OPTIMAL CONTROL FOR ACTIVE IDENTIFICATION OF UNKNOWN SYSTEMS

M. Baglietto, G. Cannata, L. Scardovi, and R. Zoppoli

Department of Communications, Computer and System Sciences, DIST–University of Genoa, Via Opera Pia 13, 16145 Genova, Italy. E-mail: {mbaglietto,cannata,lucas,rzop}@dist.unige.it

Abstract: In this paper we consider the Problem of actively identifying the measurement equation of a linear time invariant dynamic system by means of a suitable control law. A sufficient and necessary condition will be given on the identifiability of the system. Under such a condition, the problem can be formulated as an optimal control one, in which the minimization of a suitable "uncertainty measure" is performed. To this end, the use of the Shannon Entropy is proposed and motivated.

Keywords: Dynamic Systems, Entropy, System Identification, Parameter Estimation, Optimal Control.

1. INTRODUCTION

Usually, system identification consists in estimating the unknown parameters for a given control law, or even for a given control signal. As a matter of fact, the control signal influences the properties of convergence of the parameter estimate. Therefore, it is interesting to asses which control strategy could maximize the information gained by a parameter estimator. Such a paradigm where control is also finalized to identification is called *Active Identification*.

In statistics a similar problem is the Optimal Experiment Design (OED), in which one has to design an experiment in order to infer about an unknown parameterized system (Fedorov, 1972; Ghahramani *et al.*, 1996). Also in machine learning a similar problem arises when one can choose the input patterns to optimize an approximator (active learning) (MacKay, 1992). In robotics, the problem of environment exploration can be formulated as a particular case of the Active Identification Problem, and it has been studied from an heuristic point of view (Baglietto *et al.*, 2002; Baglietto *et al.*, 2003; Yamauchi, 1997). In these last years some researchers have used information theoretic concepts to study control problems (see (Saridis, 1988; Loparo *et al.*, 1997)). In (Baglietto *et al.*, 2004) the problem is stated in a general framework and an approximate solution is presented. In this paper we formulate the problem in an information theoretic setting by using the Shannon entropy as a measure of information about a set of unknown parameters. Our attention is on LTI systems with an unknown measurement channel. This restriction allows us to take on an identifiability study of the system, and to exploit in a simple form the information measure. This allows us to find the control sequence that extracts the maximum possible amount of information about the unknown parameters.

This paper is organized as follows: in Section 2, the problem formulation is given. In Section 3 a condition for the identifiability of the system is stated. In particular it will be proven that this property depends from the reachability of the system. In Section 4 the active identification problem is solved by means of a entropy minimization approach. In Section 5 some simulation results are presented. The proofs of the theorems are reported in the Appendix.

2. PROBLEM FORMULATION

Let us consider a discrete-time linear system:

$$\boldsymbol{x}_{t+1} = A\boldsymbol{x}_t + B\boldsymbol{u}_t\,, \quad t = 0, \ldots, T-1 \tag{1a}$$

$$\boldsymbol{y}_t = C(\boldsymbol{\theta})\boldsymbol{x}_t + \boldsymbol{\eta}_t\,, \quad t = 0, \ldots, T \tag{1b}$$

where $t = 1, \ldots, T$ is the time instant, $\boldsymbol{x}_t \in \mathbb{R}^n$ is the state vector, $\boldsymbol{y}_t \in \mathbb{R}^h$ is the vector of measures, $\boldsymbol{u}_t \in U \subseteq \mathbb{R}^m$ is the control vector, $\boldsymbol{\theta} \in \mathbb{R}^{nh}$ is a vector of unknown parameters and $\boldsymbol{\eta}_t$ is a disturbance vector. Let us assume in the following that $\boldsymbol{x}_0 = \bar{\boldsymbol{x}}$ is a known initial condition. The matrix $C(\boldsymbol{\theta})$ in equation (1b) is defined as

$$C(\boldsymbol{\theta}) = [\theta_{ij}]\,.$$

Since equation (1b) is bilinear in x_t and θ_{ij}, system (1) can be rewritten as

$$\boldsymbol{x}_{t+1} = A\boldsymbol{x}_t + B\boldsymbol{u}_t\,, \quad t = 0, \ldots, T-1 \tag{2a}$$

