m+\cdots+m^\mu)} \right\} \subset \mathcal{L}^2(Y,n).$$

Owing to the following sequence of inclusions:

$$L(Y^1,n) \subset L(Y^2,n) \subset \cdots \subset L(Y^\mu,n) \subset \mathcal{L}^2(Y,n),$$

projecting onto subspaces of polynomials of higher order, the estimate of X improves, in terms of the error variance.

Moreover, coming from its definition, an expression of $\widetilde{X}_\mu = \Pi\big[X|L(Y^\mu,n)\big]$ is given by a (2.2)-like equation, using the μ-th degree extended measurements vector Y^μ instead of Y_e:

$$\widetilde{X}_\mu = \Pi\Big[X|L(Y^\mu,n)\Big] = m_X + \Psi_{XY^\mu}\Psi_{Y^\mu}^\dagger(Y^\mu - m_{Y^\mu}). \tag{2.3}$$

Now, suppose to have a random sequence $\big\{X(j), \ j \in I\!\!N\big\}$ to be filtered. In order to estimate the element $X(k)$ using the measurements $\big\{Y(i) \in \mathcal{L}^{2\mu}(\mathcal{F},m), \ i = 0,\ldots,k, \ \mu \in I\!\!N\big\}$, it is convenient to define an *augmented measurements* vector:

$$Y_k = \begin{pmatrix} Y(0) \\ Y(1) \\ \vdots \\ Y(k) \end{pmatrix} \in I\!\!R^{km},$$

so that a suboptimal μ-th degree polynomial estimate of $X(k)$ is given by the projection of $X(k)$ onto $L(Y_k^\mu,n)$:

$$\widetilde{X}_\mu(k) = \Pi\Big[X(k)|L\big(Y_k^\mu,n\big)\Big].$$

However, this method is highly inefficient, because the dimension of the space $L\big(Y_k^\mu,n\big)$ has a more-than-linear growth, and therefore no iterative estimation algorithm can be built up [12].

In order to overcome this difficulty, we consider the projection of the variable $X(k)$ onto the subclass of polynomial functions of the measurements defined below:

Definition 2.3. Let Δ,μ be integers in $I\!\!N$. Consider the random variable $Y_k^{\mu,\Delta}$ defined as follows

$$Y_k^{\mu,\Delta} = \begin{bmatrix} Y^{\mu,\Delta}(0) \\ \vdots \\ Y^{\mu,\Delta}(k) \end{bmatrix}, \tag{2.4}$$

8.

where

$$Y^{\mu,\Delta}(k) = \left\{ Y^{[l_0]}(0) \otimes \cdots \otimes Y^{[l_k]}(k), \quad 0 \le l_0 + \cdots + l_k \le \mu \right\}, \quad \text{for} \quad 0 \le k < \Delta,$$

$$Y^{\mu,\Delta}(k) = \left\{ Y^{[l_{k-\Delta}]}(k-\Delta) \otimes \cdots \otimes Y^{[l_{k-1}]}(k-1) \otimes Y^{[l_k]}(k), \quad 0 \le l_{k-\Delta} + \cdots + l_k \le \mu \right\},$$

$$\text{for} \quad k \ge \Delta.$$

(2.5)

The set $L\big(Y_k^{\mu,\Delta}, n\big)$ of the linear transformations of the random variable $Y_k^{\mu,\Delta}$ is denoted as a Δ-set of polynomials up to the μ-th degree associated to the measurements $\big\{Y(0), \cdots, Y(k)\big\}$.

The projection of $X(k)$ onto $L\big(Y_k^{\mu,\Delta}, n\big)$ is denoted the μ-th degree Δ-polynomial estimate.

Remark 2.4. The μ-th degree Δ-polynomial estimate can be put in a recursive form [12].

This geometric approach has led to important results concerning the filtering problems for non-Gaussian linear or bilinear explicit systems (see [12] and [13]). The following Sections develop the filtering theory for non-Gaussian, linear, singular systems.

3. Solvable linear singular systems

This section reports some concepts and properties of linear singular systems, denoted also desriptor systems, that are useful for the derivation of the polynomial filter theory. In particular, the class of solvable linear singular systems is introduced.

Definition 3.1. Given a sequence of matricx triples $\big\{J(k+1), A(k), F(k), \quad k \ge k_0\big\}$, with $J(k+1), A(k) \in I\!\!R^{m \times n}$ and $F(k) \in I\!\!R^{m \times p}$, a descriptor linear system is given by the following non-empty set of pairs:

$$\mathcal{S}_{k_0} = \Big\{ \big(x(k), v_k\big) \in I\!\!R^n \times I\!\!R^p, \quad k \ge k_0 : J(k+1)x(k+1) = A(k)x(k) + F(k)v_k, \quad k \ge k_0 \Big\}.$$

(3.1)

$x(k) \in I\!\!R^n$ is the *descriptor vector* and $\big\{v_k\big\}$ is the *input sequence*.

Definition 3.2. (Luenberger) A descriptor linear system $\mathcal{S}_{k_0}$ is *solvable*, if:

$$\forall \big\{ v_k, k \ge k_0 \big\}, \ v_k \in I\!\!R^p, \quad \exists \big\{ x(k), k \ge k_0 \big\}, \ x(k) \in I\!\!R^n :$$
$$\Big\{ \big(x(k), v_k\big), \quad k \ge k_0 \Big\} \in \mathcal{S}_{k_0}.$$

This definition has been proposed by Luenberger in [17], with a slightly different notation, for square singular systems. In this paper it is useful to define the following subclass of solvable descriptor linear systems.

Definition 3.3. A descriptor linear system $\mathcal{S}_{k_0}$ is V_{k_0}-*causally solvable*, if:

$$\exists V_{k_0} \subseteq I\!\!R^n : \forall x_0 \in V_{k_0}, \quad \forall \big\{ v_k \in I\!\!R^p, k \ge k_0 \big\}, \quad \exists \big\{ x(k) \in I\!\!R^n, k \ge k_0; \ x(k_0) = x_0 \big\} :$$
$$\Big\{ \big(x(k), v_k\big), \quad k \ge k_0 \Big\} \in \mathcal{S}_{k_0}.$$

When V_{k_0} coincides with $I\!\!R^n$, then $\mathcal{S}_{k_0}$ is said to be a causal solvable system.

Theorem 3.4. *If a descriptor linear system is V_{k_0}-causally solvable, then the subset V_{k_0} is a linear subspace of $\mathbb{R}^n$.*

Proof. It has to be proved that $\forall x_0^a, x_0^b \in V_{k_0}$, then also $x_0^c = \alpha x_0^a + \beta x_0^b$ is in V_{k_0}, $\forall \alpha, \beta \in \mathbb{R}$, that is, for any given input sequence $\{v_k \in \mathbb{R}^p, k \geq k_0\}$, exists $\{x^c(k), k \geq k_0\}$:

$$\begin{cases} x^c(k_0) = x_0^c, \\ J(k+1)x^c(k+1) = A(k)x^c(k) + F(k)v_k, \qquad \forall k \geq k_0. \end{cases} \tag{3.2}$$

Let $x_0^a, x_0^b \in V_{k_0}$ and $\{v_k \in \mathbb{R}^p, k \geq k_0\}$ a given input sequence. This means that:

$$\exists \{x^a(k) \in \mathbb{R}^n, k \geq k_0; \ x^a(k_0) = x_0^a\}: \qquad J(k+1)x^a(k+1) = A(k)x^a(k) + F(k)v_k, \quad \forall k \geq k_0, \tag{3.3}$$

$$\exists \{x^b(k) \in \mathbb{R}^n, k \geq k_0; \ x^b(k_0) = x_0^b\}: \qquad J(k+1)x^b(k+1) = A(k)x^b(k) + F(k)v_k, \quad \forall k \geq k_0. \tag{3.4}$$

Moreover, for any null input sequence, exist two sequences $\{\bar{x}^a(k), k \geq k_0\}$ and $\{\bar{x}^b(k), k \geq k_0\}$ associated respectively to x_0^a and x_0^b so that:

$$J(k+1)\bar{x}^a(k+1) = A(k)\bar{x}^a(k), \qquad \forall k \geq k_0, \tag{3.5}$$

$$J(k+1)\bar{x}^b(k+1) = A(k)\bar{x}^b(k), \qquad \forall k \geq k_0. \tag{3.6}$$

Now, let

$$x^c(k) = \frac{1}{2}x^a(k) + \left(\alpha - \frac{1}{2}\right)\bar{x}^a(k) + \frac{1}{2}x^b(k) + \left(\beta - \frac{1}{2}\right)\bar{x}^b(k). \tag{3.7}$$

Then $x^c(k_0) = \alpha x_0^a + \beta x_0^b = x_0^c$ and moreover:

$$\begin{aligned} J(k+1)x^c(k+1) &= \frac{1}{2}J(k+1)x^a(k+1) + \left(\alpha - \frac{1}{2}\right)J(k+1)\bar{x}^a(k+1) \\ &\quad + \frac{1}{2}J(k+1)x^b(k+1) + \left(\beta - \frac{1}{2}\right)J(k+1)\bar{x}^b(k+1) \\ &= \frac{1}{2}A(k)x^a(k) + \frac{1}{2}F(k)v_k + \left(\alpha - \frac{1}{2}\right)A(k)\bar{x}^a(k) \\ &\quad + \frac{1}{2}A(k)x^b(k) + \frac{1}{2}F(k)v_k + \left(\beta - \frac{1}{2}\right)A(k)\bar{x}^b(k) \\ &= A(k)x^c(k) + F(k)v_k, \end{aligned}$$

so that $x_0^c \in V_{k_0}$.

$\blacksquare$

Remark 3.5. Causally solvable singular systems are particularly suitable to model stochastic descriptor, linear systems, where $\{v_k\}$ and $x(k_0)$ are respectively a white noise sequence and a random variable.

Theorem 3.6. *A sufficient condition for causal solvability is that $J(k)$ is a full row rank matrix in $\mathbb{R}^{m \times n}$, that is:*

$$\text{rank}\Big[J(k)\Big] = m \leq n, \qquad \forall k > k_0. \tag{3.8}$$

Proof. Owing to the rank condition (3.8), according to the Rouchè-Capelli theorem, the descriptor equation:

$$J(k+1)x(k+1) = A(k)x(k) + F(k)v_k \tag{3.9}$$

has solutions $x(k+1)$ for any given $x(k) \in \mathbb{R}^n$ and any given $v_k \in \mathbb{R}^p$, for all $k \geq k_0$. A full row rank condition concerning $J(k)$ means that V_{k_0} is equal to $\mathbb{R}^n$. ∎

Remark 3.7. Solvability hypothesis guarantees the existence of solutions for the implicit system (3.9) for any given input sequence $\{v_k\}$ and, in case of causally solvable systems, also for any given admissible initial descriptor vector $x(k_0)$ in a given subspace of $\mathbb{R}^n$. Uniqueness needs further informations concerning the outputs of the system.

Definition 3.8. Consider the following V_{k_0}-causally solvable, singular, linear system, endowed with a measurement equation:

$$\begin{cases} J(k+1)x(k+1) = A(k)x(k) + F(k)v_k, \\ x(k_0) = \bar{x} \in V_{k_0}, \qquad\qquad\qquad\qquad\qquad k \geq k_0, \\ y(k) = C(k)x(k) + G(k)w_k, \end{cases} \tag{3.10}$$

with $x(k) \in \mathbb{R}^n$, $v_k \in \mathbb{R}^p$, $y(k) \in \mathbb{R}^q$, $w_k \in \mathbb{R}^r$ and $J(k), A(k) \in \mathbb{R}^{m \times n}$, $F(k) \in \mathbb{R}^{m \times p}$, $C(k) \in \mathbb{R}^{q \times n}$, $G(k) \in \mathbb{R}^{q \times r}$.

The regular system described by the following explicit form:

$$\begin{cases} \xi(k+1) = M(k)\xi(k) + N(k)v_k + T(k)y(k+1) + S(k)w_{k+1}, \\ \xi(k_0) = \bar{x}, \end{cases} \qquad k \geq k_0, \tag{3.11}$$

defined by a sequence of matrices:

$$\Big(M(k) \in \mathbb{R}^{n \times n}, N(k) \in \mathbb{R}^{n \times m}, T(k) \in \mathbb{R}^{n \times q}, S(k) \in \mathbb{R}^{n \times r}, \quad k \geq k_0 \Big) \tag{3.12}$$

is a *Complete Regular System* (*CRS*) for (3.10) if and only if $\forall\{v_k, k \geq k_0\}$, $\forall\{w_k, k \geq k_0\}$, it results, $\forall k \geq k_0$:

   i) $J(k+1)\xi(k+1) = A(k)\xi(k) + F(k)v_k$;
   ii) $y(k) = C(k)\xi(k) + G(k)w_k$.

The class of singular systems here considered is precisely that one described by the following definition [10].

