

A m -dimensional stochastic estimator

R. Palma Orozco

*Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada del IPN,
Legaria No. 694, México, D.F. 11500, México.*

J. de J. Medel Juárez

*Centro de Investigación en Computación del IPN,
Av. Juan de Dios Bátiz s/n casi esq. Miguel Othón de Mendizábal.
Unidad Profesional Adolfo López Mateos México D.F. 07738, México.*

G. Garrido Aguilar

*Escuela Superior de Ingeniería Mecánica y Eléctrica del IPN,
Unidad Culhuacán, Av. Santa Ana No. 1000 México D.F. 04430, México.*

Recibido el 28 de junio de 2011; aceptado el 14 de diciembre de 2011

This paper shows the development of a optimal stochastic estimator for a black-box system in a m -dimensional space, observing noise with an unknown dynamics model. The results are based in state space, described by a discrete stochastic estimator and noise characterization. The proposed result gives an algorithm to construct diagonal form for the state space system. It is a new technique for a instrumental variable tool, and a diagonalization process avoiding the calculation of pseudo-inverse matrices is presented with a linear computational complexity $O(j)$ and j as the diagonal matrix dimension. The results show that it is possible to reconstruct the observable signal with a probability approximation.

Keywords: Linear algebra; matrix theory; control theory; stochastic processes.

Este artículo muestra el desarrollo de un estimador estocástico óptimo para un modelo de sistemas tipo caja negra con ruido en un espacio m -dimensional. Se propone un algoritmo para evaluar y construir la forma diagonal del sistema en espacio de estados para estimar las ganancias internas. El algoritmo permite eliminar el cálculo de matrices pseudoinversas y tiene una complejidad computacional de orden lineal $O(j)$, donde j es la dimensión de la matriz diagonal y que computacionalmente representa una menor complejidad que los métodos utilizados tradicionalmente a través de la pseudoinversa. Los resultados muestran que es posible reconstruir la señal observable con una buena aproximación en un sentido de probabilidad.

Descriptores: Álgebra lineal; teoría de matrices; teoría de control; procesos estocásticos.

PACS: 02.10.Ud; 02.10.Yn; 02.30.Yy; 02.50.Ey; 02.70.-c

1. Introduction

Often in the real world one expects is to find linear relationships between variables. In stochastic processes the linear description or linear fit, is a mathematical method considering $Y_k \in \mathbb{R}^{[m \times 1]}$ and $Y_{k-1} \in \mathbb{R}^{[m \times 1]}$ with a random term $W_k \in \mathbb{R}^{[m \times 1]}$, i.e., $\{Y_k\} \propto \{W_k\}$ and $k \in [1, m]$, $m \in \mathbb{N}$ [3]. The linear model relates to the variables $\{Y_k\}$ and $\{W_k\}$ or any transformation thereof, which generates a hyperplane $A_k \in \mathbb{R}^{[m \times m]}$ with unknown parameters. W_k is a random disturbance which includes neither controllable nor observable factors and therefore is associated with the randomness giving it the stochastic character [3]. The values of A_k are generated inside the system without knowing its gains. The basic model has an ARMA(1, 1) form (1).

$$Y_k = A_k Y_{k-1} + W_k \quad (1)$$

The estimator A_k is linear unbiased, wherein the overall concept is the instrumental variable selection [8,9]. Breaking the temporal correlation between the regressors and disturbances, justifies the search for an alternative estimator [6]. The second moment applied in an estimator provides consis-

tency and is computationally viable considering the recursive form [7,10].

The instrumental variable method is a procedure requiring linear algebra calculus, determining the best fit line. A careful proof analysis will show that the method has an application in several dimensions. Instead of finding the best fit line by a pseudo-inverse method, it is better to find that given by any finite linear combination of specified functions that allows minimizing the complexity. Therefore, the general problem is given by the set functions f_1, \dots, f_j , to find the internal coefficient values $\hat{a}_1, \dots, \hat{a}_j$ such that the linear combination knows the observable signal, $y = a_1 f_1 + \dots + a_j f_j$, $j \in \mathbb{N}$ that is the best approximation [1,2]. The solutions depend on the constraint functions f_j , $j \in \mathbb{N}$. If there is exactly one equation for each measurement, and the functions f_j have linear combinations with their elements, the solution is discussed under linear regression.

For non-linear systems the number of unknown parameters, could be described as instrumental variable applied iteratively to a linearized function form until convergence is achieved in some sense. However, it is often possible to linearize a non-linear function at the outset and still use linear

methods determining fit parameters without resorting to iterative procedures. This approach commonly violates the implicit assumption that the errors distribution is normal, but often still gives acceptable results using normal equations with pseudo-inverse techniques.

A common use of the Moore-Penrose method is to compute a best fit solution to a linear equation system that lacks a unique solution. Another use is to find the minimum Euclidean norm solution [23]. The pseudo-inverse is defined and unique for all matrices whose entries are real numbers. It can be computed using singular value decomposition. Indeed, the pseudo-inverse matrix with singular value decomposition could be described as $M=UPV^*$ is $M^+=V^+P^+U^*$ where P^+ is that formed by replacing every non zero diagonal entry by its reciprocal and then transposing the resulting matrix. This method is one way to solve linear instrumental variable problems selecting eigenvalues and eigenvectors without knowing the internal evolution system [24].

The singular decomposition matrix value $M^{[n \times n]}$, $n \in \mathbb{N}$, is typically computed by a two-step procedure. In the first step, the matrix is reduced to a bi-diagonal matrix, and has $O(n^3)$ floating-point operations (flops). The second step is computed by singular decomposition matrix value. This step can only be made by an iterative method. In practice it suffices to compute the singular value decomposition up to a certain precision. If the precision is considered constant, then the second step takes $O(n)$ iterations costing $O(n)$ flops [20,25]. Thus, the first step was more expensive, and the overall cost $O(n^3)$.