$$\boldsymbol{y}_t = X_t'\boldsymbol{\theta} + \boldsymbol{\eta}_t\,, \quad t = 0, \ldots, T \tag{2b}$$

where

$$X_t' \triangleq \begin{bmatrix} \boldsymbol{x}_t' & 0 & \ldots & 0 \\ 0 & \boldsymbol{x}_t' & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \boldsymbol{x}_t' \end{bmatrix}$$

and

$$\boldsymbol{\theta} \triangleq \left[\, \theta_{1,1}\ \theta_{1,2}\ \ldots\ \theta_{1,n}, \theta_{2,1}, \ldots, \theta_{n,n} \,\right]'$$

is the stacking of the rows (transposed) of the matrix $C(\boldsymbol{\theta})$. It is worth noting that (2b) is the linear regression form of (1b).

We will adopt the following notation

$$\boldsymbol{u}_0^{T-1} = \mathrm{col}\,[\boldsymbol{u}_0, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_{T-1}]\,.$$

The vector $\boldsymbol{\theta}$ is here considered as a vector of initially unknown parameters to be estimated. In order to do this, the Best Linear Unbiased Estimator is assumed to be used.

The key problem of Active Identification addressed in this paper is to control the system (1a) in order to obtain an *optimal* regressor matrix. The term *optimal* is intended here with respect to the minimization of a suitable *information measure* $\mathcal{L}_T(\cdot)$. The study of the information measure $\mathcal{L}_T(\cdot)$ will be the argument of Section 4.

Formally the problem can be stated as follows

Problem 1. (*A*ctive identification problem) Given a suitable information measure $\mathcal{L}_T(\cdot)$ (related to the vector $\boldsymbol{\theta}$), find the optimal control sequence $(\boldsymbol{u}_0^{T-1})^\circ$ such that

$$(\boldsymbol{u}_0^{T-1})^\circ = \arg\min_{\boldsymbol{u}_0^{T-1}} \mathcal{L}_T(\boldsymbol{u}_0^{T-1})$$

subject to (1a) and

$$\boldsymbol{u}_0^{T-1} \in \mathcal{U}$$

$$\boldsymbol{x}_t \in X \subseteq \mathbb{R}^n$$

where $\mathcal{U}$ is the bounded set of the admissible control sequences. □

Problem 1 implicitly assumes the identifiability of the vector $\boldsymbol{\theta}$ and intuitively requires some degree of reachability of the system (1). Indeed in the next section it will be proven that the identifiability of the system is strictly related to its reachability.

3. IDENTIFIABILITY ANALYSIS

The following theorem follows directly from the definition of a *completely observable system* (see for example (Jazwinski, 1970)).

Theorem 1. System (1) is *completely identifiable* (in $t \geq T$ stages) iff a time horizon T and a control sequence $\boldsymbol{u}_0^{T-1}$ exist such that:

$$\sum_{t=0}^{T} X_t X_t' > 0. \tag{3}$$

□

By defining

$$\Phi_T' \triangleq \left[\, X_0\ X_1\ \ldots\ X_T \,\right]$$

equation (3) can be rewritten in a more concise form as

$$\Phi_T'\Phi_T > 0.$$

If $\Phi_T'\Phi_T$ is singular then certain linear combinations of the elements of $\boldsymbol{\theta}$ cannot be determined; in this case, no information about them can be extracted from the data $\{\boldsymbol{y}_0, \ldots, \boldsymbol{y}_T\}$.

In the following it will be proved, that complete identifiability (for short *identifiability*) is a structural property. Indeed, the following result states the equivalence of reachability and identifiability.

Theorem 2. The system (1) is identifiable in n stages (and consequently in $T > n$ stages) from the state zero iff it is completely reachable, i.e., iff

$$rank\,(K) = n$$

where

$$K = \left[\, B\ AB\ \ldots\ A^{n-1}B \,\right].$$

□

Proof (see the Appendix).