Definition 3.9. (Darouach et al. [10]) Let (3.10) be a singular, V_{k_0}-causally solvable, linear system. It is said to be *estimable from the measurements* if the evolution of $x(k)$ is univocally determined by the output sequence $\{y(k), \ k \geq k_0\}$.

For the reader convenience it will be recalled the following important result about singular systems which will be useful for the sequel.

Theorem 3.10. (*Darouach et al.* [10]) *A singular, causally solvable, linear system* (3.10) *is estimable from the measurements if and only if the matrix* $\begin{bmatrix} J(k) \\ C(k) \end{bmatrix}$ *has full column rank, that is*

$$\text{rank} \begin{bmatrix} J(k) \\ C(k) \end{bmatrix} = n, \qquad \forall k > k_0. \tag{3.13}$$

In order to study filtering problems for singular systems it is also useful to reinterpret known results about singular systems as given by the following Proposition.

Proposition 3.11. *For each CRS* (3.11), *associated to a linear singular system* (3.10), *the evolution of* $\xi(k)$ *is invariant with respect to each sequence of matrices* (3.12) *if and only if the system is estimable from the measurements. In this case* $\xi(k)$ *is equal to the unique evolution of the descriptor vector.*

Proof. Estimability guarantees that there is only one sequence $\{\xi(k)\}$ compatible with the singular recursive equation i) and the measurement equation ii) and it necessarily coincides with the exact evolution of the singular system.

∎

Finally a useful characterization of a CRS associated to singular systems estimable from the measurements is reported by the following Theorem.

Theorem 3.12. *The class of CRS's associated to a causally solvable linear singular system, estimable from the measurements is given by the following sequence of matrices:*

$$M(k) = H^+(k) \begin{bmatrix} A(k) \\ O_{q \times n} \end{bmatrix}, \qquad N(k) = H^+(k) \begin{bmatrix} F(k) \\ O_{q \times p} \end{bmatrix},$$

$$T(k) = H^+(k) \begin{bmatrix} O_{m \times q} \\ I_q \end{bmatrix}, \qquad S(k) = H^+(k) \begin{bmatrix} O_{m \times r} \\ -G(k+1) \end{bmatrix}, \tag{3.14}$$

with $H^+(k)$ *any given sequence of left-inverse of* $H(k) = \begin{bmatrix} J(k+1) \\ C(k+1) \end{bmatrix}$, *i.e.* $H^+(k)H(k) = I_n$.

Proof. Let $\{\bar{x}(k), \ k \geq k_0\}$ be the sequence which produces the measurements, compatible with the singular equations of the system (3.10). Its uniqueness is guaranteed by the estimability condition. Let $\{\xi(k), \ k \geq k_0\}$ be the sequence associated to the system generated by (3.14). The equivalence between the sequences $\{\bar{x}(k)\}$ and $\{\xi(k)\}$ is given by induction. As trivially $\xi(k_0) = \bar{x}(k_0)$, it has to be shown that:

$$\xi(k) = \bar{x}(k) \implies \xi(k+1) = \bar{x}(k+1), \qquad \forall k > k_0.$$

Let $\xi(k) = \bar{x}(k)$ for a given $k \geq k_0$:

$$\xi(k+1) = M(k)\xi(k) + N(k)v_k + T(k)y(k+1) + S(k)w_{k+1}$$

$$= H^+(k) \begin{bmatrix} A(k) \\ O_{q \times n} \end{bmatrix} \bar{x}(k) + H^+(k) \begin{bmatrix} F(k) \\ O_{q \times p} \end{bmatrix} v_k + H^+(k) \begin{bmatrix} O_{m \times q} \\ I_q \end{bmatrix} C(k+1)\bar{x}(k+1)$$

$$+ H^+(k) \begin{bmatrix} O_{m \times q} \\ I_q \end{bmatrix} G(k+1)w_{k+1} - H^+(k) \begin{bmatrix} O_{m \times q} \\ I_q \end{bmatrix} G(k+1)w_{k+1}$$

$$= H^+(k) \begin{bmatrix} A(k)\bar{x}(k) + F(k)v_k \\ C(k+1)\bar{x}(k+1) \end{bmatrix} = H^+(k) \begin{bmatrix} J(k+1)\bar{x}(k+1) \\ C(k+1)\bar{x}(k+1) \end{bmatrix}$$

$$= H^+(k) \begin{bmatrix} J(k+1) \\ C(k+1) \end{bmatrix} \bar{x}(k+1) = H^+(k)H(k)\bar{x}(k+1) = \bar{x}(k+1).$$

■

4. Polynomial filtering

Consider a discrete-time, stochastic, linear, V_{k_0}-causally solvable, singular system described by the following recursive equations:

$$\begin{cases} J(k+1)x(k+1) = A(k)x(k) + B(k)u(k) + f_k, \\ x(k_0) = x_0, \\ y(k) = C(k)x(k) + g_k, \end{cases} \qquad k \geq k_0, \qquad (4.1)$$

where:

i) $x(k) \in I\!R^n$ is the descriptor vector, $u(k) \in I\!R^p$ is the control input and $y(k) \in I\!R^q$ is the measured output;

ii) $J(k), A(k) \in I\!R^{m \times n}$, $B(k) \in I\!R^{m \times p}$ and $C(k) \in I\!R^{q \times n}$;

iii) state and measurement noise sequences $\{f_k \in I\!R^m, \ k \geq k_0\}$ and $\{g_k \in I\!R^q, \ k \geq k_0\}$ are independent, zero-mean, white sequences (f_k and f_h are independent $\forall h \neq k$, and so are g_k and g_h; moreover f_k is independent of g_h $\forall k, h$).

iv) the following moments:

$$I\!E\left[f_k^{[i]}\right] = \zeta_f^i(k) \in I\!R^{m^i}, \qquad I\!E\left[g_k^{[i]}\right] = \zeta_g^i(k) \in I\!R^{q^i}, \qquad k \geq k_0, \qquad i = 1, \ldots, 2\nu,$$

are finite and available;

v) x_0 is a random variable of mean $\bar{x}$ assuming values in the subspace $V_{k_0} \subseteq I\!R^n$ and, together with the state and measurement noise sequences, it forms a set of independent random variables. Moreover, the following central moments

$$I\!E\left[\left(x_0 - \bar{x}\right)^{[i]}\right] = \zeta_0^i \in I\!R^{n^i}, \qquad i = 1, \ldots, 2\nu,$$

are finite and available.

Remark 4.1. Of course, the knowledge of the kind of the distribution allows the computation of any order moments. On the other hand, here it is requested just a finite number of them, which is, in general, a weaker information.

Theorem 4.2. *Let the stochastic system (4.1) be estimable from the measurements. Then a class of CRS's associated to (4.1) is given by the following:*

$$\begin{cases} x(k+1) = \mathcal{A}(k)x(k) + \mathcal{B}(k)u(k) + \mathcal{D}(k+1)y(k+1) + \mathcal{N}_k^f, \\ x(k_0) = x_0, \\ y(k) = C(k)x(k) + \mathcal{N}_k^g, \end{cases} \qquad k \geq k_0 \qquad (4.2)$$

with the matrices $\mathcal{A}(k) \in I\!R^{n \times n}$, $\mathcal{B}(k) \in I\!R^{n \times p}$ and $\mathcal{D}(k+1) \in I\!R^{n \times q}$, defined as below:

$$\mathcal{A}(k) = H^+(k)\begin{bmatrix} A(k) \\ O_{q \times n} \end{bmatrix}, \qquad \mathcal{B}(k) = H^+(k)\begin{bmatrix} B(k) \\ O_{q \times p} \end{bmatrix}, \qquad \mathcal{D}(k+1) = H^+(k)\begin{bmatrix} O_{m \times q} \\ I_q \end{bmatrix}, \qquad (4.3)$$

where $H^+(k)$ is any sequence of left-inverses of $H(k) = \begin{bmatrix} J(k+1) \\ C(k+1) \end{bmatrix}$, and where

$$\mathcal{N}_k^f = H^+(k) \begin{pmatrix} f_k \\ -g_{k+1} \end{pmatrix} \in I\!\!R^n, \qquad \mathcal{N}_k^g = g_k \in I\!\!R^q. \qquad (4.4)$$

Proof. It is obtained by direct substitution, following the same passages in the proof of Theorem 3.12.

∎

Remark 4.3. The new state noise sequence $\{\mathcal{N}_k^f,\ k \geq k_0\}$ is a zero mean, white sequence, but it is not independent of the measurement noise sequence $\{\mathcal{N}_k^g,\ k \geq k_0\}$, in that $\mathcal{N}_{k+1}^g$ is correlated to $\mathcal{N}_k^f$.

Remark 4.4. The regular system described in (4.2) is not strictly causal, in that the linear recursive equation needs $y(k+1)$ to get $x(k+1)$. Nevertheless it admits an interesting decomposition described by the following proposition.

Proposition 4.5. *System (4.2) can be split into two subsystems providing the descriptor variable as follows:*

$$x(k) = x_{nc}(k) + x_c(k), \qquad (4.5)$$

where

$$\begin{cases} x_{nc}(k+1) = \mathcal{A}(k)x_{nc}(k) + \mathcal{B}(k)u(k) + \mathcal{D}(k+1)y(k+1), & k \geq k_0 \\ x_{nc}(k_0) = \bar{x}, \end{cases} \qquad (4.6)$$

and

$$\begin{cases} x_c(k+1) = \mathcal{A}(k)x_c(k) + \mathcal{N}_k^f, \\ x_c(k_0) = x_0 - \bar{x}, & k \geq k_0, \\ y_c(k) = C(k)x_c(k) + \mathcal{N}_k^g, \end{cases} \qquad (4.7)$$

with $y_c(k) = y(k) - C(k)x_{nc}(k)$. $x_{nc}(k)$ is the non strictly causal component, whose evolution is deterministically given by measurements realizations and inputs, starting from the mean value of the initial state, while $x_c(k)$, the causal component of $x(k)$, evolves according to the stochastic equations (4.7).

Proof. The proof is an immediate consequence of linearity.

∎

From the previously described decomposition it can be noted that, being $x_{nc}(k)$ a linear function of the observations from k_0 to k, the best estimate of $x_{nc}(k)$, in the sense of minimum variance, is $x_{nc}(k)$ itself. On the other hand, in the general non-Gaussian case, there not exists a recursive filter for the optimal estimation of $x_c(k)$. From these facts in this section a suboptimal filter is considered in order to estimate $x_c(k)$.

Definition 4.6. A ν-th degree Δ-polynomial estimate for the singular system (4.1) is intended to be the following:

$$\hat{x}^{\nu,\Delta}(k) = x_{nc}(k) + \hat{x}_c^{\nu,\Delta}(k), \qquad (4.8)$$

14.

where $\hat{x}_c^{\nu,\Delta}(k) = \Pi\left[x_c(k)|L(Y_{ck}^{\nu,\Delta}, n)\right].$

Remark 4.7. Note that the error covariance matrix of $\hat{x}^{\nu,\Delta}(k)$ is such that

$$\mathrm{Cov}\left(x(k) - \hat{x}^{\nu,\Delta}(k)\right) = \mathrm{Cov}\left(x_c(k) - \hat{x}_c^{\nu,\Delta}(k)\right), \qquad (4.9)$$

and therefore it depends only on the causal component of the state.

As stressed in Remark 4.3 there is a one-step correlation between state and output noises of system (4.7). On the other hand, the polynomial filtering algorithm developed in [12] assumes independent state and output noises. This is the reason why the extended system defined below is needed.

Lemma 4.8. *The strictly causal system (4.7) can be put in the following extended state form:*

$$\begin{cases} \mathcal{X}_e(k+1) = \mathcal{A}_e(k)\mathcal{X}_e(k) + \mathcal{F}_e(k)\mathcal{N}_k, \\ \mathcal{X}_e(k_0) = \bar{\mathcal{X}}_e = \begin{pmatrix} x_c(k_0) \\ g_{k_0} \end{pmatrix}, \qquad\qquad k \geq k_0, \\ y_c(k) = \mathcal{C}_e(k)\mathcal{X}_e(k), \end{cases} \qquad (4.10)$$

with:

- $\mathcal{X}_e(k) = \begin{pmatrix} x_c(k) \\ \mathcal{N}_k^g \end{pmatrix} \in I\!R^\eta$ *the extended state, with $\eta = n + q$;*

- $\mathcal{N}_k = \begin{pmatrix} f_k \\ g_{k+1} \end{pmatrix} \in I\!R^\varrho$ *the extended noise, with $\varrho = m + q$;*

- *the matrices $\mathcal{A}_e(k) \in I\!R^{\eta \times \eta}$, $\mathcal{C}_e(k) \in I\!R^{q \times \eta}$ and $\mathcal{F}_e(k) \in I\!R^{\eta \times \varrho}$ are defined as follows:*

$$\mathcal{A}_e(k) = \begin{bmatrix} \mathcal{A}(k) & O_{n \times q} \\ O_{q \times n} & O_{q \times q} \end{bmatrix}, \qquad \mathcal{C}_e(k) = \begin{bmatrix} C(k) & I_q \end{bmatrix}, \qquad (4.11)$$

$$\mathcal{F}_e(k) = \begin{bmatrix} H^+(k) \begin{bmatrix} I_m \\ O_{q \times m} \end{bmatrix} & H^+(k) \begin{bmatrix} O_{m \times q} \\ -I_q \end{bmatrix} \\ O_{q \times m} & I_q \end{bmatrix}. \qquad (4.12)$$

Moreover, $\{\mathcal{N}_k, \ k \geq k_0\}$ is a white noise sequence, whose moments up to the 2ν-th order are given by:

$$I\!E\left[\mathcal{N}_k^{[i]}\right] = \sum_{r=0}^{i} \mathcal{M}_{m,q}^{r,i}\left(\zeta_f^r(k) \otimes \zeta_g^{i-r}(k+1)\right), \qquad k \geq k_0, \qquad (4.13)$$

with:

$$\mathcal{M}_{m,q}^{r,i} = M_r^i(\varrho)\left(\begin{bmatrix} I_m \\ O_{q \times m} \end{bmatrix}^{[r]} \otimes \begin{bmatrix} O_{m \times q} \\ I_q \end{bmatrix}^{[i-r]}\right) \in I\!R^{\varrho^i \times (m^r q^{i-r})}, \qquad (4.14)$$

where $M_r^i(\varrho) \in I\!R^{\varrho^i \times \varrho^i}$ are the binomial coefficient of a Kronecker power (see Lemma A.5 in Appendix).