Traditional form considers Householder reflections with a cost of $4mn^2 - 4n^3/3$ flops, $m, n \in \mathbb{N}$, assuming that only singular values and not singular vectors are needed [19]. If $m \gg n$ then it is advantageous to first reduce the matrix M to a triangular matrix with QR decomposition and then use Householder reflections to further reduce the matrix to bi-diagonal form; the combined cost was $2mn^2 + 2n^3$ flops [19,20,25]. The second step can be made by a variant of the QR algorithm for eigenvalues description (Golub & Kahan) [22,21].

To understand the dynamic system is a well known problem. In all cases, it is necessary to include some unknown initial values to compute the first estimation. The second estimation uses the old values to get the next best approximation. The iteration continues until the diagonal matrix dimension. It is expected that the estimator values are convenient to define best estimations. Therefore, it is necessary to compute the eigenvalues and eigenvectors system, requiring an evaluation of some pseudo-inverse method that is expensive because of computational complexity. Thus, this procedure is not suitable at all. To avoid this, a method for designing an optimal stochastic estimator considering the observable signals given by the system is shown. There is a reduction in the computational complexity, because there is no necessity to implement any Penrose procedure. In order to get the best stochastic estimator, Medel *et al.* have proposed some descriptions [15-18].

The purpose of this paper is to show an inverse allowing good estimators for a black-box system (1). A diagonal representation, providing the best computational complexity allows inverse matrix. It is structured in the following manner: Section 2 describes the basic formalism of the m -dimensional stochastic estimator and Sec. 3 presents the implementation results. Section 4 discusses these results. Section 5 determines the conclusions, and finally, Sec. 6 is the theorem's proof annex.

2. Main results

In (1), the vectors $Y_k, W_k \in \mathbb{R}^{[m \times 1]}$ are random variables $N(\mu, \sigma^2 < \infty)$ and $\mathcal{Y}_k, \mathcal{W}_k \in \mathbb{R}^{[n \times (m \times m)]}$, $n, m \in \mathbb{N}$ with all non-zero entries, respectively, corresponding to the diagonal matrices in a Jordan canonical form having n blocks with $m \times m$ entries each one [16]. The system gain matrix $A_k \in \mathbb{R}^{[m \times m]}$ has the corresponding diagonal matrix $\mathcal{A}_k \in \mathbb{R}^{[n \times (m \times m)]}$ with all non-zero entries. Any diagonal matrix is symmetric, triangular and normal if its entries are in a \mathbb{R} field. Let $D = \text{diag}(d_i) \in \mathbb{R}^{[n \times (m \times m)]}$ and $\det(D) = \prod_i d_i$ with $i \in [1, nm]$. The diagonal matrix is invertible if and only if every $d_i \neq 0$, hence there are $\mathcal{Y}_k^{-1}, \mathcal{W}_k^{-1} \in \mathbb{R}^{[n \times (m \times m)]}$ [14]. The system described in (1), in diagonal form is given by (2).

$$\mathcal{Y}_k = \mathcal{A}_k \mathcal{Y}_{k-1} + \mathcal{W}_k \quad (2)$$

Let \mathcal{Y}_k and \mathcal{W}_k be vectors related to the state space and $\vartheta_k \in \mathbb{R}^{[n \times (m \times m)]}$ be the instrumental variable. Therefore, $\{\mathcal{Y}_k\} \propto \{\mathcal{W}_k\}$ if and only if there exists an affine linear space $\mathcal{Y}_k \in \mathbb{R}^{[n \times (m \times m)]}$ such that $\mathcal{A}_k, \mathcal{I} \in \mathbb{R}^{[n \times (m \times m)]}$ are constants, satisfying the linear combination (2) [12,13]. The diagonal identification form is given by (3).

$$\widehat{\mathcal{Y}}_k = \widehat{\mathcal{A}}_k \widehat{\mathcal{Y}}_{k-1} + \mathcal{W}_k \quad (3)$$

Thus, the relationship (3) defines the black-box system identification where $\widehat{\mathcal{A}}_k$ is unknown internal parameters and it is described in Theorem 1 and the proof is in Sec. 6. Theorem 2 describes the stochastic estimator (5) having an optimal convergence region.

Theorem 1 *Let $\mathcal{Y}_k \in \mathbb{R}^{[n \times (m \times m)]}$ be bounded output with $N(\mu, \sigma^2 < \infty)$. There is a diagonal stochastic estimator $\widehat{\mathcal{A}}_k \in \mathbb{R}^{[n \times (m \times m)]}$ given by (4).*

$$\widehat{\mathcal{A}}_k = E\{\mathcal{X}_k \mathcal{M}_{k-1}^\diamond\} \quad (4)$$

Where the operator E represents a mathematical expectation and is a linear operator on the state space, $\mathcal{X}_k \in \mathbb{R}^{[n \times (m \times m)]}$ and $\mathcal{M}_{k-1}^\diamond \in \mathbb{R}^{[n \times (m \times m)]}$ are diagonals innovation process and correlation matrix, respectively.

Proof 1. See annex in Sec. 6.

Theorem 2 *The diagonal stochastic estimator (4) has an optimal convergence given by (5) in almost all points (a.a.p.).*

$$\widehat{\mathcal{A}}_k \xrightarrow{\text{a.a.p.}} \mathcal{A}_k \pm \epsilon \tag{5}$$

Where $\epsilon \in \mathbb{R}^{[n \times (m \times m)]}$.