Given a identifiable system, only a subset of the control sequences guarantee the identification of the parameters. These sequences will be called *proper sequences.*

Let us consider the following constructive procedure to fix a control sequence:

Procedure 1. Let us define

$$U \triangleq \begin{bmatrix} \boldsymbol{u}_0 & \boldsymbol{u}_1 & \boldsymbol{u}_2 & \dots & \boldsymbol{u}_{n-1} \\ 0 & \boldsymbol{u}_0 & \boldsymbol{u}_1 & \dots & \boldsymbol{u}_{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \boldsymbol{u}_0 & \boldsymbol{u}_1 \\ 0 & 0 & 0 & \dots & \boldsymbol{u}_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots, \boldsymbol{v}_n \end{bmatrix}, \tag{4}$$

where

$$\boldsymbol{v}_i, \quad i = 1, \dots, n$$

are the column vectors of U. We want to find a control sequence $\boldsymbol{u}_0^{n-1}$ such that

$$\boldsymbol{v}_i \notin ker\,(K)\ , \quad i = 0, \dots, n-1 \tag{5}$$

Such sequences can be constructed in the following two steps:

- take $\boldsymbol{u}_0 \notin ker\,(B)$, such a vector always exists (except for the trivial case of B null);
- for each $i = 1, \dots, n-1$ take $\boldsymbol{u}_i$ such that the following condition holds

  $$B\boldsymbol{u}_i \neq \sum_{j=1}^{i} A^j B \boldsymbol{u}_{i-j}.$$

  Note that these vectors $\boldsymbol{u}_i$ always exist (except for the trivial case of B null).

□

Corollary 1. If the system (1) is completely reachable then $\boldsymbol{u}_0^{T-1}$ guarantees the identifiability of the parameter vector $\boldsymbol{\theta}$ if it satisfies the conditions expressed in (5) (for short, sequences satisfying (5) will called *proper*). Moreover it can be constructed following the lines of Procedure 1. □

4. ENTROPY BASED ACTIVE IDENTIFICATION

Under the reachability assumption, it is possible to gain all the information about all the parameters. Now we address the problem of finding the control policy which extracts the maximum information. The first step is to define mathematically what the *information* is. We now recall some concepts from statistics, that will be useful in the following. Given a general measurement equation

$$\boldsymbol{y} = g(\boldsymbol{\theta}) + \boldsymbol{\eta}$$

where $\boldsymbol{y}$ is the measure vector, $\boldsymbol{\theta}$ is an unknown parameter vector and $\boldsymbol{\eta}$ is a generic noise, the Fisher Information Matrix (FIM) is defined as the matrix M whose components are

$$m_{i,j}(\boldsymbol{\theta}) = -\operatorname*{E}_{\boldsymbol{y}}\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log p(\boldsymbol{y}|\boldsymbol{\theta})\right]. \tag{6}$$

When we have a prior information about the distribution of the unknown parameters it is possible to extend the definition of the FIM in order to take into consideration also such an information. Then we have:

$$m_{i,j} = -\operatorname*{E}_{\boldsymbol{y},\boldsymbol{\theta}}\left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log p(\boldsymbol{y}, \boldsymbol{\theta})\right], \tag{7}$$

where expectation is taken with respect to the observations and the parameters. In the following we will consider, without loss of generality, the Bayesian setting, and equation (7) will define the Bayesian Fisher Information Matrix (BFIM). A fundamental result and the most important application of the FIM is the Cramer-Rao Theorem, which gives us the lowest possible variance for all unbiased estimators $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ (see for example (Soderstrom and Stoica, 1988)

Theorem 3. For every unbiased estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$, if M is not singular, the following relation holds:

$$Cov(\hat{\boldsymbol{\theta}}) \geq M^{-1}$$

where the inequality indicates that the matrix $Cov(\hat{\boldsymbol{\theta}}) - M^{-1}$ is a non negative definite matrix. □

We now use it to define the most efficient estimator.