Proof. System (4.10) comes simply considering the extended vectors and the matrices defined in (4.11) and (4.12). Whiteness, referred to the extended noise, easily descends from the independence between state and measurement noise of system (4.1). Finally, the computation of the moments of the noise $\mathcal{N}_k$ gives, according to Kronecker algebra:

$$
\begin{aligned}
\mathbb{E}\left[\mathcal{N}_k^{[i]}\right] &= \mathbb{E}\left[\left(\begin{array}{c} f_k \\ g_{k+1} \end{array}\right)^{[i]}\right] = \mathbb{E}\left[\left(\left[\begin{array}{c} I_m \\ O_{q\times m} \end{array}\right] f_k + \left[\begin{array}{c} O_{m\times q} \\ I_q \end{array}\right] g_{k+1}\right)^{[i]}\right] \\
&= \sum_{r=0}^{i} M_r^i(\varrho)\mathbb{E}\left[\left(\left[\begin{array}{c} I_m \\ O_{q\times m} \end{array}\right]^{[r]} f_k^{[r]}\right) \otimes \left(\left[\begin{array}{c} O_{m\times q} \\ I_q \end{array}\right]^{[i-r]} g_{k+1}^{[i-r]}\right)\right] \\
&= \sum_{r=0}^{i} M_r^i(\varrho)\left(\left[\begin{array}{c} I_m \\ O_{q\times m} \end{array}\right]^{[r]} \otimes \left[\begin{array}{c} O_{m\times q} \\ I_q \end{array}\right]^{[i-r]}\right) \cdot \mathbb{E}\left[f_k^{[r]} \otimes g_{k+1}^{[i-r]}\right] \\
&= \sum_{r=0}^{i} \mathcal{M}_{m,q}^{r,i}\mathbb{E}\left[f_k^{[r]}\right] \otimes \mathbb{E}\left[g_{k+1}^{[i-r]}\right] = \sum_{r=0}^{i} \mathcal{M}_{m,q}^{r,i}\left(\zeta_f^r(k) \otimes \zeta_g^{i-r}(k+1)\right).
\end{aligned}
$$

$\blacksquare$

Let $\Delta \in \mathbb{N}$. A Δ-polynomial estimate of $x_c(k)$ is so given by the first n components of $\mathcal{X}_e(k)$ projected onto the subspace of all the Δ-polynomial transformations of the measurements $\{y_c(j),\ j = k_0,\ldots,k\}$. In order to achieve a ν-th degree Δ-polynomial estimate of $\mathcal{X}_e(k)$ as a projection on $L(Y_k^{\nu,\Delta}, n)$, as described in section 2, the following vectors are introduced:

$$
\begin{cases}
\mathcal{X}_\Delta(k) = \begin{pmatrix} \mathcal{X}_e(k) \\ y_c(k-1) \\ \vdots \\ y_c(k-\Delta) \end{pmatrix}, & \Delta > 0, \\
\mathcal{X}_\Delta(k) = \mathcal{X}_e(k), & \Delta = 0,
\end{cases}
\qquad \mathcal{X}_\Delta(k) \in \mathbb{R}^\sigma, \qquad \sigma = \eta + q\Delta, \qquad (4.15)
$$

$$
\begin{cases}
\mathcal{Y}_\Delta(k) = \begin{pmatrix} y_c(k) \\ y_c(k-1) \\ \vdots \\ y_c(k-\Delta) \end{pmatrix}, & \Delta > 0, \\
\mathcal{Y}_\Delta(k) = y_c(k), & \Delta = 0,
\end{cases}
\qquad \mathcal{Y}_\Delta(k) \in \mathbb{R}^\gamma, \qquad \gamma = q(\Delta+1), \qquad (4.16)
$$

with $y_c(k) = 0$ for $k < k_0$.

Then:

$$
\begin{cases}
\mathcal{X}_\Delta(k+1) = \mathcal{A}_\Delta(k)\mathcal{X}_\Delta(k) + \mathcal{F}_\Delta(k)\mathcal{N}_k, \\
\mathcal{X}_\Delta(k_0) = \bar{\mathcal{X}}_\Delta, \\
\mathcal{Y}_\Delta(k) = \mathcal{C}_\Delta(k)\mathcal{X}_\Delta(k),
\end{cases}
\qquad (4.17)
$$

with the matrices $\mathcal{A}_\Delta \in \mathbb{R}^{\sigma\times\sigma}$, $\mathcal{C}_\Delta(k) \in \mathbb{R}^{\gamma\times\sigma}$ and $\mathcal{F}_\Delta(k) \in \mathbb{R}^{\sigma\times\varrho}$:

$$
\begin{cases}
\mathcal{A}_\Delta(k) = \begin{bmatrix}
\mathcal{A}_e(k) & O_{\eta\times q} & \cdots & \cdots & O_{\eta\times q} \\
\mathcal{C}_e(k) & O_{q\times q} & \cdots & \cdots & O_{q\times q} \\
O_{q\times\eta} & I_q & \ddots & & \vdots \\
\vdots & & \ddots & \ddots & \vdots \\
O_{q\times\eta} & \cdots & \cdots & I_q & O_{q\times q}
\end{bmatrix}, & \Delta > 0, \\
\mathcal{A}_\Delta(k) = \mathcal{A}_e(k), & \Delta = 0,
\end{cases}
\qquad (4.18)
$$

$$\begin{cases} \mathcal{C}_\Delta(k) = \begin{bmatrix} \mathcal{C}_e(k) & O_{q\times q} & \cdots & O_{q\times q} \\ O_{q\times\eta} & I_q & \ddots & \vdots \\ \vdots & \ddots & \ddots & O_{q\times q} \\ O_{q\times\eta} & \cdots & O_{q\times q} & I_q \end{bmatrix}, & \Delta > 0, \\ \mathcal{C}_\Delta(k) = \mathcal{C}_e(k), & \Delta = 0, \end{cases} \tag{4.19}$$

$$\begin{cases} \mathcal{F}_\Delta(k) = \begin{bmatrix} \mathcal{F}_e(k) \\ O_{q\Delta\times\varrho} \end{bmatrix}, & \Delta > 0, \\ \mathcal{F}_\Delta(k) = \mathcal{F}_e(k), & \Delta = 0, \end{cases} \tag{4.20}$$

and the initial state $\bar{\mathcal{X}}_\Delta \in I\!\!R^\sigma$:

$$\begin{cases} \bar{\mathcal{X}}_\Delta = \begin{pmatrix} \bar{\mathcal{X}}_e \\ 0 \end{pmatrix}, & \Delta > 0, \\ \bar{\mathcal{X}}_\Delta = \bar{\mathcal{X}}_e, & \Delta = 0. \end{cases}$$

In order to obtain a polynomial system, the following vectors have to be considered:

$$\mathcal{X}_\Delta^\nu(k) = \begin{pmatrix} \mathcal{X}_\Delta(k) \\ \mathcal{X}_\Delta^{[2]}(k) \\ \vdots \\ \mathcal{X}_\Delta^{[\nu]}(k) \end{pmatrix} \in I\!\!R^{\sigma_\nu}, \qquad \text{with} \quad \sigma_\nu = \sigma + \sigma^2 + \cdots + \sigma^\nu, \tag{4.21}$$

$$\mathcal{Y}_\Delta^\nu(k) = \begin{pmatrix} \mathcal{Y}_\Delta(k) \\ \mathcal{Y}_\Delta^{[2]}(k) \\ \vdots \\ \mathcal{Y}_\Delta^{[\nu]}(k) \end{pmatrix} \in I\!\!R^{\gamma_\nu}, \qquad \text{with} \quad \gamma_\nu = \gamma + \gamma^2 + \cdots + \gamma^\nu. \tag{4.22}$$

Theorem 4.9. *The processes* $\{\mathcal{X}_\Delta^\nu(k),\ k \geq k_0\}$ *and* $\{\mathcal{Y}_\Delta^\nu(k),\ k \geq k_0\}$, *defined in* (4.21) *and* (4.22) *satisfy the following equations:*

$$\begin{cases} \mathcal{X}_\Delta^\nu(k+1) = \mathcal{A}_\Delta^\nu(k)\mathcal{X}_\Delta^\nu(k) + \mathcal{U}_\Delta^\nu(k) + \xi_\Delta^\nu(k), \\ \mathcal{X}_\Delta^\nu(k_0) = \bar{\mathcal{X}}_\Delta^\nu, & k \geq k_0, \\ \mathcal{Y}_\Delta^\nu(k) = \mathcal{C}_\Delta^\nu(k)\mathcal{X}_\Delta^\nu(k), \end{cases} \tag{4.23}$$

where:

i) $\bar{\mathcal{X}}_\Delta^\nu = \begin{pmatrix} \bar{\mathcal{X}}_\Delta \\ \bar{\mathcal{X}}_\Delta^{[2]} \\ \vdots \\ \bar{\mathcal{X}}_\Delta^{[\nu]} \end{pmatrix} \in I\!\!R^{\sigma_\nu}$;

ii) matrices $\mathcal{A}_\Delta^\nu(k) \in I\!\!R^{\sigma_\nu \times \sigma_\nu}$ *and* $\mathcal{C}_\Delta^\nu(k) \in I\!\!R^{\gamma_\nu \times \sigma_\nu}$ *are defined as below:*

$$\mathcal{A}_\Delta^\nu(k) = \begin{bmatrix} H_{1,1}(k) & O_{\sigma\times\sigma^2} & \cdots & O_{\sigma\times\sigma^\nu} \\ H_{2,1}(k) & H_{2,2}(k) & \cdots & O_{\sigma^2\times\sigma^\nu} \\ \vdots & \vdots & \ddots & \vdots \\ H_{\nu,1}(k) & H_{\nu,2}(k) & \cdots & H_{\nu,\nu}(k) \end{bmatrix}, \tag{4.24}$$

$$
\mathcal{C}_\Delta^\nu(k) = \begin{bmatrix} \mathcal{C}_\Delta(k) & O_{\gamma\times\sigma^2} & \cdots & O_{\gamma\times\sigma^\nu} \\ O_{\gamma^2\times\sigma} & \mathcal{C}_\Delta^{[2]}(k) & \cdots & O_{\gamma^2\times\sigma^\nu} \\ \vdots & \vdots & \ddots & \vdots \\ O_{\gamma^\nu\times\sigma} & O_{\gamma^\nu\times\sigma^2} & \cdots & \mathcal{C}_\Delta^{[\nu]}(k) \end{bmatrix}, \tag{4.25}
$$

with:

$$
H_{i,l}(k) = M_{i-l}^i(\sigma)\Big(\mathcal{F}_\Delta^{[i-l]}(k) \otimes \mathcal{A}_\Delta^{[l]}(k)\Big) \cdot \Big(I\!E\Big[\mathcal{N}_k^{[i-l]}\Big] \otimes I_{\sigma,l}\Big) \in I\!\!R^{\sigma^i\times\sigma^l}; \tag{4.26}
$$

iii) $\mathcal{U}_\Delta^\nu(k), \xi_\Delta^\nu(k) \in I\!\!R^{\sigma^\nu}$ *are respectively deterministic and stochastic input sequences, whose expressions are given by:*

$$
\mathcal{U}_\Delta^\nu(k) = \begin{pmatrix} H_{1,0}(k) \\ H_{2,0}(k) \\ \vdots \\ H_{\nu,0}(k) \end{pmatrix}; \qquad \xi_\Delta^\nu(k) = \begin{pmatrix} \phi_1(k) \\ \phi_2(k) \\ \vdots \\ \phi_\nu(k) \end{pmatrix}, \quad \phi_i(k) \in I\!\!R^{\sigma^i}, \tag{4.27}
$$

with:

$$
\phi_i(k) = \sum_{l=0}^{i-1} M_{i-l}^i(\sigma)\Big(\mathcal{F}_\Delta^{[i-l]}(k) \otimes \mathcal{A}_\Delta^{[l]}(k)\Big) \cdot \Big(\Big(\mathcal{N}_k^{[i-l]} - I\!E\Big[\mathcal{N}_k^{[i-l]}\Big]\Big) \otimes I_{\sigma,l}\Big)\mathcal{X}_\Delta^{[l]}(k). \tag{4.28}
$$