Proof 2. See annex in Section 6.

3. Simulation results

This section shows two examples with $m = 2$ and $m = 3$. In both cases, the observable signals, y_{i1k} , as random real num-

bers with a normal distribution function $N(0, 1)$ were proposed. The noise, w_{i1k} , is bounded by $[0, 0.01]$, $i \in [1, m]$. The system for $m = 2$ is given by (6) and for $m = 3$ by (8), respectively. Figure 1, shows a simulation for $m = 2$ and Fig. 2, for $m = 3$. The observable signals are in blue and the estimated signals in red. The estimated signals have good approximation in probability with $\{|a_{ij}|\} \leq 1$ and $\sigma^2\{w_{ij}\} < \infty$.

The model (1) for the case $m = 2$ is written as (6), where $Y_k, Y_{k-1}, W_k \in \mathbb{R}^{[2 \times 1]}$ and $A_k \in \mathbb{R}^{[2 \times 2]}$.

$$\begin{bmatrix} y_{11k} \\ y_{21k} \end{bmatrix} = \begin{bmatrix} a_{11k} & a_{12k} \\ a_{21k} & a_{22k} \end{bmatrix} \begin{bmatrix} y_{11k-1} \\ y_{21k-1} \end{bmatrix} + \begin{bmatrix} w_{11k} \\ w_{21k} \end{bmatrix}. \tag{6}$$

The diagonal form for (6) is given by (7).

$$\begin{bmatrix} y_{11k} - w_{11k} & 0 & | & 0 & 0 \\ 0 & y_{21k} - w_{21k} & | & 0 & 0 \\ \hline 0 & 0 & | & y_{21k} - w_{21k} & 0 \\ 0 & 0 & | & 0 & y_{11k} - w_{11k} \end{bmatrix} = \begin{bmatrix} a_{11k} & 0 & | & 0 & 0 \\ 0 & a_{22k} & | & 0 & 0 \\ \hline 0 & 0 & | & a_{12k} & 0 \\ 0 & 0 & | & 0 & a_{21k} \end{bmatrix} \begin{bmatrix} y_{11k-1}^2 & 0 & | & 0 & 0 \\ 0 & y_{21k-1}^2 & | & 0 & 0 \\ \hline 0 & 0 & | & y_{21k-1}^2 & 0 \\ 0 & 0 & | & 0 & y_{11k-1}^2 \end{bmatrix}. \tag{7}$$

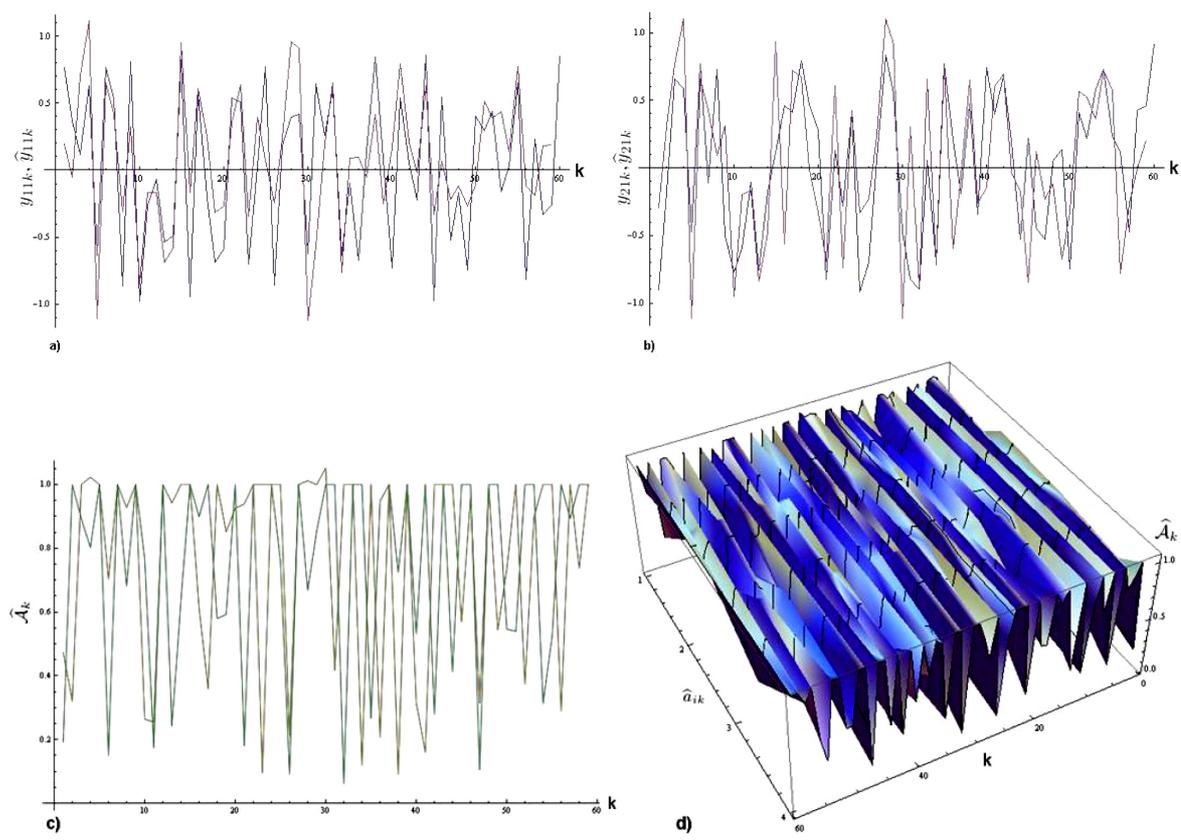


FIGURE 1. Case $m=2$. Observable signal and its identification components. **a)** First, y_{11k}, \hat{y}_{11k} , **b)** Second, y_{21k}, \hat{y}_{21k} , **c)** Estimated values for $\hat{\mathcal{A}}_k$, **d)** Three-dimensional plot of a surface representing $\hat{\mathcal{A}}_k$.