Definition 1. An unbiased estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ is said to be *efficient* if it meets the Cramer-Rao bound with equality, i.e., if $Cov(\hat{\boldsymbol{\theta}}) = M^{-1}$. □

The Fisher information matrix is therefore a measure of the amount of information about $\boldsymbol{\theta}$ that is present in the data. It gives a lower bound on the error incurred when estimating $\boldsymbol{\theta}$ from the data. However, it is possible that there does not exist an estimator meeting this lower bound. When the disturbance vectors $\boldsymbol{\eta}_0, \boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_T$ have a Gaussian distribution and are independent and identically distributed, after straightforward calculations the information matrix for our problem becomes:

$$M_T(\boldsymbol{u}_0^{T-1}) = \Phi_T'(\boldsymbol{u}_0^{T-1}) R^{-1} \Phi_T(\boldsymbol{u}_0^{T-1}) \tag{8}$$

where R is the covariance matrix of the vector $\boldsymbol{\eta}$. In this case, the BFIM turns out to be

$$M_T(\boldsymbol{u}_0^{T-1}) = \Sigma^{-1} + \Phi_T'(\boldsymbol{u}_0^{T-1}) R^{-1} \Phi_T(\boldsymbol{u}_0^{T-1}) \tag{9}$$

where Σ is the covariance of the prior parameter density function. It is worth noting that M_T

depends on the whole trajectory of the system, i.e. the sequence of states of the controlled system and consequently on the control sequence $\boldsymbol{u}_0^{T-1}$.

If we consider an efficient unbiased estimator (we choose to use the Best Linear Unbiased Estimator - BLUE), we can consider as a Loss function the *entropy* of the estimator. The entropy measure is related to the dispersion of a random variable X (Cover and Thomas, 1991) and is defined as follows,

$$H(\boldsymbol{X}) = -\int_{\mathcal{X}} p(\boldsymbol{x}) \log p(\boldsymbol{x}). \tag{10}$$

where $\mathcal{X}$ is the support set of the random variable. In the Gaussian case, the entropy of the estimator takes on a very simple form.

$$H_T(\hat{\boldsymbol{\theta}}) = c + \log\det\left(M^{-1}\right)$$

where c is a suitable constant (that will not be considered in the following), and M is the FIM. In this case $H_T(\hat{\boldsymbol{\theta}})$ will be equal to

$$\log\det\left(\Phi_T'(\boldsymbol{u}_0^{T-1})R^{-1}\Phi_T(\boldsymbol{u}_0^{T-1})\right)^{-1}.$$

If the BFIM is considered, $H_T(\hat{\boldsymbol{\theta}})$ will be equal to

$$\log\det\left(\Sigma^{-1} + \Phi_T'(\boldsymbol{u}_0^{T-1})R^{-1}\Phi_T(\boldsymbol{u}_0^{T-1})\right)^{-1}.$$

Hence the entropy is related to the FIM (or BFIM) and the minimization of the entropy of the estimator is equivalent to the minimization of the determinant of the FIM (or the BFIM). It is worth noting that minimizing the mean square error is equivalent to minimize the trace of the BFIM, while minimizing the entropy of the estimator is equivalent to minimize the determinant of the BFIM. In the particular case where which the measurement is a scalar is possible to derive a simple form for the determinant of the FIM.

Theorem 4. Consider the system (1) where $y_t \in \mathbb{R}$ and $C(\boldsymbol{\theta}) = \boldsymbol{\theta}'$ and σ^2 is the disturbance variance. Then

$$\arg\min_{\boldsymbol{u}_0^{T-1}} \log\det\left(M_T^{-1}\right) =$$
$$\arg\min_{\boldsymbol{u}_0^{T-1}} \left\{-\sum_{t=0}^{T} ln(1 + \frac{1}{\sigma^2}\boldsymbol{x}_t' M_{i-1}^{-1}\boldsymbol{x}_t)\right\}$$

where $M_{-1}^{-1} = \Sigma_0$ is the a priori covariance matrix of the parameters. □

Proof (see the Appendix)

The active identification problem can be stated as:

Problem 2. (Active identification problem)
Find the optimal control sequence $(\boldsymbol{u}_0^{T-1})^{\circ}$ such that

$$\left(\boldsymbol{u}_0^{T-1}\right)^{\circ} = \arg\min_{\boldsymbol{u}_0^{T-1}} \left\{-\sum_{t=0}^{T}\boldsymbol{x}_t' M_{i-1}^{-1}\boldsymbol{x}_t\right\}$$

subject to eq. (1a) and

$$\boldsymbol{u}_0^{T-1} \in \mathcal{U}$$
$$\boldsymbol{x}_t \in X \subset \mathbb{R}^n$$

□

This problem can be solved by means of constrained nonlinear programming techniques.