Moreover $\xi_\Delta^\nu(k)$ *is a zero-mean, white sequence, whose covariance matrix is:*

$$
I\!E\Big[\xi_\Delta^\nu(k)\xi_\Delta^{\nu T}(k)\Big] = \mathcal{Q}(k) = \begin{bmatrix} \mathcal{Q}_{1,1}(k) & \mathcal{Q}_{1,2}(k) & \cdots & \mathcal{Q}_{1,\nu}(k) \\ \mathcal{Q}_{2,1}(k) & \mathcal{Q}_{2,2}(k) & \cdots & \mathcal{Q}_{2,\nu}(k) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{Q}_{\nu,1}(k) & \mathcal{Q}_{\nu,2}(k) & \cdots & \mathcal{Q}_{\nu,\nu}(k) \end{bmatrix}, \tag{4.29}
$$

with $\mathcal{Q}_{r,s}(k) \in I\!\!R^{\sigma^r\times\sigma^s}$, $1 \le r,s \le \nu$ *defined as:*

$$
\begin{aligned}
\mathcal{Q}_{r,s}(k) = \sum_{l=0}^{r-1}\sum_{m=0}^{s-1} M_{r-l}^r(\sigma)\Big(\mathcal{F}_\Delta^{[r-l]}(k) \otimes \mathcal{A}_\Delta^{[l]}(k)\Big)P_{l,m}^{r,s}(k) \\
\cdot \Big(\mathcal{F}_\Delta^{[s-m]}(k) \otimes \mathcal{A}_\Delta^{[m]}(k)\Big)^T \big(M_{s-m}^s(\sigma)\big)^T
\end{aligned} \tag{4.30}
$$

and $P_{l,m}^{r,s}(k)$ *is equal to:*

$$
\begin{aligned}
P_{l,m}^{r,s}(k) = st^{-1}\Bigg(\Big(I_{\varrho,s-m} \otimes C_{\varrho^{r-l},\sigma^m}^T \otimes I_{\eta,l}\Big) \\
\cdot \Big(\Big(I\!E\Big[\mathcal{N}_k^{[s+r-m-l]}\Big] - I\!E\Big[\mathcal{N}_k^{[s-m]}\Big] \otimes I\!E\Big[\mathcal{N}_k^{[r-l]}\Big]\Big) \otimes C_{1,\sigma^m} \otimes I_{\sigma,l}\Big) \\
\cdot I\!E\Big[\mathcal{X}_\Delta^{[l+m]}(k)\Big]\Bigg).
\end{aligned} \tag{4.31}
$$

Proof. The proof comes immediately from Theorem 3.3.2 of reference [12], in that all the theorem hypotheses are satisfied.

∎

18.

Remark 4.10. From expression (4.26) it follows that, $\forall i \geq 1$:

$$H_{1,0}(k) = 0, \qquad H_{i,0}(k) = \mathcal{F}_\Delta^{[i]}(k)I\!E\left[\mathcal{N}_k^{[i]}\right], \qquad H_{i,i}(k) = \mathcal{A}_\Delta^{[i]}(k). \qquad (4.32)$$

Remark 4.11. Moments up to the 2ν-th order, referred to $\mathcal{N}_k$ in (4.26), (4.28) and (4.31) can be computed using (4.13) and (4.14) in Lemma 4.8.

In order to compute the mean values of $\mathcal{X}_\Delta^{[i]}(k)$, $i = 0, \ldots, 2\nu$, in (4.31) the following lemma has to be considered:

Lemma 4.12. *Let the vector:*

$$\mu^{2\nu}(k) = \begin{pmatrix} \mu_1^{2\nu}(k) \\ \mu_2^{2\nu}(k) \\ \vdots \\ \mu_{2\nu}^{2\nu}(k) \end{pmatrix}, \qquad \mu_i^{2\nu}(k) = I\!E\left[\mathcal{X}_\Delta^{[i]}(k)\right] \in I\!R^{\sigma^i}. \qquad (4.33)$$

Then, $\mu^{2\nu}(k)$ evolves following the equations of the system:

$$\begin{cases} \mu^{2\nu}(k+1) = \mathcal{A}_\Delta^{2\nu}(k)\mu^{2\nu}(k) + \mathcal{U}_\Delta^{2\nu}(k), \\ \mu^{2\nu}(k_0) = \bar{\mu}^{2\nu}, \end{cases} \qquad k \geq k_0, \qquad (4.34)$$

where $\mathcal{A}_\Delta^{2\nu}(k)$ and $\mathcal{U}_\Delta^{2\nu}$ are defined as in Theorem 4.9 and:

$$\bar{\mu}_i^{2\nu} = I\!E\left[\bar{\mathcal{X}}_\Delta^{[i]}\right] = \begin{bmatrix} I_\eta \\ O_{q\Delta \times \eta} \end{bmatrix}^{[i]} \sum_{r=0}^{i} \mathcal{M}_{n,q}^{r,i}\left(\zeta_0^r \otimes \zeta_g^{i-r}(k_0)\right). \qquad (4.35)$$

Proof. Also in this case, the proof can be found in reference [12]. The expression of $\bar{\mu}_i^{2\nu}(k_0)$ is derived from:

$$\bar{\mathcal{X}}_\Delta = \begin{pmatrix} \bar{\mathcal{X}}_e \\ 0 \end{pmatrix} = \begin{bmatrix} I_\eta \\ O_{q\Delta \times \eta} \end{bmatrix} \bar{\mathcal{X}}_e \implies \bar{\mu}_i^{2\nu} = I\!E\left[\bar{\mathcal{X}}_\Delta^{[i]}\right] = \begin{bmatrix} I_\eta \\ O_{q\Delta \times \eta} \end{bmatrix}^{[i]} I\!E\left[\bar{\mathcal{X}}_e^{[i]}\right].$$

Applying (4.13) and (4.14) of Lemma 4.8, (4.35) comes.

∎

System (4.23) can be filtered using the Kalman algorithm, that provides the best estimate of the vector $\mathcal{X}_\Delta^\nu(k)$ among all linear transformations of the measurements $\mathcal{Y}_\Delta^\nu(k_0), \ldots, \mathcal{Y}_\Delta^\nu(k)$ or, that is the same, among all the ν-th degree, Δ-polynomials of the measurements $y_c(k_0), \ldots, y_c(k)$. The best Δ-polynomial estimate of $\mathcal{X}_e(k)$, i.e. $\Pi\left[\mathcal{X}_c(k)|L(Y_{c_k}^{\nu,\Delta}, \eta)\right]$, is given by the first η components of $\widehat{\mathcal{X}}_\Delta^\nu(k)$. Due to the Kalman filter structure, such a polynomial estimate is computed by a recursive algorithm.

Theorem 4.13. *Consider the following vectors:*

$$\widehat{X}_\Delta^\nu(k) = \begin{pmatrix} x_{nc}(k) \\ \widehat{\mathcal{X}}_\Delta^\nu(k) \end{pmatrix} \in I\!\!R^{n+\sigma_\nu} \tag{4.36}$$

and

$$Y_\Delta^\nu(k) = \begin{pmatrix} y(k) \\ \mathcal{Y}_\Delta^\nu(k) \end{pmatrix} \in I\!\!R^{q+\gamma_\nu}, \tag{4.37}$$

where $x_{nc}(k)$ is the non strictly causal component of the descriptor vector and $\widehat{\mathcal{X}}_\Delta^\nu(k)$ is the best linear estimate of $\mathcal{X}_\Delta^\nu(k)$, the state of system (4.23). Then $\hat{x}^{\nu,\Delta}(k)$, that is the ν-th degree, Δ-polynomial estimate of $x(k)$ as in Definition 4.6, is given by the output of the following system:

$$\begin{cases} \widehat{X}_\Delta^\nu(k+1) = \widehat{X}_\Delta^\nu(k+1|k) + K_\Delta^\nu(k+1)\Big[Y_\Delta^\nu(k+1) - C_\Delta^\nu(k+1)\widehat{X}_\Delta^\nu(k+1|k)\Big], \\ \widehat{X}_\Delta^\nu(k+1|k) = A_\Delta^\nu(k)\widehat{X}_\Delta^\nu(k) + U_\Delta^\nu(k), \\ \widehat{X}_\Delta^\nu(k_0|k_0-1) = \bar{X}_\Delta^\nu = \begin{pmatrix} \bar{x} \\ \bar{\mu}^\nu \end{pmatrix}, \\ \hat{x}^{\nu,\Delta}(k) = \mathcal{R}_\Delta^\nu \widehat{X}^\nu(k), \end{cases} \qquad k \geq k_0, \tag{4.38}$$

where:

$$A_\Delta^\nu(k) = \begin{bmatrix} \mathcal{A}(k) & O_{n\times\sigma_\nu} \\ O_{\sigma_\nu\times n} & \mathcal{A}_\Delta^\nu(k) \end{bmatrix} \in I\!\!R^{(n+\sigma_\nu)\times(n+\sigma_\nu)}, \tag{4.39}$$

$$C_\Delta^\nu(k) = \begin{bmatrix} O_{q\times n} & O_{q\times\sigma_\nu} \\ O_{\gamma_\nu\times n} & \mathcal{C}_\Delta^\nu(k) \end{bmatrix} \in I\!\!R^{(q+\gamma_\nu)\times(n+\sigma_\nu)}, \tag{4.40}$$

$$U_\Delta^\nu(k) = \begin{pmatrix} \mathcal{B}(k)u(k) \\ \mathcal{U}_\Delta^\nu(k) \end{pmatrix} \in I\!\!R^{n+\sigma_\nu}, \qquad \mathcal{R}_\Delta^\nu = \begin{bmatrix} I_n & I_n & O_{n\times(\sigma_\nu-n)} \end{bmatrix} \in I\!\!R^{n\times(n+\sigma_\nu)}, \tag{4.41}$$

the gain matrix $K_\Delta^\nu(k) \in I\!\!R^{(n+\sigma_\nu)\times(q+\gamma_\nu)}$:

$$\begin{cases} K_\Delta^\nu(k) = \begin{bmatrix} \mathcal{D}(k) & O_{n\times\gamma_\nu} \\ O_{\sigma_\nu\times q} & \mathcal{K}_\Delta^\nu(k) \end{bmatrix}, & k > k_0, \\ K_\Delta^\nu(k_0) = \begin{bmatrix} O_{n\times q} & O_{n\times\gamma_\nu} \\ O_{\sigma_\nu\times q} & \mathcal{K}_\Delta^\nu(k_0) \end{bmatrix}, \end{cases} \tag{4.42}$$

and the following Riccati equations:

$$\begin{cases} \mathcal{P}_P^{\nu,\Delta}(k+1) = \mathcal{A}_\Delta^\nu(k)\mathcal{P}_\Delta^\nu(k)\mathcal{A}_\Delta^{\nu T}(k) + \mathcal{Q}(k), \\ \mathcal{K}_\Delta^\nu(k+1) = \mathcal{P}_P^{\nu,\Delta}(k+1)\mathcal{C}_\Delta^{\nu T}(k+1)\Big[\mathcal{C}_\Delta^\nu(k+1)\mathcal{P}_P^{\nu,\Delta}(k+1)\mathcal{C}_\Delta^{\nu T}(k+1)\Big]^\dagger, \\ \mathcal{P}_\Delta^\nu(k+1) = \Big[I_{\sigma_\nu} - \mathcal{K}_\Delta^\nu(k+1)\mathcal{C}_\Delta^\nu(k+1)\Big]\mathcal{P}_P^{\nu,\Delta}(k+1), \\ \mathcal{P}_P^{\nu,\Delta}(k_0) = \text{Cov}(\bar{\mathcal{X}}_\Delta^\nu), \end{cases} \qquad k \geq k_0. \tag{4.43}$$

Proof. It comes applying the Kalman filter to system (4.23) in order to estimate $\mathcal{X}_\Delta^\nu(k)$ among the linear transformations of $\{\mathcal{Y}_\Delta^\nu(k_0), \ldots, \mathcal{Y}_\Delta^\nu(k)\}$. Using Definition 4.6, the filter is obtained adding the non causal component $x_{nc}(k)$ to the first n components of $\widehat{\mathcal{X}}_\Delta^\nu(k)$.

$\blacksquare$

20.