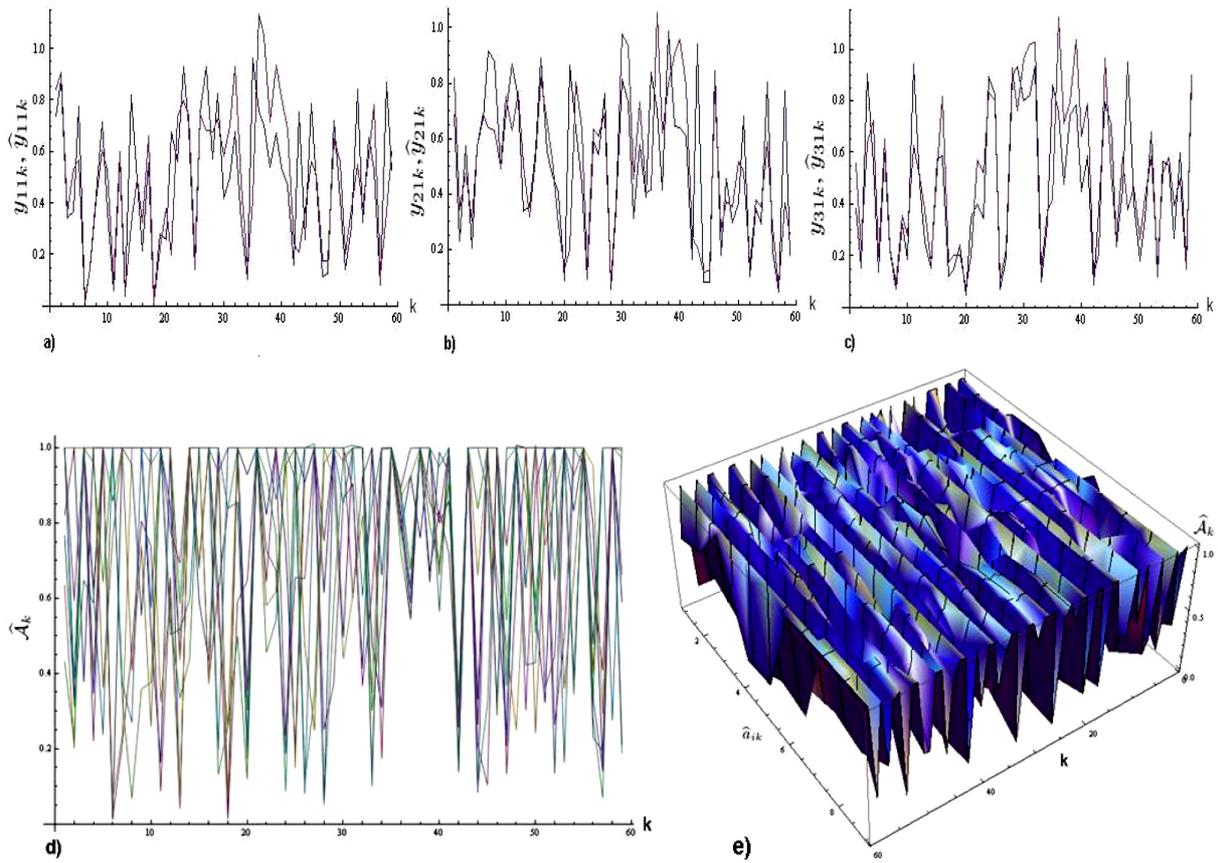


FIGURE 2. Case $m=3$. Observable signal and its identification components. **a)** First, y_{11k}, \hat{y}_{11k} , **b)** Second, y_{21k}, \hat{y}_{21k} , **c)** Third, y_{31k}, \hat{y}_{31k} , **d)** Estimated values for \hat{A}_k , **e)** Three-dimensional plot of a surface representing \hat{A}_k .

The identification proposed by Medel *et al.* [16] considering (3) with respect to (7) is shown in Fig. 1.

For case $m = 2$, Fig. 1 separately shows each of the system components. Figure 1.a) shows in blue, the first observable, y_{11k} , and in red, the first identification, \hat{y}_{11k} . Figure 1.b) shows in blue, the second observable, y_{21k} , and in red, the second identification, \hat{y}_{21k} . As shown, the estimator is very accurate, because it keeps the each change rate signal. Figure 1.c) shows the estimated values for \hat{A}_k . Finally, Fig. 1.d) shows the surfaces corresponding to each of the data $\hat{a}_{11k}, \hat{a}_{12k}, \hat{a}_{21k}, \hat{a}_{22k}$.