In the particular case, where the set the state vectors belong to takes on the form $X = \{\boldsymbol{x} : \boldsymbol{x}'\boldsymbol{x} \le c\}$, $c > 0$, a *greedy policy* can be obtained by solving the following problem.

Problem 3.

$$\min_{\boldsymbol{u}_t}\left\{-\boldsymbol{x}_{t+1}' M_t^{-1}\boldsymbol{x}_{t+1}\right\}$$

subject to:

$$\boldsymbol{x}_{t+1} = A\boldsymbol{x}_t + B\boldsymbol{u}_t$$
$$\boldsymbol{u}_t \in U \subset \mathbb{R}^m$$
$$\boldsymbol{x}_t \in X \in \mathbb{R}^n$$

where $X = \{\boldsymbol{x} : \boldsymbol{x}'\boldsymbol{x} \le c\}$, $c > 0$, and $t = 0, 1, \ldots, T-1$.

If $rank\,(B) = n$, by using a simple geometric interpretation (see Fig (1)), it is possible to give an analytic solution of the problem. At every step the optimal control vector is given by the following relation:

$$\boldsymbol{u}_t^{\circ} = (B'B)^{-1}B'\boldsymbol{x}_{t+1}^* - (B'B)^{-1}B'A\boldsymbol{x}_t$$
$$\boldsymbol{x}_{t+1}^* = \frac{c\boldsymbol{v}_t^{\max}}{\|\boldsymbol{v}_t^{\max}\|}, \quad t = 0, \ldots, T-1$$

where $\boldsymbol{v}_t^{\max}$ is the eigenvector relative to the maximum eigenvalue of the matrix M_t^{-1} and $M_{-1} = \Sigma_0^{-1}$.

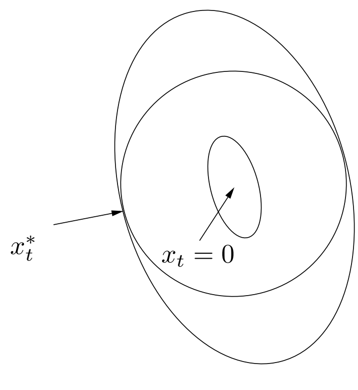

Fig. 1. Geometrical interpretation of the state selection

5. SIMULATION RESULTS

In this section, a statistical analysis of the proposed approach is given. The system considered is the following:

$$\begin{aligned} \boldsymbol{x}_{t+1} &= A\boldsymbol{x}_t + B\boldsymbol{u}_t\,, & t &= 0,\dots,T-1 \\ y_t &= \boldsymbol{\theta}'\boldsymbol{x}_t + \eta_t\,, & t &= 0,\dots,T \end{aligned}$$

where $\boldsymbol{x}_t \in \mathbb{R}^3$, $\boldsymbol{\theta} \in \mathbb{R}^3$ and

$$\boldsymbol{u}_t \in \{\boldsymbol{u} \in \mathbb{R}^2 : \boldsymbol{u}'\boldsymbol{u} \le 9\}.$$

The matrices A and B are extracted randomly to form reachable and stable LTI systems, 1000 instances of the problem have been considered. The noise variance is set to 1. The estimation square error comparing our approach (based on the solution of 2) with a randomly selected control vector is shown in the box plots presented in figure 2. The time horizon T is set to 20 .

From a comparison of the two strategies, one can easily see that our approach improves significantly the parameter estimation with random inputs.

In figure 3 the optimal control is compared with a sinusoidal control. Here the time horizon T is set to 40.

Even in this case the optimal control improves significantly the parameter estimation effectiveness.