Remark 4.14. The initial one-step prediction error covariance is equal to:

$$\mathcal{P}_P^{\nu,\Delta}(k_0) = I\!\!E\left[\left(\bar{\mathcal{X}}_\Delta^\nu - I\!\!E[\bar{\mathcal{X}}_\Delta^\nu]\right)\left(\bar{\mathcal{X}}_\Delta^\nu - I\!\!E[\bar{\mathcal{X}}_\Delta^\nu]\right)^T \right] = I\!\!E\left[st^{-1}\left(\left(\bar{\mathcal{X}}_\Delta^\nu - I\!\!E[\bar{\mathcal{X}}_\Delta^\nu]\right)^{[2]}\right)\right]$$

$$= st^{-1}\left(I\!\!E\left[\left(\bar{\mathcal{X}}_\Delta^\nu - I\!\!E[\bar{\mathcal{X}}_\Delta^\nu]\right)^{[2]}\right]\right) = st^{-1}\left(I\!\!E\left[\bar{\mathcal{X}}_\Delta^{\nu[2]}\right] - \left(I\!\!E\left[\bar{\mathcal{X}}_\Delta^\nu\right]\right)^{[2]}\right)$$

$$= st^{-1}\left(\bar{\mu}^{2\nu} - \bar{\mu}^{\nu[2]}\right).$$

Remark 4.15. The use of the pseudo-inverse for the computation of the Kalman gain $\mathcal{K}_\Delta^\nu(k+1)$ of (4.43) instead of standard inversion, is necessary when $\nu > 1$; in this case the matrix $\mathcal{C}_\Delta^\nu(k+1)\mathcal{P}_P^{\nu,\Delta}(k+1)\mathcal{C}_\Delta^{\nu T}(k+1)$ is singular, due to the redundancy of the components of the vector $\mathcal{Y}_\Delta^\nu(k+1)$.

5. The linear estimate: minimum variance versus maximum likelihood

This Section is devoted to the analysis of the behavior of the proposed filter with respect to the behavior of filtering algorithms existing in literature for descriptor systems. Since only linear algorithms can be found in literature, the comparative analysis will be done by considering only the first order polynomial filter. Some of the main important methods for filtering, as mentioned in the introduction, have a statistical meaning, precisely that ones based on the maximum likelihood ([27], [28]) for which the Gaussian hypothesis is unavoidable; on the other hand, some other methods that are based on the minimization of a suitable defined functional do not need Gaussianity but unfortunately they have not a precise statistical meaning [8]. The proposed linear algorithm achieves the goal of giving an answer to the optimal linear filtering of non-Gaussian singular systems, in that it provides the minimum variance solution, according to Definition 4.6. It will be proved that, under the extra assumption of Gaussianity, it coincides with the maximum likelihood filter [27], so that it can be considered as a proper extension of it.

In the following, among all CRS's associated to the descriptor system (4.1) that are characterized in Theorem 4.2 using any right-inverse of H, the one defined by the Moore-Penrose pseudoinverse of H as a particular right-inverse will be considered. That means that in the following the matrices $\mathcal{A}(k)$, $\mathcal{B}(k)$, $\mathcal{D}(k)$ and the sequences N_k^f, $\mathcal{N}_k^g$, defined in Theorem 4.2, must be intended with $H^+ = H^\dagger$.

Theorem 5.1. *The linear estimation of the descriptor vector $x(k)$ of the singular system* (4.1), *that is $\hat{x}(k) = x_{nc}(k) + \hat{x}_c(k)$ according to Definition* 4.6, *is given by the following filter:*

$$\begin{cases} \hat{x}(k+1) = \hat{x}(k+1|k) + \mathcal{D}(k+1)y(k+1) + \\ \qquad + K(k+1)\Big[y(k+1) - C(k+1)\mathcal{D}(k+1)y(k+1) - C(k+1)\hat{x}(k+1|k)\Big], \\ \hat{x}(k+1|k) = \mathcal{A}(k)\hat{x}(k) + \mathcal{B}(k)u(k), \qquad\qquad\qquad k > k_0 \qquad (5.1) \\ \hat{x}(k_0) = \hat{x}(k_0|k_0-1) + K(k_0)\Big[y(k_0) - C(k_0)\hat{x}(k_0|k_0-1)\Big] \\ \hat{x}(k_0|k_0-1) = \bar{x}, \end{cases}$$

with the gain $K(k)$, the filter error covariance matrix $P(k) = \mathrm{Cov}\big(x(k) - \hat{x}(k)\big)$ and the one-step prediction error covariance matrix $P_P(k) = \mathrm{Cov}\big(x(k) - \hat{x}(k|k-1)\big)$ given by:

$$
\begin{cases}
M_P(k+1) = \begin{bmatrix} A(k)P(k)A^T(k) + Q_f(k) & O_{m\times q} \\ O_{q\times m} & Q_g(k+1) \end{bmatrix}, \\[4pt]
K(k+1) = -H^\dagger(k)M_P(k+1)L^T(k)\Big[L(k)M_P(k+1)L^T(k)\Big]^\dagger, \\[4pt]
P(k+1) = \big[H^\dagger(k) + K(k+1)L(k)\big]\, M_P(k+1)H^{\dagger T}(k), \\[4pt]
P_P(k+1) = H^\dagger(k)M_P(k+1)H^{\dagger T}(k), \\[2pt]
P_P(k_0) = \Psi_{x_0} \\[4pt]
K(k_0) = \Psi_{x_0}C^T(k_0)\Big[C(k_0)\Psi_{x_0}C^T(k_0) + Q_g(k_0)\Big]^\dagger \\[4pt]
P(k_0) = \big[I_n - K(k_0)C(k_0)\big]\Psi_{x_0},
\end{cases}
\qquad k \geq k_0 \qquad (5.2)
$$

with:

$$
L(k) = \begin{bmatrix} O_{q\times m} & I_q \end{bmatrix} \cdot \Big(I_{m+q} - H(k)H^\dagger(k)\Big) \in I\!R^{q\times(m+q)} \qquad (5.3)
$$

Matrices $\{Q_f(k), Q_g(k),\ k \geq k_0\}$ are, respectively, the covariance matrices of state and measurement noises, while Ψ_{x_0} is the a priori covariance matrix of the initial state.

Proof. Following Definition 4.6, a linear estimation for a singular system at time k is given by the sum of the non causal component $x_{nc}(k)$ and the best linear estimate of the strictly causal component $x_c(k)$ among all the linear transformations of the causal measurements $y_c(k_0),\ldots,y_c(k)$. Owing to the fact that $\nu = 1$, then $\Delta = 0$.

Applying Theorem 4.13, the filter (4.38) can be written as:

$$
\begin{cases}
\widehat{X}_0^1(k+1) = \widehat{X}_0^1(k+1|k) + K_0^1(k+1)\Big[Y_0^1(k+1) - C_0^1(k+1)\widehat{X}_0^1(k+1|k)\Big], \\[4pt]
\widehat{X}_0^1(k+1|k) = A_0^1(k)\widehat{X}_0^1(k) + U_0^1(k), \\[4pt]
\widehat{X}_0^1(k_0|k_0-1) = \bar{X}_0^1 = \begin{pmatrix} \bar{x} \\ \bar{\mu}^1 \end{pmatrix}, \\[4pt]
\hat{x}(k) = \hat{x}^{1,0}(k) = \mathcal{R}_0^1\widehat{X}_0^1(k),
\end{cases}
\qquad k \geq k_0,
$$

$$(5.4)$$

with:

$$
\widehat{X}_0^1(k) = \begin{pmatrix} x_{nc}(k) \\ \widehat{\mathcal{X}}_0^1(k) \end{pmatrix} = \begin{pmatrix} x_{nc}(k) \\ \widehat{\mathcal{X}}_e(k) \end{pmatrix} \in I\!R^{n+\eta}, \qquad Y_0^1(k) = \begin{pmatrix} y(k) \\ y_c(k) \end{pmatrix} \in I\!R^{2q} \qquad (5.5)
$$

and:

$$
A_0^1(k) = \begin{bmatrix} \mathcal{A}(k) & O_{n\times\eta} \\ O_{\eta\times n} & \mathcal{A}_e(k) \end{bmatrix} \in I\!R^{(n+\eta)\times(n+\eta)}, \qquad C_0^1(k) = \begin{bmatrix} O_{q\times n} & O_{q\times\eta} \\ O_{q\times n} & \mathcal{C}_e(k) \end{bmatrix} \in I\!R^{2q\times(n+\eta)}, \quad (5.6)
$$

$$
U_0^1(k) = \begin{pmatrix} \mathcal{B}(k)u(k) \\ 0 \end{pmatrix} \in I\!R^{n+\eta}, \qquad \mathcal{R}_0^1 = \begin{bmatrix} I_n & I_n & O_{n\times q} \end{bmatrix} \in I\!R^{n\times(n+\eta)}, \qquad (5.7)
$$

so that, premultiplying the filter equations by $\mathcal{R}_0^1$, equations (5.1) easily come with the Kalman gain $K(k) = \begin{bmatrix} I_n & O_{n\times q} \end{bmatrix} \mathcal{K}_0^1(k)$. According to (4.43), $\mathcal{K}_0^1(k)$ is given by the following Riccati equations:

$$\begin{cases} \mathcal{P}_P^{1,0}(k+1) = \mathcal{A}_e(k)\mathcal{P}_0^1(k)\mathcal{A}_e^T(k) + \mathcal{F}_e(k)\mathcal{W}(k)\mathcal{F}_e^T(k), \\ \mathcal{K}_0^1(k+1) = \mathcal{P}_P^{1,0}(k+1)\mathcal{C}_e^T(k+1)\Big[\mathcal{C}_e(k+1)\mathcal{P}_P^{1,0}(k+1)\mathcal{C}_e^T(k+1)\Big]^\dagger, \\ \mathcal{P}_0^1(k+1) = \Big[I_\eta - \mathcal{K}_0^1(k+1)\mathcal{C}_e(k+1)\Big]\mathcal{P}_P^{1,0}(k+1), \\ \mathcal{P}_P^{1,0}(k_0) = \Psi_{k_0}, \end{cases} \qquad k \geq k_0. \qquad (5.8)$$

$\mathcal{W}(k)$ is the covariance matrix of the extended noise $\mathcal{N}_k$ which, owing to the independence between state and measurements noises, can be written as:

$$\mathcal{W}(k) = \mathrm{Cov}(\mathcal{N}_k) = st^{-1}\Big(I\!\!E\big[\mathcal{N}_k^{[2]}\big]\Big) = \begin{bmatrix} Q_f(k) & O_{m\times q} \\ O_{q\times m} & Q_g(k+1) \end{bmatrix}, \qquad (5.9)$$

with $Q_f(k) = \mathrm{Cov}(f_k)$ and $Q_g(k) = \mathrm{Cov}(g_k)$, as previously written.

Ψ_{k_0} is the a priori covariance matrix of the extended initial state, so that:

$$\Psi_{k_0} = \mathrm{Cov}(\bar{\mathcal{X}}_e) = st^{-1}\Big(I\!\!E\big[\bar{\mathcal{X}}_e^{[2]}\big]\Big) = \begin{bmatrix} \Psi_{x_0} & O_{n\times q} \\ O_{q\times n} & Q_g(k_0) \end{bmatrix}. \qquad (5.10)$$

In order to achieve the Riccati equations in the form (5.2), the following partition is considered:

$$\mathcal{P}_P^{1,0}(k) = \begin{bmatrix} P_P(k) & S_P(k) \\ S_P^T(k) & W_P(k) \end{bmatrix}, \qquad \mathcal{P}_0^1(k) = \begin{bmatrix} P(k) & S(k) \\ S^T(k) & W(k) \end{bmatrix}. \qquad (5.11)$$

where $P_P(k), P(k) \in I\!\!R^{n\times n}$, $S_P(k), S(k) \in I\!\!R^{n,q}$ and $W_P(k), W(k) \in I\!\!R^{q\times q}$.

After substitution in (5.8), direct computations lead to Riccati equations (5.2). ∎

Remark 5.2. The estimability condition (3.13) needs a number of measurements at least equal to $n - m$. In the case $q = n - m$, it can be seen that $L(k)$ is a null matrix, $\forall k \geq k_0$. This happens because $H(k)$ is square and non singular and from (5.3):

$$L(k)H(k) = \begin{bmatrix} O_{q\times m} & I_q \end{bmatrix} \cdot \Big(I_{m+q} - H(k)H^\dagger(k)\Big)H(k)$$

$$= \begin{bmatrix} O_{q\times m} & I_q \end{bmatrix} \cdot \Big(H(k) - H(k)H^\dagger(k)H(k)\Big) = O_{q\times n}.$$

Lemma 5.3. *If $m + q = n$, then the filter equation (5.1) reduces to:*

$$\hat{x}(k+1) = \mathcal{A}(k)\hat{x}(k) + \mathcal{B}(k)u(k) + \mathcal{D}(k+1)y(k+1), \qquad k > k_0, \qquad (5.12)$$

and the filtered estimate $\hat{x}(k)$ is statistically equivalent to the one-step prediction $\hat{x}(k|k-1)$.