The model (1) for the case $m = 3$ is written as (8), where $Y_k, Y_{k-1}, W_k \in \mathbb{R}^{[3 \times 1]}$ and $A_k \in \mathbb{R}^{[3 \times 3]}$.

$$\begin{bmatrix} y_{11k} \\ y_{21k} \\ y_{31k} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} y_{11k-1} \\ y_{21k-1} \\ y_{31k-1} \end{bmatrix} + \begin{bmatrix} w_{11k} \\ w_{21k} \\ w_{31k} \end{bmatrix}. \tag{8}$$

The diagonal form for (8) is given by (9).

$$\begin{bmatrix} y_{11k} - w_{11k} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & y_{21k} - w_{21k} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & y_{31k} - w_{31k} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & y_{21k} - w_{21k} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & y_{31k} - w_{31k} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & y_{11k} - w_{11k} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & y_{31k} - w_{31k} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & y_{11k} - w_{11k} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & y_{21k} - w_{21k} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11k} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & a_{22k} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{33k} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{12k} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{23k} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{31k} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a_{13k} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{21k} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{32k} \end{bmatrix} \cdot \begin{bmatrix} y_{11k-1}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & y_{21k-1}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & y_{31k-1}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & y_{21k-1}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & y_{31k-1}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & y_{11k-1}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & y_{31k-1}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & y_{11k-1}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & y_{21k-1}^2 \end{bmatrix}. \tag{9}$$

The identification proposed by Medel *et al.* [16] considering (3) with respect to (9) is shown in Fig. 2.

For case $m = 3$, Fig. 2 separately shows each of the system components. Figure 2.a) shows in blue, the first observable, y_{11k} , and in red, the first identification, \hat{y}_{11k} . Figure 2.b) shows in blue, the second observable, y_{21k} , and in red, the second identification, \hat{y}_{21k} . Figure 2.c) shows in blue, the third observable, y_{31k} , and in red, the third identification, \hat{y}_{31k} . Figure 2.d) shows the estimated values for \hat{A}_k . As in the case $m = 2$, the estimator is very accurate. Finally, Fig. 2.e) shows the surfaces corresponding to each of the data $\hat{a}_{11k}, \hat{a}_{12k}, \hat{a}_{13k}, \hat{a}_{21k}, \hat{a}_{22k}, \hat{a}_{23k}, \hat{a}_{31k}, \hat{a}_{32k}, \hat{a}_{33k}$.

4. Discussion

The m -dimensional state space representation for Y_k and W_k , is a linear combination of (1). The diagonal is defined using the Jordan canonical form [12,24]. Each diagonal entry has $\mathcal{Y}_k = \text{diag}\{Y_{k_i}\}_{i=1}^{nm}$, $\mathcal{W}_k = \text{diag}\{W_{k_i}\}_{i=1}^{nm}$ and $\mathcal{A}_k = \text{diag}\{A_{k_i}\}_{i=1}^{nm}$ then the diagonal representation for the state space is given by (3).

For this case, the instrumental variable was proposed as $\vartheta_k = \mathcal{Y}_{k-1}^T$, that provides statistical properties for the estimator (4), where $\mathcal{M}_{k-1} = \mathcal{Y}_{k-1}\vartheta_k$ is invertible, *i.e.*, \mathcal{M}_{k-1}^{-1} exists. This representation reduced the computational complexity, because to compute \mathcal{M}_{k-1}^{-1} is simply to evaluate $1/(\mathcal{Y}_{ik-1}\vartheta_{ik})$ in diagonal structure and takes $O(j), j \in \mathbb{N}$.

In the final procedure, computing the dot products, the estimator \hat{A}_k is just to multiply each diagonal entry that is $O(1)$ by the total dimension j having $O(j)$.

The computational cost for implementing the pseudo-inverse in a traditional instrumental variable shown in Fig. 3.a). For each j in pseudo-inverse case element in the observable signal increases in order of j^3 and compared with the diagonalization proposed in this paper shows a linear dependence with those elements, see Fig. 3.b), observing a lower computational complexity. This has important implications for data storage and computing when evaluating very large systems.

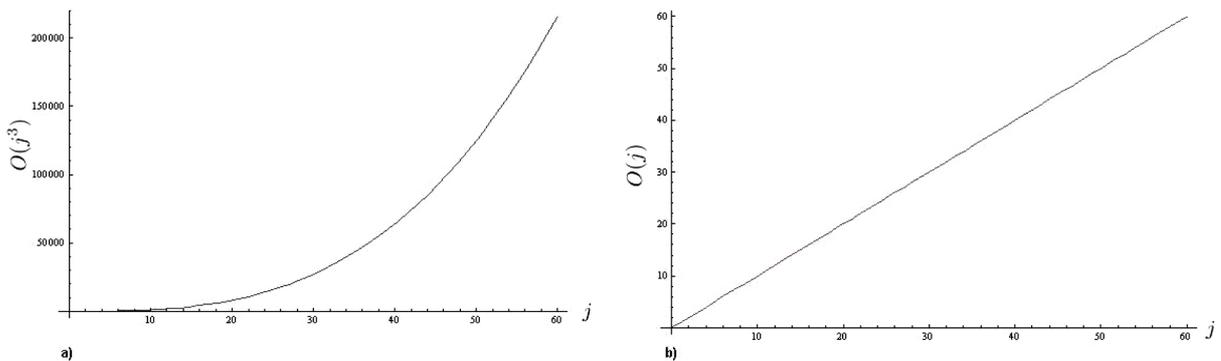


FIGURE 3. Computational complexity methods: **a)** Pseudo-inverse, **b)** Diagonal-inverse.

5. Conclusions

This paper described an estimator for a m -dimensional system without using the pseudo-inverse traditional calculation. The system was transformed in diagonal form, having an invertible condition estimate the matrix parameters with lower resources than traditional techniques. It is noteworthy that the proposed method does not require calculating eigenvalues and eigenvectors which is the principal contribution of this work. The theorems allow defining a stochastic estimator for black-box systems in a m -dimensional state space.