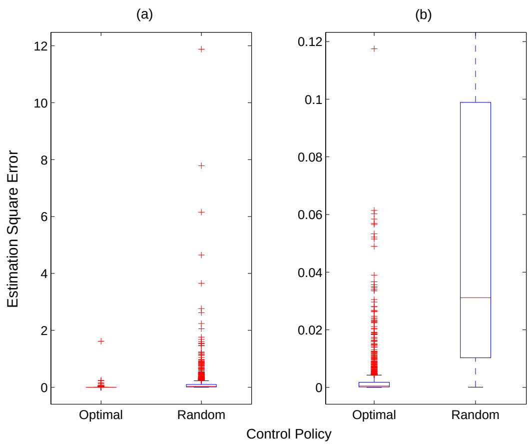

Fig. 2. Box plots of the estimation square error (a) and their enlargement (b)

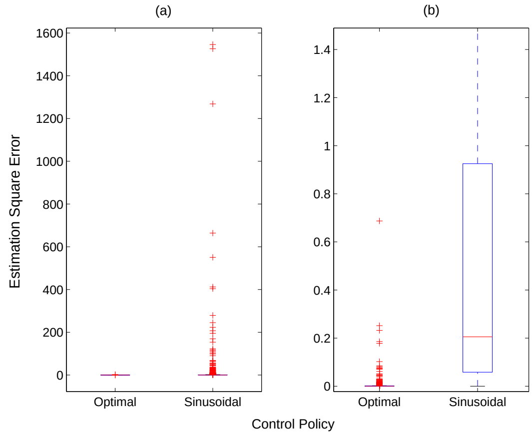

Fig. 3. Box plots of the estimation square error (a) and their enlargement (b)

REFERENCES

Baglietto, M., L. Scardovi and R. Zoppoli (2004). Active identification of unknown systems: an information theoretic approach. In: *42nd IEEE American Control Conference*. Boston. to appear.

Baglietto, M., M. Paolucci, L. Scardovi and R. Zoppoli (2002). Information based multi-agent exploration. In: *IEEE Third International Workshop on Robot Motion and Control*. Bukowy Dworek, Poland. pp. 173–179.

Baglietto, M., M. Paolucci, L. Scardovi and R. Zoppoli (2003). Entropy-based environment exploration and stochastic optimal control. In: *42nd IEEE Conference on Decision and Control*. Maui, Hawaii. pp. 2938–2941.

Cover, T. M. and J. A. Thomas (1991). *Elements of Information Theory*. Wiley.

Fedorov, V. V. (1972). *Theory of optimal experiments*. Academic Press.

Ghahramani, Z., D. A. Cohn and M. I. Jordan (1996). Active learning with statistical models. *J. of Artificial Intelligence Research* **4**, 129–145.

Jazwinski, A. H. (1970). *Filtering Theory*. Academic Press.

Loparo, K. A., X. Feng and Y. Fang (1997). Optimal state estimation for stochastic systems: an information theoretic approach. *IEEE Trans. Automatic Control* **42**, 771–785.

MacKay, D. (1992). Information-based objective functions for active data selection. *Neural Computation* **4**, 590–604.

Saridis, G. N. (1988). Entropy formulation of optimal and adaptive control. *IEEE Trans. Automatic Control* **33**, 713–721.

Soderstrom, T. and P. Stoica (1988). *Sistem Identification*. Prentice-Hall.

Yamauchi, B. (1997). A frontier based approach for autonomous exploration. In: *IEEE International Symposium on Computational Intelligence in Robotics and Automation*. Monterey, CA. pp. 146–151.

Zhang, F. (1999). *Matrix Theory*. Springer-Verlag.

APPENDIX

The following results are required to prove theorem 2 and 4. We recall two Propositions from Linear Algebra (see for example (Zhang, 1999))

Proposition 1. Given two real matrices, $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, the following relations hold:

$$rank\,(AB) \le min(rank\,(A)\,, rank\,(B)) \quad (12a)$$
$$rank\,(AB) \ge rank\,(A) + rank\,(B) - n \quad (12b)$$

□

Proposition 2. Given a matrix $A \in \mathbb{R}^{n\times n}$ and two vectors $\boldsymbol{x}$ and $\boldsymbol{y} \in \mathbb{R}^n$, the following relation hold:

$$\det(A + \boldsymbol{x}\boldsymbol{y}') = det(A)(1 + \boldsymbol{x}'A^{-1}\boldsymbol{y}) \quad (13)$$

□

Lemma 1.

$$\sum_{t=1}^{T} X_t X_t' > 0 \Leftrightarrow \sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t' > 0. \quad (14)$$

Lemma (1) follows from the block diagonal structure of (3).