Proof. From Remark 5.2 it follows that $L(k) = O_{q\times n}$, $\forall k \geq k_0$, and consequently $K(k) = O_{n\times q}$, $\forall k > k_0$, as it comes from the Riccati equations in 5.2, so that (5.12) is obtained and moreover, $P(k) = P_P(k)$. ∎

In order to prove the above announced equivalence we need to state the following three lemmas, in which the dependence on k is dropped for brevity.

Lemma 5.4. *The following relations, concerning the matrices L and $I_{m+q} - HH^\dagger$, stand:*

    i) $L = L(I_{m+q} - HH^\dagger)$;

    ii) $I_{m+q} - HH^\dagger = L^\dagger L$;

Proof.

    i) The proof is a direct consequence of the fact that $I_{m+q} - HH^\dagger$ is an orthogonal projector onto $\mathcal{N}(H^T)$:

$$L = \begin{bmatrix} O_{q\times m} & I_q \end{bmatrix}(I_{m+q} - HH^\dagger) = \begin{bmatrix} O_{q\times m} & I_q \end{bmatrix}(I_{m+q} - HH^\dagger)(I_{m+q} - HH^\dagger)$$
$$= L(I_{m+q} - HH^\dagger).$$

    ii) The item is proved if it is verified that the projectors $I_{m+q} - HH^\dagger$ and $L^\dagger L$ have the same null space. According to the first item, already proved:

$$x \in \mathcal{N}(I_{m+q} - HH^\dagger) \implies (I_{m+q} - HH^\dagger)x = 0 \implies Lx = L(I_{m+q} - HH^\dagger)x = 0,$$

so that: $\mathcal{N}(I_{m+q} - HH^\dagger) \subseteq \mathcal{N}(L^\dagger L) = \mathcal{N}(L)$. The two null spaces are equal if $\text{rank}(L) = \text{rank}(I_{m+q} - HH^\dagger)$. Let $I_{m+q} - HH^\dagger$ be partitioned as:

$$I_{m+q} - HH^\dagger = \begin{bmatrix} G \\ L \end{bmatrix}, \qquad \text{with} \quad G \in \mathbb{R}^{m\times(m+q)} \text{ and } L \text{ as in (5.3)}.$$

The aim is to show that the first m rows of $I_{m+q} - HH^\dagger$, that is matrix G, are dependent on the last q rows, that is L. From the hypothesis concerning the full row rank condition for J (see Section 3 on causally solvable, singular systems), and from the identity: $H^T(I_{m+q} - HH^\dagger) = O_{n\times(m+q)}$ (see [2] for more details), it comes:

$$J^T G + C^T L = O_{n\times(m+q)} \implies G = -(JJ^T)^{-1}JC^T L,$$

                                                                  ■

Lemma 5.5. *Let Q_1 and Q_2 be a pair of symmetric, non-negative definite matrices in $\mathbb{R}^{q\times q}$. If $\mathcal{R}(Q_1) = \mathcal{R}(Q_2)$, then:*

$$Q_2^\dagger - Q_1^\dagger = -Q_2^\dagger(Q_2 - Q_1)Q_1^\dagger. \tag{5.13}$$

    Proof. Let $r = \text{rank}(Q_1) = \text{rank}(Q_2) \leq q$. Owing to the hypothesis concerning the ranges of Q_1 and Q_2, there can be found a couple of non-singular matrices $\widetilde{Q}_1, \widetilde{Q}_2 \in \mathbb{R}^{r\times r}$ and a full-rank matrix $U \in \mathbb{R}^{q\times r}$, so that:

$$Q_1 = U\widetilde{Q}_1 U^T, \qquad Q_2 = U\widetilde{Q}_2 U^T, \qquad U^T U = I_r. \tag{5.14}$$

Of course, from (5.14) follows:

$$\widetilde{Q}_1 = U^T Q_1 U, \qquad \widetilde{Q}_2 = U^T Q_2 U. \tag{5.15}$$

It is also easy to verify, using the Penrose definition that:

$$Q_1^\dagger = U\widetilde{Q}_1^{-1}U^T, \qquad Q_2^\dagger = U\widetilde{Q}_2^{-1}U^T, \qquad \implies \qquad Q_2^\dagger - Q_1^\dagger = U\left(\widetilde{Q}_2^{-1} - \widetilde{Q}_1^{-1}\right)U^T, \tag{5.16}$$

so that:

$$-Q_2^\dagger(Q_2 - Q_1)Q_1^\dagger = -U\widetilde{Q}_2^{-1}U^T\left(U(\widetilde{Q}_2 - \widetilde{Q}_1)U^T\right)U\widetilde{Q}_1^{-1}U^T = -U\left(\widetilde{Q}_1^{-1} - \widetilde{Q}_2^{-1}\right)U^T$$
$$= Q_2^\dagger - Q_1^\dagger.$$

                                                                  ■

Lemma 5.6. *Let Π be a projector and A a square matrix of suitable dimension. Then:*

$$\left(\Pi A\right)^{\dagger}\Pi A = \left(\Pi A\right)^{\dagger}A. \tag{5.17}$$

Proof. The proof is achieved by using the following property concerning the pseudo-inverses of a product (see [2] for more details):

$$\text{if} \quad Q^{T}QBB^{T} = BB^{T}Q^{T}Q \quad \text{then} \quad \left(QB\right)^{\dagger} = B^{\dagger}Q^{\dagger}. \tag{5.18}$$

Thanks to (5.18), used with $Q = \Pi$ and $B = \Pi A$, and recalling that $\Pi = \Pi^{T} = \Pi\Pi^{T} = \Pi^{\dagger}$, it follows

$$\left(\Pi A\right)^{\dagger} = \left(\Pi\Pi A\right)^{\dagger} = \left(\Pi A\right)^{\dagger}\Pi.$$

Postmultiplying by A, the thesis follows.

∎

Theorem 5.7. *The filter* (5.1) *is equivalent to the maximum likelihood filter proposed by Nikoukhah et al. in [27], namely:*

$$\begin{cases} \check{x}(k+1) = \begin{bmatrix} O_{n\times(m+q)} & I_{n} \end{bmatrix} \begin{bmatrix} \check{M}_{P}(k+1) & H(k) \\ H^{T}(k) & O_{n\times n} \end{bmatrix}^{\dagger} \begin{pmatrix} \check{\xi}(k+1|k) \\ 0 \end{pmatrix}, \\[4mm] \check{\xi}(k+1|k) = \begin{pmatrix} A(k)\check{x}(k) + B(k)u(k) \\ y(k+1) \end{pmatrix}, & k > k_{0}, \\[4mm] \check{x}(k_{0}) = \begin{bmatrix} O_{n\times n} & O_{n\times q} & I_{n} \end{bmatrix} \begin{bmatrix} \Psi_{x_{0}} & O_{n\times q} & I_{n} \\ O_{q\times n} & Q_{g}(k_{0}) & C(k_{0}) \\ I_{n} & C^{T}(k_{0}) & O_{n\times n} \end{bmatrix}^{\dagger} \begin{pmatrix} \bar{x} \\ y(k_{0}) \\ 0 \end{pmatrix}, \end{cases} \tag{5.19}$$

with the covariance matrix $\check{P}(k)$ given by:

$$\begin{cases} \check{P}(k+1) = -\begin{bmatrix} O_{n\times(m+q)} & I_{n} \end{bmatrix} \begin{bmatrix} \check{M}_{P}(k+1) & H(k) \\ H^{T}(k) & O_{n\times n} \end{bmatrix}^{\dagger} \begin{bmatrix} O_{(m+q)\times n} \\ I_{n} \end{bmatrix}, \\[4mm] \check{M}_{P}(k+1) = \begin{bmatrix} A(k)\check{P}(k)A^{T}(k) + Q_{f}(k) & O_{m\times q} \\ O_{q\times m} & Q_{g}(k+1) \end{bmatrix}, & k > k_{0}, \\[4mm] \check{P}(k_{0}) = -\begin{bmatrix} O_{n\times n} & O_{n\times q} & I_{n} \end{bmatrix} \begin{bmatrix} \Psi_{x_{0}} & O_{n\times q} & I_{n} \\ O_{q\times n} & Q_{g}(k_{0}) & C(k_{0}) \\ I_{n} & C^{T}(k_{0}) & O_{n\times n} \end{bmatrix}^{\dagger} \begin{bmatrix} O_{n\times n} \\ O_{n\times q} \\ I_{n} \end{bmatrix}. \end{cases} \tag{5.20}$$

Proof. The proof is articulated in two steps: it is first proved that the a priori error covariance matrix of the maximum likelihood filter, $\check{P}(k_{0})$ in (5.20), is equal to the a priori error covariance matrix of the best linear estimate, $P(k_{0})$ in (5.2). The equivalence between the two error covariance matrix for any given instant k is, then, proved by induction.

First step. According to the Riccati equations (5.2), after some computations, using properties of generalized inverses of partitioned matrices (see [2] for more details), one has

$$P(k_{0}) = \Psi_{x_{0}} - \Psi_{x_{0}}C^{T}Q_{1}^{\dagger}C\Psi_{x_{0}}, \quad \text{with} \quad Q_{1} = Q_{g} + C\Psi_{x_{0}}C^{T}, \tag{5.21}$$

and

$$\check{P}(k_0) = \left(\begin{bmatrix} I_n & C^T \end{bmatrix} \begin{bmatrix} I_n + \Psi_{x_0} & C^T \\ C & Q_g + CC^T \end{bmatrix}^\dagger \begin{bmatrix} I_n \\ C \end{bmatrix} \right)^\dagger$$
$$\cdot \left(I_n - \begin{bmatrix} I_n & C^T \end{bmatrix} \begin{bmatrix} I_n + \Psi_{x_0} & C^T \\ C & Q_g + CC^T \end{bmatrix}^\dagger \begin{bmatrix} I_n \\ C \end{bmatrix} \right).$$
$$(5.22)$$

Moreover, it is

$$\begin{bmatrix} I_n & C^T \end{bmatrix} \begin{bmatrix} I_n + \Psi_{x_0} & C^T \\ C & Q_g + CC^T \end{bmatrix}^\dagger \begin{bmatrix} I_n \\ C \end{bmatrix} \qquad (5.23)$$
$$= \left(I_n + \Psi_{x_0} \right)^{-1} + \left(I_n + \Psi_{x_0} \right)^{-1} \Psi_{x_0} C^T Q_2^\dagger C \Psi_{x_0} \left(I_n + \Psi_{x_0} \right)^{-1},$$

where the symmetric, positive semidefinite matrix Q_2 is defined as

$$Q_2 = Q_g + CC^T - C\left(I_n + \Psi_{x_0} \right)^\dagger C^T = Q_g + C\Psi_{x_0} \left(I_n + \Psi_{x_0} \right)^{-1} C^T, \qquad (5.24)$$

(see [2]). Substituting (5.23) in (5.22):

$$\check{P}(k_0) = \left(I_n + \Psi_{x_0} C^T Q_2^\dagger C \Psi_{x_0} \left(I_n + \Psi_{x_0} \right)^{-1} \right)^{-1} \cdot \left(I_n + \Psi_{x_0} \right)$$
$$\cdot \left(I_n - \left(I_n + \Psi_{x_0} \right)^{-1} - \left(I_n + \Psi_{x_0} \right)^{-1} \Psi_{x_0} C^T Q_2^\dagger C \Psi_{x_0} \left(I_n + \Psi_{x_0} \right)^{-1} \right)$$
$$= \left(I_n + \Psi_{x_0} C^T Q_2^\dagger C \Psi_{x_0} \left(I_n + \Psi_{x_0} \right)^{-1} \right)^{-1} \cdot \left(\Psi_{x_0} - \Psi_{x_0} C^T Q_2^\dagger C \Psi_{x_0} \left(I_n + \Psi_{x_0} \right)^{-1} \right)$$
$$= \left(I_n + \Psi_{x_0} \right) \left(I_n + \Psi_{x_0} + \Psi_{x_0} C^T Q_2^\dagger C \Psi_{x_0} \right)^{-1}$$
$$\cdot \left(\Psi_{x_0} + \Psi_{x_0}^2 - \Psi_{x_0} C^T Q_2^\dagger C \Psi_{x_0} \right) \left(I_n + \Psi_{x_0} \right)^{-1}$$
$$= \left(I_n + \Psi_{x_0} \right) \left(I_n + \Psi_{x_0} + \Psi_{x_0} C^T Q_2^\dagger C \Psi_{x_0} \right)^{-1} \left(I_n + \Psi_{x_0} \right) - I_n.$$
$$(5.25)$$

In order to verify that $P(k_0) = \check{P}(k_0)$, it is enough to show the following identity:

$$\left(\check{P}(k_0) + I_n \right)^{-1} \left(P(k_0) + I_n \right) = I_n. \qquad (5.26)$$

After some computations, substituting (5.21) and (5.25), one has:

$$\left(\check{P}(k_0) + I_n \right)^{-1} \left(P(k_0) + I_n \right) = I_n + \left(I_n + \Psi_{x_0} \right)^{-1} \Psi_{x_0} C^T \left(Q_2^\dagger - Q_1^\dagger + Q_2^\dagger (Q_2 - Q_1) Q_1^\dagger \right) C \Psi_{x_0}.$$
$$(5.27)$$

According to Lemma 5.5, if the symmetric semipositive definite matrices Q_1 and Q_2 have the same range, the right hand term of equation (5.27) becomes I_n, and eq. (5.26) is proved. Being Q_1 and Q_2 symmetric, Q_1 and Q_2 have the same range if and only if they have the same null space. This result can be proved by showing first that $\mathcal{N}(Q_1) \subseteq \mathcal{N}(Q_2)$ and next that $\mathcal{N}(Q_2) \subseteq \mathcal{N}(Q_1)$. To prove the first inclusion choose any $x \in \mathcal{N}(Q_1)$. Then $x \in \mathcal{N}(Q_g) \cap \mathcal{N}(C\Psi_{x_0} C^T)$, since Q_g and $C\Psi_{x_0} C^T$ are both positive semidefinite matrices. Moreover $\mathcal{N}(C\Psi_{x_0} C^T) = \mathcal{N}(\Psi_{x_0}^{\frac{1}{2}} C^T)$, so that:

$$x \in \mathcal{N}\left(Q_g + C\Psi_{x_0}^{\frac{1}{2}} (I_n + \Psi_{x_0}^{-1}) \Psi_{x_0}^{\frac{1}{2}} C^T \right) = \mathcal{N}(Q_2).$$

26.