The estimator is optimal because there is a convergence region satisfying Theorem 2 and is very accurate because it keeps the change rates of each signal, regardless of the original signal length or the state space dimension. In order to avoid the singularities in the system, necessary conditions are given for the diagonalization process. The computational complexity for this new calculation method is $O(j)$, where j is the diagonal matrix dimension. The figures above show the algorithm is consistent and easy to implement.

6. Annex

This section gives the proofs for Theorem 1 and Theorem 2.

Proof 1 (Theorem 1). *By induction on state space dimension.*

Case $n = 1$. *The system (1) in a canonical form is represented by (10) with $i = 1$.*

$$y_{i1k} = a_{i1k}y_{i1k-1} + w_{i1k} \tag{10}$$

Let $v_{i1k} = y_{i1k} - w_{i1k}$ be the innovation process, a_{11k} a constant and considering the second probability moment in (10) with the instrumental variable ϑ_k as a function of y_{11k-1} , having (11).

$$E\{v_{11k}\vartheta_k\} = E\{a_{11k}y_{11k-1}\vartheta_k\} \tag{11}$$

Let $x_{11k} = v_{11k}\vartheta_k$.

$$\begin{aligned} E\{x_{11k}\} &= \hat{a}_{11k}E\{y_{11k-1}\vartheta_k\} \\ \hat{a}_{11k} &= E\{x_{11k}\}E\{(y_{11k-1}\vartheta_k)^{-1}\} \end{aligned}$$

Let $m_{11k-1} = (y_{11k-1}\vartheta_k)^{-1}$, having (12).

$$\begin{aligned} \hat{a}_{11k} &= E\{x_{11k}\}E\{m_{11k-1}\} \\ &= E\{x_{11k}m_{11k-1}\} \end{aligned} \tag{12}$$

Let

$$\mathcal{X}_k = \text{diag}(X_{nk}), \mathcal{A}_k = \text{diag}(A_{nk}), \mathcal{M}_{k-1}^\circ = \text{diag}(m_{nk-1})$$

with $i = 1$, having (13).

$$\hat{\mathcal{A}}_k = E\{\mathcal{X}_k \mathcal{M}_{k-1}^\circ\} \tag{13}$$

Case $n = 2$. *The system (1) in an extended canonical form is represented by (14).*

$$\begin{bmatrix} y_{11k} \\ y_{21k} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{11k-1} \\ y_{21k-1} \end{bmatrix} + \begin{bmatrix} w_{11k} \\ w_{21k} \end{bmatrix} \tag{14}$$

Let $v_{i1k} = y_{i1k} - w_{i1k}$ be the innovation process with $i \in [1, 2]$ and $V_k = A_k Y_{k-1}$ in a matricial form is given by (15).

$$\begin{bmatrix} v_{11k} \\ v_{21k} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{11k-1} \\ y_{21k-1} \end{bmatrix} \tag{15}$$

Let A_k a constant symmetric matrix and considering the second probability moment in (15) with the instrumental variable ϑ_k as a function of Y_{k-1}^T , having (16).

$$E\{V_k \vartheta_k\} = E\{A_k Y_{k-1} \vartheta_k\} \tag{16}$$

Let (17) be the diagonal form for the system without losing its properties.

$$E \left\{ \begin{bmatrix} x_{11k} & 0 & | & 0 & 0 \\ 0 & x_{22k} & | & 0 & 0 \\ \hline 0 & 0 & | & x_{12k} & 0 \\ 0 & 0 & | & 0 & x_{21k} \end{bmatrix} \right\}$$

$$= \begin{bmatrix} a_{11k} & 0 & | & 0 & 0 \\ 0 & a_{22k} & | & 0 & 0 \\ \hline 0 & 0 & | & a_{12k} & 0 \\ 0 & 0 & | & 0 & a_{21k} \end{bmatrix} E \left\{ \begin{bmatrix} y_{11k-1}\vartheta_{11k} & 0 & | & 0 & 0 \\ 0 & y_{21k-1}\vartheta_{21k} & | & 0 & 0 \\ \hline 0 & 0 & | & y_{21k-1}\vartheta_{21k} & 0 \\ 0 & 0 & | & 0 & y_{11k-1}\vartheta_{11k} \end{bmatrix} \right\} \quad (17)$$

Let (18) be the simplified form for (17), where $\mathcal{X}_k = \text{diag}(X_{nk}), \mathcal{A}_k = \text{diag}(A_{nk}), \mathcal{Y}_{k-1} = \text{diag}(Y_{nk-1})$, with $n = 1, 2$.

$$E\{\mathcal{X}_k\} = \widehat{\mathcal{A}}_k E\{\mathcal{Y}_{k-1}\vartheta_k\}. \quad (18)$$

Since \mathcal{A}_k is constant and $\det(\mathcal{Y}_{k-1}\vartheta_k) \neq 0$ therefore $\mathcal{Y}_{k-1}\vartheta_k$ is invertible, having (19).

$$\widehat{\mathcal{A}}_k = E\{\mathcal{X}_k\}E\{(\mathcal{Y}_{k-1}\vartheta_k)^{-1}\} \quad (19)$$

Let $\mathcal{M}_{k-1}^\diamond = (\mathcal{Y}_{k-1}\vartheta_k)^{-1}$, then from (19) results (20).

$$\widehat{\mathcal{A}}_k = E\{\mathcal{X}_k\mathcal{M}_{k-1}^\diamond\} \quad (20)$$

Case $n = m$. Let (21) be the extended canonical form for the system (1).