A similar analysis can be used to show that $\mathcal{N}(Q_2) \subseteq \mathcal{N}(Q_1)$.

Second step. The proof is achieved verifying that:

$$\text{if } P(k) = \check{P}(k) \text{ for some } k \quad \Longrightarrow \quad P(k+1) = \check{P}(k+1).$$

From (5.2) and (5.20), easily follows that $P(k) = \check{P}(k)$ implies $M_P(k+1) = \check{M}_P(k+1)$. After some computations concerning pseudo-inverses of partitioned matrices (see [2] for more details):

$$P(k+1) = H^\dagger(k)\left\{I_{m+q} - M_P(k+1)L^T(k)\left(L(k)M_P(k+1)L^T(k)\right)^\dagger L(k)\right\}M_P(k+1)H^{\dagger T}(k),$$
$$(5.28)$$

$$\check{P}(k+1) = H^\dagger(k)\left\{I_{m+q} - M_P(k+1)\left[\left(I_{m+q} - H(k)H^\dagger(k)\right)M_P(k+1)\right.\right.$$
$$\left.\left. \cdot \left(I_{m+q} - H(k)H^\dagger(k)\right)\right]^\dagger\right\}M_P(k+1)H^{\dagger T}(k),$$
$$(5.29)$$

The two covariance matrices are equal if the following identity is verified:

$$M_P^{\frac{1}{2}}L^T\left(LM_PL^T\right)^\dagger LM_P^{\frac{1}{2}} = M_P^{\frac{1}{2}}\left(\left(I_{m+q} - HH^\dagger\right)M_P\left(I_{m+q} - HH^\dagger\right)\right)^\dagger M_P^{\frac{1}{2}}. \qquad (5.30)$$

From (5.18), the left side of (5.30) becomes:

$$M_P^{\frac{1}{2}}L^T\left(LM_PL^T\right)^\dagger LM_P^{\frac{1}{2}} = M_P^{\frac{1}{2}}L^T\left(LM_P^{\frac{1}{2}}M_P^{\frac{1}{2}}L^T\right)^\dagger LM_P^{\frac{1}{2}} = M_P^{\frac{1}{2}}L^T\left(M_P^{\frac{1}{2}}L^T\right)^\dagger\left(LM_P^{\frac{1}{2}}\right)^\dagger LM_P^{\frac{1}{2}}.$$
$$(5.31)$$

Applying Lemma 5.4, item ii), and (5.18), the right side of (5.30) becomes:

$$M_P^{\frac{1}{2}}\left(\left(I_{m+q} - HH^\dagger\right)M_P\left(I_{m+q} - HH^\dagger\right)\right)^\dagger M_P^{\frac{1}{2}} = M_P^{\frac{1}{2}}\left(L^\dagger LM_P^{\frac{1}{2}}M_P^{\frac{1}{2}}L^\dagger L\right)^\dagger M_P^{\frac{1}{2}}$$
$$= M_P^{\frac{1}{2}}\left(M_P^{\frac{1}{2}}L^\dagger L\right)^\dagger\left(L^\dagger LM_P^{\frac{1}{2}}\right)^\dagger M_P^{\frac{1}{2}}.$$
$$(5.32)$$

Substituting (5.31) and (5.32), the identity (5.30) is verified, if it is shown that:

$$\left(LM_P^{\frac{1}{2}}\right)^\dagger LM_P^{\frac{1}{2}} = \left(L^\dagger LM_P^{\frac{1}{2}}\right)^\dagger M_P^{\frac{1}{2}}. \qquad (5.33)$$

The left side is a projector onto $\mathcal{R}\left(\left(LM_P^{\frac{1}{2}}\right)^\dagger\right)$, and it is easy to verify that it is equal to $\left(L^\dagger LM_P^{\frac{1}{2}}\right)^\dagger L^\dagger LM_P^{\frac{1}{2}}$, a projector onto $\mathcal{R}\left(\left(L^\dagger LM_P^{\frac{1}{2}}\right)^\dagger\right)$, by proving the equivalence of the two ranges:

$$\mathcal{R}\left(\left(L^\dagger LM_P^{\frac{1}{2}}\right)^\dagger\right) = \mathcal{R}\left(\left(L^\dagger LM_P^{\frac{1}{2}}\right)^T\right) = \mathcal{R}\left(M_P^{\frac{1}{2}}L^\dagger L\right) = \mathcal{R}\left(M_P^{\frac{1}{2}}L^T\right) = \mathcal{R}\left(\left(LM_P^{\frac{1}{2}}\right)^T\right)$$
$$= \mathcal{R}\left(\left(LM_P^{\frac{1}{2}}\right)^\dagger\right),$$

so that (5.33) becomes:

$$\left(L^\dagger LM_P^{\frac{1}{2}}\right)^\dagger L^\dagger LM_P^{\frac{1}{2}} = \left(L^\dagger LM_P^{\frac{1}{2}}\right)^\dagger M_P^{\frac{1}{2}}. \qquad (5.34)$$

The identity (5.34) is immediately verified by applying Lemma 5.6.

∎

6. Simulations

This section presents computer simulations that show the improvements of a quadratic filter with respect to a linear one. They refer to singular systems deriving from unknown input systems, whose singularity derives from a loss of knowledge (see Darouach et al. in [10]). The following third order time-invariant system

$$\begin{cases} \tilde{x}(k+1) = \widetilde{A}\tilde{x}(k) + \widetilde{B}u(k) + f_k, \\ y(k) = \widetilde{C}\tilde{x}(k) + g_k, \end{cases} \tag{6.1}$$

where

$$\widetilde{A} = \begin{bmatrix} 0.7 & 0.1 & 0 \\ 0 & 0.4 & 0.5 \\ 0 & 0 & 0.8 \end{bmatrix}, \qquad \widetilde{B} = \begin{bmatrix} 1 \\ -1.5 \\ 2 \end{bmatrix}, \qquad \widetilde{C} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & -1 & 0 \end{bmatrix}, \tag{6.2}$$

has been considered, where the input $u(k)$ is unknown.

According to item iii) and iv) of system (4.1), state and output noises, f_k and g_k respectively, are supposed to be independent, zero-mean, white sequences, whose moments are available up to the 4-th order. Moreover, to show the efficiency of the algorithm also for systems affected by noises with singular covariance matrices (a very common case for singular systems, see Nikoukhaha in [27]), the state and measurement noises used in the simulations have been defined as the following white sequences

$$f_k = \begin{pmatrix} f_k^1 \\ f_k^2 \\ f_k^3 \end{pmatrix}, \qquad g_k = \begin{pmatrix} g_k^1 \\ g_k^2 \end{pmatrix},$$

where

$$P(f_k^1 = 0) = 1$$

$$P(f_k^2 = -\sqrt{6}) = 0.4, \qquad\qquad P(f_k^2 = 2\sqrt{\tfrac{2}{3}}) = 0.6,$$

$$P(f_k^3 = -\sqrt{\tfrac{5}{6}}) = 0.8, \qquad\qquad P(f_k^3 = 4\sqrt{\tfrac{5}{6}}) = 0.2,$$

and

$$P(g_k^1 = -\sqrt{\tfrac{5}{3}}) = 0.9, \qquad\qquad P(g_k^1 = 3\sqrt{5}) = 0.1,$$

$$P(g_k^2 = 0) = 1.$$

The initial state is a random vector independent of f_k and g_k with central moments available up to the 4-th order. $u(k)$ is an unknown, deterministic input.

It is well known in literature (see [10] for more details) that systems with unknown inputs can be treated as singular systems through the definition of an extended state $x(k)$ that contains the uncertain input:

$$x(k) = \begin{pmatrix} \tilde{x}(k) \\ u(k-1) \end{pmatrix}. \tag{6.3}$$

With this position, system (6.1) can be rewritten in the form (4.1), with:

$$J = \begin{bmatrix} I_n & -\widetilde{B} \end{bmatrix} \qquad\qquad C = \begin{bmatrix} \widetilde{C} & O_{q\times 1} \end{bmatrix} \tag{6.4}$$

Moreover, according to its construction, it results to be a causal solvable system, estimable from measurements, so that the filtering approach proposed in this paper can be applied.

Both linear and quadratic filters have been implemented ($\Delta = 0$ has been chosen). The plot in fig. 6.1 shows the improvements in state estimation by increasing the order of the filtering algorithm.

The asymptotic values of the error covariance matrices allow to appreciate the improvement of the filter:

$$P_{\nu=2} = \begin{bmatrix} 0.5612 & -0.1411 & -0.1414 & 0.3281 \\ -0.1411 & 0.5835 & 0.5833 & -0.0332 \\ -0.1414 & 0.5833 & 0.5835 & -0.0332 \\ 0.3281 & -0.0332 & -0.0332 & 0.3525 \end{bmatrix}$$

$$P_{\nu=1} = \begin{bmatrix} 0.6174 & -0.0493 & -0.0493 & 0.3520 \\ -0.0493 & 0.7684 & 0.7684 & 0.0266 \\ -0.0493 & 0.7684 & 0.7684 & 0.0266 \\ 0.3520 & 0.0266 & 0.0266 & 0.3814 \end{bmatrix}$$
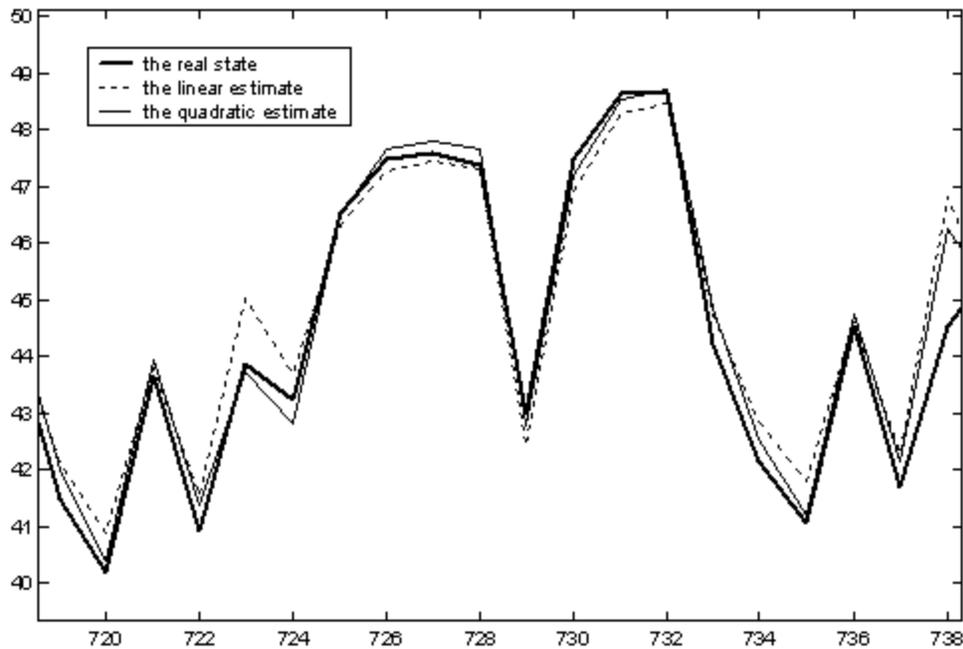

Fig. 6.1

7. Conclusions

Filtering theory for descriptor systems has been widely investigated in literature in presence of Gaussian noises. This paper proposes a geometric approach to estimate the descriptor vector of a non-Gaussian singular system. The novelty of this paper is that the proposed filtering algorithm is based on the minimum error variance criterion. It consists of a polynomial filter, whose performances improve by increasing the polynomial order, that can be considered as a proper extension of the filter presented in [27], because its restriction to first order polynomials in the case of Gaussian noises gives exactly the same results of the maximum likelihood estimator developed in [27].