$$\begin{bmatrix} y_{11k} \\ y_{21k} \\ \vdots \\ y_{m1k} \end{bmatrix} = \begin{bmatrix} a_{11k} & a_{12k} & \cdots & a_{1mk} \\ a_{21k} & a_{22k} & \cdots & a_{2mk} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1k} & a_{m2k} & \cdots & a_{mmk} \end{bmatrix} \begin{bmatrix} y_{11k-1} \\ y_{21k-1} \\ \vdots \\ y_{m1k-1} \end{bmatrix} + \begin{bmatrix} w_{11k} \\ w_{21k} \\ \vdots \\ w_{m1k} \end{bmatrix} \quad (21)$$

Let $v_{i1k} = y_{i1k} - w_{i1k}$ be the innovation process with $i \in [1, m]$ and $V_k = A_k Y_{k-1}$ in a matricial form is given by (22).

$$\begin{bmatrix} v_{11k} \\ v_{21k} \\ \vdots \\ v_{m1k} \end{bmatrix} = \begin{bmatrix} a_{11k} & a_{12k} & \cdots & a_{1mk} \\ a_{21k} & a_{22k} & \cdots & a_{2mk} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1k} & a_{m2k} & \cdots & a_{mmk} \end{bmatrix} \begin{bmatrix} y_{11k-1} \\ y_{21k-1} \\ \vdots \\ y_{m1k-1} \end{bmatrix} \quad (22)$$

Let A_k a constant symmetric matrix and considering the second probability moment in (22) with the instrumental variable ϑ_k as a function of Y_{k-1}^T , having (23).

$$E\{V_k\vartheta_k\} = E\{A_k Y_{k-1}\vartheta_k\} \quad (23)$$

Then, $\det(\mathcal{Y}_{k-1}\vartheta_k) \neq 0 \Rightarrow (\mathcal{Y}_{k-1}\vartheta_k)^{-1}$ exists, therefore (19) and (20) holds for $n = m$.

Case $n = m + 1$. Let (24) be the extended canonical form for the system (1).

$$\begin{bmatrix} y_{11k} \\ y_{21k} \\ \vdots \\ y_{m1k} \\ y_{(m+1)1k} \end{bmatrix} = \begin{bmatrix} a_{11k} & a_{12k} & \cdots & a_{1mk} \\ a_{21k} & a_{22k} & \cdots & a_{2mk} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1k} & a_{m2k} & \cdots & a_{mmk} \\ a_{(m+1)1k} & a_{(m+1)2k} & \cdots & a_{(m+1)(m+1)k} \end{bmatrix} \begin{bmatrix} y_{11k-1} \\ y_{21k-1} \\ \vdots \\ y_{m1k-1} \\ y_{(m+1)1k-1} \end{bmatrix} + \begin{bmatrix} w_{11k} \\ w_{21k} \\ \vdots \\ w_{m1k} \\ w_{(m+1)1k} \end{bmatrix} \quad (24)$$

Let $v_{i1k} = y_{i1k} - w_{i1k}$ be the innovation process with $i \in [1, m + 1]$ and $V_k = A_k Y_{k-1}$ in a matricial form is given by (25).

$$\begin{bmatrix} v_{11k} \\ v_{21k} \\ \vdots \\ v_{m1k} \\ v_{(m+1)1k} \end{bmatrix} = \begin{bmatrix} a_{11k} & a_{12k} & \cdots & a_{1mk} \\ a_{21k} & a_{22k} & \cdots & a_{2mk} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1k} & a_{m2k} & \cdots & a_{mmk} \\ a_{(m+1)1k} & a_{(m+1)2k} & \cdots & a_{(m+1)(m+1)k} \end{bmatrix} \begin{bmatrix} y_{11k-1} \\ y_{21k-1} \\ \vdots \\ y_{m1k-1} \\ y_{(m+1)1k-1} \end{bmatrix} \quad (25)$$

Let A_k a constant symmetric matrix and considering the second probability moment in (25) with the instrumental variable ϑ_k as a function of Y_{k-1}^T , having (26).

$$E\{V_k\vartheta_k\} = E\{A_k Y_{k-1}\vartheta_k\} \quad (26)$$

As in case $n = m$, $\det(\mathcal{Y}_{k-1}\vartheta_k) \neq 0 \Rightarrow (\mathcal{Y}_{k-1}\vartheta_k)^{-1}$ exists, therefore (19) and (20) holds for $n = m + 1$.

System model (1) case, uses an estimated representation given by (4), then the diagonal estimator $\widehat{\mathcal{A}}_k$ is defined by (27).

$$\widehat{\mathcal{A}}_k = E\{\mathcal{X}_k\mathcal{M}_{k-1}^\diamond\} \quad (27)$$

■

Proof 2 (Theorem 2). Let e_k be the error defined by $e_k = \mathcal{Y}_k - \widehat{\mathcal{Y}}_k$, where $\mathcal{Y}_k = \mathcal{A}_k \mathcal{Y}_{k-1} + \mathcal{W}_k$ and $\widehat{\mathcal{Y}}_k = \widehat{\mathcal{A}}_k \widehat{\mathcal{Y}}_{k-1} + \mathcal{W}_k$. Let $J_k = E\{e_k e_k^T\}$ be the functional error, its gradient with respect to $\widehat{\mathcal{A}}_k$ is $\nabla_{\widehat{\mathcal{A}}_k} J_k$ and is needed to minimize it evaluating $\nabla_{\widehat{\mathcal{A}}_k} J_k = 0$.

$$0 = \nabla_{\widehat{\mathcal{A}}_k} J_k \quad (28)$$

$$\begin{aligned} &= \nabla_{\widehat{\mathcal{A}}_k} E\{(\mathcal{A}_k \mathcal{Y}_{k-1} - \widehat{\mathcal{A}}_k \widehat{\mathcal{Y}}_{k-1})^2\} \\ &= 2\widehat{\mathcal{A}}_k \widehat{\mathcal{Y}}_{k-1}^2 - 2\widehat{\mathcal{Y}}_{k-1} \mathcal{A}_k \mathcal{Y}_{k-1} - 2\widehat{\mathcal{Y}}_{k-1} \widehat{\mathcal{W}}_k^2 \\ &= \widehat{\mathcal{A}}_k \widehat{\mathcal{Y}}_{k-1}^2 - \widehat{\mathcal{Y}}_{k-1} \mathcal{A}_k \mathcal{Y}_{k-1} - \widehat{\mathcal{Y}}_{k-1} \widehat{\mathcal{W}}_k^2 \\ \widehat{\mathcal{A}}_k \widehat{\mathcal{Y}}_{k-1}^2 &= \widehat{\mathcal{Y}}_{k-1} \mathcal{A}_k \mathcal{Y}_{k-1} + \widehat{\mathcal{Y}}_{k-1} \widehat{\mathcal{W}}_k^2 \\ \widehat{\mathcal{A}}_k &= \widehat{\mathcal{Y}}_{k-1} \mathcal{A}_k \mathcal{Y}_{k-1} (\widehat{\mathcal{Y}}_{k-1}^2)^{-1} + \widehat{\mathcal{Y}}_{k-1} \widehat{\mathcal{W}}_k^2 (\widehat{\mathcal{Y}}_{k-1}^2)^{-1} \\ \widehat{\mathcal{A}}_k &\sim \mathcal{A}_k \mathcal{Y}_{k-1}^2 (\widehat{\mathcal{Y}}_{k-1}^2)^{-1} + \widehat{\mathcal{W}}_k^2 (\widehat{\mathcal{Y}}_{k-1}^2)^{-1} \end{aligned} \quad (29)$$

$$\sim \mathcal{A}_k + \widehat{\mathcal{W}}_k^2 \widehat{\mathcal{Y}}_{k-1}^{-1} \quad (30)$$

In this case the estimation converges in almost all points (a. a. p.).

$$\therefore \widehat{\mathcal{A}}_k \xrightarrow{a.a.p.} \mathcal{A}_k \pm \epsilon; \quad \epsilon = \widehat{\mathcal{W}}_k^2 \widehat{\mathcal{Y}}_{k-1}^{-1} \quad (31)$$

■

Acknowledgement

This document was checked by English Teacher Lawrence Whitehill, England.

-
1. J.L. Devore, *Probabilidad y Estadística para Ingeniería y Ciencias* (International Thomson Editores. México, 2002).
 2. G. Casella and R. Berger, *Statistical Inference* 2nd edit., (Duxbury Advanced Series, Pacific Grove, CA 2002).
 3. A.C. Aitken, *Proc. R. Soc. Edinburg* **55** (1935) 42-48.
 4. R.L. Plackett, *Biometrika* **37** (1950) 149-157.
 5. H. Abdi, Least-squares, in: M. Lewis-Beck, A. Bryman, T. Futing (Eds.), *Encyclopedia for research methods for the social sciences*. (Thousand Oaks, CA, Sage 2003). pp. 792-795.
 6. J. Angrist, A. Krueger, *J. Econ. Perspect.* **15** (2001)69-85.
 7. J. Pearl, *Causality: Models, Reasoning and Inference* (Cambridge University Press, 2000).
 8. G. Imbens, J. Angrist, *Econometrica* **62** (1994) 467-476.
 9. C.R. Nelson and R. Startz, *Econometrica* **58** (1990) 967-976.
 10. J. Stock, J. Wright and M. Yogo, *J. Am. Stat. Assoc.* **20** (2002) 518-29.
 11. M. Drton, B. Sturmfels, and S. Sullivant, *Lectures on Algebraic Statistics* (Springer, 2009).
 12. P.L. Falb, *Methods of algebraic geometry in control theory* (Birkhauser, 1990).
 13. D.A. Cox, B. Sturmfels, D.N. Manocha, *Applications of computational algebraic geometry* (AMS Bookstore, 1997).
 14. B. Juttler and R. Piene, *Geometric modeling and algebraic geometry* (Springer, 2007).
 15. J.. Medel Juárez, J.C. García Infante, and R. Urbieto Parrazales, *Rev. Mex. Fís.* **57** (2011) 0413.
 16. J. de J. Medel Juárez, R.U. Parrazales, and R.P. Orozco, *Rev. Mex. Fís.* **57** (2011) 0204.
 17. J.J. Medel Juárez, C.V. García Mendoza, *Rev. Mex. Fís.* **56** (2010) 0054.
 18. J.J. Medel Juárez, M.T. Zagaceta Álvarez, *Rev. Mex. Fís.* **56** (2010) 0001.
 19. A.S. Householder, *Unitary Triangularization of a Nonsymmetric Matrix* (1958).
 20. L.N. Trefethen, *Society for Industrial and Applied Mathematics* (1997).
 21. J. Demmel, W. Kahan, *Journal on Scientific and Statistical Computing* **11** (1990) 873-912.
 22. G.H. Golub, W. Kahan, *Calculating the singular values and pseudo-inverse of a matrix, Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis* **2** (1965) 205-224.
 23. R.A. Horn, C.R. Johnson, *Matrix Analysis* (Sec. 7.3. Cambridge University Press, 1985).
 24. R.A. Horn, C.R. Johnson, *Topics in Matrix Analysis* (Chapter 3., Cambridge University Press, 1991).
 25. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes: The Art of Scientific Computing* (Sec. 2.6 New York, Cambridge University Press, 2007).