Appendix: the Kronecker algebra

For the ease of the reader, in this Appendix are reported some useful results on the Kronecker algebra. The proofs and other further details can be found in [30].

Let M and N be matrices of dimensions $r \times s$ and $p \times q$ respectively, then the Kronecker product $M \otimes N$ is defined as the $(r \cdot p) \times (s \cdot q)$ matrix

$$M \otimes N = \begin{bmatrix} m_{11}N & \dots & m_{1s}N \\ \vdots & \ddots & \vdots \\ m_{r1}N & \dots & m_{rs}N \end{bmatrix}, \tag{A.1}$$

where the m_{ij} are the entries of M.

Definition A.1. Let M be an $r \times s$ matrix:

$$M = \begin{bmatrix} m_1 & m_2 & \dots & m_s \end{bmatrix}, \tag{A.2}$$

where m_i denotes the i-th column of M. The stack of M is defined as the $r \cdot s$ vector:

$$\text{st}(M) = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_s \end{bmatrix}. \tag{A.3}$$

Observe that a vector as in (A.3) can be reduced to a matrix M as in (A.2), once it is known the number of the rows r of the original matrix, by considering the inverse operation of the stack denoted by st^{-1}. More generally, let m be a vector in $\mathbb{R}^\mu$, and r be a divisor of μ. Then the $r \times (\mu/r)$ matrix given by $M = \text{st}^{-1}(m, r)$ is defined so that:

$$\text{st}(M) = m. \tag{A.4}$$

In presence of vectors $m \in \mathbb{R}^{(\mu^2)}$, that is their length is given by a square, the notation $\text{st}^{-1}(m)$ has to be considered as a short version of $\text{st}^{-1}(m, \mu)$.

In case of vectors Kronecker products, it is easy to verify that, if $u \in \mathbb{R}^r$ and $v \in \mathbb{R}^s$, the i-th entry of $u \otimes v$ is given by

$$(u \otimes v)_i = u_l \cdot v_m; \quad l = \left[\frac{i-1}{s} \right] + 1, \quad m = |i-1|_s + 1, \tag{A.5}$$

where $[\cdot]$ and $|\cdot|_s$ denote integer part and s-modulo respectively. Moreover, the Kronecker power of M is defined as

$$\begin{aligned} M^{[0]} &= 1 \in \mathbb{R}, \\ M^{[l]} &= M \otimes M^{[l-1]} \quad l \geq 1. \end{aligned} \tag{A.6}$$

Some useful properties of the Kronecker product and stack operation are the following

$$
\begin{aligned}
(A + B) \otimes (C + D) &= A \otimes C + A \otimes D \\
&\quad + B \otimes C + B \otimes D \\
A \otimes (B \otimes C) &= (A \otimes B) \otimes C \\
(A \cdot C) \otimes (B \cdot D) &= (A \otimes B) \cdot (C \otimes D) \\
(A \otimes B)^T &= A^T \otimes B^T \\
st(A \cdot B \cdot C) &= (C^T \otimes A) \cdot st(B) \\
u \otimes v &= st(v \cdot u^T) \\
tr(A \otimes B) &= tr(A) \cdot tr(B)
\end{aligned}
$$

Other useful properties can be found in [1].

According to its definition (A.1), the Kronecker product is not commutative. However, the following result holds:

Lemma A.2. *For any given pair of matrices $A \in I\!R^{r \times s}$, $B \in I\!R^{n \times m}$, it is:*

$$
B \otimes A = C_{r,n}^T (A \otimes B) C_{s,m}, \tag{A.7}
$$

where $C_{r,n}$, $C_{s,m}$ are defined so that, denoted $\{C_{u,v}\}_{h,l}$ their (h, l) entries:

$$
\{C_{u,v}\}_{h,l} = \begin{cases} 1, & \text{if } l = (|h-1|_v)u + \left(\left[\dfrac{h-1}{v} \right] + 1 \right); \\[2mm] 0, & \text{otherwise.} \end{cases} \tag{A.8}
$$

Remark A.3. *Observe that $C_{1,1} = 1$, hence in the vector case when $a \in I\!R^r$ and $b \in I\!R^n$,* (A.7) *becomes*

$$
b \otimes a = C_{r,n}^T (a \otimes b). \tag{A.9}
$$

Moreover, in the vector case the commutation matrices satisfy also the following recursive formula.

Lemma A.4. *Let $a, b \in I\!R^n$ and $l \in I\!N$. Then*

$$
b^{[l]} \otimes a = G_l(n)(a \otimes b^{[l]}), \tag{A.10}
$$

with the sequence $\{G_l(n) = C_{n,n^l}^T\}$ given by the following recursive equations

$$
\begin{aligned}
G_1(n) &= C_{n,n}^T, \\
G_l(n) &= (I_{n,1} \otimes G_{l-1}(n)) \cdot (G_1(n) \otimes I_{n,l-1}), \qquad l > 1,
\end{aligned} \tag{A.11}
$$

where $I_{n,r}$ is the identity matrix in $I\!R^{n^r}$.

A binomial formula can be found for the Kronecker power, which generalizes the classical Newton one.

Lemma A.5. *Let $a, b \in \mathbb{R}^n$. For any integer $h \geq 0$ the matrix coefficients of the following binomial power formula:*

$$(a + b)^{[h]} = \sum_{k=0}^{h} M_k^h(n)(a^{[k]} \otimes b^{[h-k]}) \tag{A.12}$$

constitute a set of matrices $\{M_0^h(n), ..., M_h^h(n); \quad M_k^h(n) \in \mathbb{R}^{n^h \times n^h}\}$ such that:

$$
\begin{aligned}
M_h^h(n) =& M_0^h(n) = I_{n,h}, \\
M_j^h(n) =& (M_j^{h-1}(n) \otimes I_{n,1}) + (M_{j-1}^{h-1}(n) \otimes I_{n,1}) \cdot (I_{n,j-1} \otimes G_{h-j}(n)), \quad 1 \leq j \leq h-1,
\end{aligned}
\tag{A.13}
$$

where $G_l(n)$ and $I_{n,l}$ are as in Lemma A.4.

References

[1] R. Bellman, *Introduction to matrix analysis*, McGraw-Hill, 1970.

[2] S. L. Campbell and C. D. Meyer Jr, *Generalized inverses of linear transformations*, New York, Dover, 1979.

[3] S. L. Campbell, *Singular Systems of Differential Equations*, Research Notes in Math., No. 40, San Francisco, CA:Pitman, 1980.

[4] S. L. Campbell, *Singular Systems of Differential Equations II*, Research Notes in Math., No. 61, San Francisco, CA:Pitman, 1982.

[5] S. H. Crandall, D. C. Karnopp, E. F. Kurtz Jr. and D. C. Pridmore-Brown, *Dynamics of Mechanical and Electromechanical Systems*, McGraw-Hill, New York, 1968.

[6] L. Dai, "State estimation schemes in singular systems", in:*Proc. 10th IFAC World Congress*, Munich, Vol. 9, pp. 211–215, 1987.

[7] L. Dai, "Filtering and LQG problems for discrete-time stochastic singular systems", *IEEE Trans. Automat. Contr.*, No. 34, pp. 1105–1108, 1989.

[8] M. Darouach and M. Zasadzinski, "State estimation for a class of singular systems", *Int. J. Systems Sci.*, Vol. 23, N. 4, pp. 517–530, 1992.

[9] M. Darouach, M. Zasadzinski and D. Mehdi, "State estimation of stochastic singular linear systems", *Int. J. Syst. Sci.*, Vol. 2, No. 2, pp. 345–354, 1993.

[10] M. Darouach, M. Zasadzinski, A. Basson Onana and S. Nowakowski, "Kalman filtering with unknown inputs via optimal state estimation of singular systems", *Int. J. Syst. Sci.*, Vol. 26, No. 10, pp. 2015–2028, 1995.

[11] A. Germani, A. De Santis and M. Raimondi, "Optimal quadratic filtering of linear discrete-time non-Gaussian systems", *IEEE Trans. Autom. Contr.*, Vol. 40, 1995.

[12] A. Germani, F. Carravetta and M. Raimondi, "Polynomial filtering for linear discrete-time non-Gaussian systems", *SIAM J. Control and Optim.*, Vol. 34, N. 5, pp. 1666–1690, 1996.

[13] A. Germani, F. Carravetta and M. Raimondi, "Polynomial filtering of discrete-time stochastic linear systems with multiplicative state noise", *IEEE Trans. Autom. Contr.*, Vol. 42, N. 8, pp. 1106–1126, 1997.

[14] R. E. Kalman, "A new approach to linear filtering and prediction problems", *ASME J. Basic Eng.*, series D, Vol. 82, 1960.

[15] F. L. Lewis, "A survey of linear singular systems", *Circuits, Syst., Signal Process.*, No. 5, pp. 3–36, 1986.

32.

[16] D. G. Luenberger and A. Arbel, "Singular dynamic Leontief systems", *Econometrica*, 1977.

[17] D. G. Luenberger, "Dynamic Equations in Descriptor Form", *IEEE Trans. Automat. Contr.*, No. 3, pp. 312–321, 1977.

[18] D. G. Luenberger, "Time-Invariant Descriptor Systems", *Automatica*, No. 14, pp. 473–480, 1978.

[19] D. G. Luenberger, "Boundary recursion for descriptor variable systems", *IEEE Trans. Automat. Contr.*, No. 34, pp. 287–292, 1989.

[20] M. A. Massoumnia, G. C. Verghese and A. W. Willsky, "Failure detection and identification", *IEEE Trans. Autom. Contr.*, Vol. 34, No. 3, pp. 316–321, 1989.

[21] J. H. Miller and S. C. Schwartz, "Robust detection of a known signal in nearly Gaussian noise", *IEEE Trans. Inform. Theory*, Vol. 17, pp. 50–56, 1971.

[22] J. H. Miller and J. B. Thomas, "Detectors for discrete-time signals in non-Gaussian noise", *IEEE Trans. Inform. Theory*, Vol. 18, pp. 241–250, 1972.

[23] R. Nikoukhah , B. C. Levy and A. S. Willsky, "Boundary-value descriptor systems: Well-posedness, reachability, and observability", *Int. J. Contr.*, No. 46, pp. 1715–1737, 1987.

[24] R. Nikoukhah, "A deterministic and stochastic theory for two-point boundary-value descriptor systems", Ph.D. Dissertation, Dept. Elec. Eng. Comp. Sci., Mass. Inst. Technol., MA:Cambridge, 1988.

[25] R. Nikoukhah, B. C. Levy and A. S. Willsky, "Stability, stochastic stationary, and generalized Lyapunov equations for two-point boundary-value descriptor systems", *IEEE Trans. Automat. Contr.*, No. 34, pp. 1141–1152, 1989.

[26] R. Nikoukhah, M. B. Adams, B. C. Levy, A. S. Willsky, "Estimation for boundary-value descriptor systems", *Circuits, Syst., Signal Proc.*, No. 8, pp. 25–48, 1989.

[27] R. Nikoukhah , B. C. Levy and A. S. Willsky, "Kalman filtering and Riccati equations for descriptor systems", *IEEE Trans. Automat. Contr.*, No. 9, pp. 1325–1342, 1992.

[28] R. Nikoukhah, S. L. Campbell and F. Delebecque, "Kalman filtering for general discrete-time linear systems", *IEEE Trans. Automat. Contr.*, Vol. 44, No. 10, pp. 1829–1839, 1999.

[29] S. S. Rappaport and L. Kurtz, "An optimal nonlinear detector for digital data transmission through non-Gaussian channels", *IEEE Trans. Comm. Tech.*, Vol. 14, pp. 266–274, 1966.

[30] G. S. Rodgers, *Matrix derivatives*, Marcel Dekker, New York, Basel, 1980

[31] S. P. Sing and R. W. Liu, "Existence of state equation representation of linear large-scale dynamical systems", *IEEE Trans. Circuit Theory.*, No. 20, pp. 239–246, 1973.

[32] G. C. Verghese, B. C. Levy and T. Kailath, "A generalized state-space for singular systems", *IEEE Trans. Automat. Contr.*, Vol. 26, No. 4, pp. 811–831, 1981.

[33] J. C. Willems, "A framework for the study of dynamical systems", in:*Realization and Modelling in System Theory*, No. 1, M. A. Kaashoek, J. H. Van Schuppen, and A. C. M. Ran, Eds., Boston, MA: Birkhauser-Verlag, pp. 43–60, 1990.

[34] E. Yaz, "Relationship between several novel control schemes proposed for a class of nonlinear stochastic systems", *Int. J. Control*, N. 45, pp. 1447–1454, 1987.

[35] E. Yaz, "A control scheme for a class of discrete nonlinear stochastic systems", *IEEE Trans. Autom. Contr.*, Vol. 32, pp. 77–80, 1987.