Lemma 2. n vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$, $\boldsymbol{x}_i \in \mathbb{R}^n$, are linear independent iff

$$rank\,(\boldsymbol{x}_1\boldsymbol{x}_1' + \boldsymbol{x}_2\boldsymbol{x}_2' + \ldots + \boldsymbol{x}_n\boldsymbol{x}_n') = n$$

□

Proof (⇒) Consider

$$Z = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \ldots & \boldsymbol{x}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix}$$

From (12a) and (12b) we have that

$$n \le rank\,(Z) \le n$$

□

(⇐) If $rank\,(Z) = n$ then, by (12a), we have

$$rank\,(\begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \ldots & \boldsymbol{x}_n \end{bmatrix}) \ge n$$

which completes the proof. □

Proof of Theorem 2 (⇐) *Identifiability implies reachability (from the null state)*

Let us fix $T = n$. Thanks to Property (14) it is sufficient to study the *rank* of the matrix

$$rank\,(F(1,n)) = rank\left(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t'\right). \quad (15)$$

For (2) the $rank\,(F(1,n))$ is equal to the number of linearly independent vectors $\boldsymbol{x}_i$, $i = 1, \ldots, n$, then the *rank* expressed in (15) is equal to

$$rank\,[\boldsymbol{x}_1\ \boldsymbol{x}_2\ \ldots\ \boldsymbol{x}_n]\,, \quad (16)$$

which, by substitution, can be written as

$$rank \begin{bmatrix} \boldsymbol{u}_0' B' \\ \boldsymbol{u}_0' B' A' + \boldsymbol{u}_1' B' \\ \vdots \\ \boldsymbol{u}_0' B' A'^{n-1} + \boldsymbol{u}_1' B' A'^{n-2} + \ldots + \boldsymbol{u}_{n-1}' B' \end{bmatrix}'$$

or, equivalently,

$$rank\,(KU)$$

where U is defined in (4).

Suppose, by contradiction, that $rank\,(K) < n$ then, using relation (12a) we obtain

$$rank\,(KU) \le min(rank\,(K)\,, rank\,(U))$$

and then

$$rank\,(KU) < n$$

for all sequences $\boldsymbol{u}_0^{n-1}$, contradicting the hypothesis. □

(⇒) *Reachability implies Identifiability (from the null state)*

We have to prove that if the system is reachable then exists a time T and a sequence $\boldsymbol{u}_0^{T-1}$ such that the quantity in (15) is equal to n. Let us fix $T = n$. Let us define the column vectors of U as

$$\boldsymbol{v}_i, \quad i = 1, \ldots, n.$$

Choose a control sequence $\boldsymbol{u}_0^{n-1}$ so that vectors $\boldsymbol{v}_i$ satisfy condition (5). This can be always done by following the construction reported in Procedure 1.

Note that

$$\boldsymbol{x}_i = K\boldsymbol{v}_i, \quad i = 1, \ldots, n$$

and assume, by contradiction, that $\boldsymbol{x}_i$ are linear independent, then $\exists \alpha_1, \ldots, \alpha_n$ with $\alpha_i \ne 0$ for some i, s.t.

$$\sum_{i=1}^{n} \alpha_i \boldsymbol{x}_i = 0$$

and, by substitution,

$$\sum_{i=1}^{n} \alpha_i K \boldsymbol{v}_i = K \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i = 0.$$

Then we have a non null linear combination of vectors belonging to both $span(K')$ and $Ker(K)$ which contradicts the hypothesis. □

Proof of Theorem 4

$$\log\det\left(M_T^{-1}\right) = \log\det\left(M_{T-1} + \frac{1}{\sigma^2}\boldsymbol{x}_T\boldsymbol{x}_T'\right)^{-1} =$$
$$-\log\det\left(M_{T-1} + \frac{1}{\sigma^2}\boldsymbol{x}_T\boldsymbol{x}_T'\right).$$

Now, by proposition 2, we can rewrite (17) as

$$-\log\det\left(M_{T-1}\right) - \log\left(1 + \frac{1}{\sigma^2}\boldsymbol{x}_T' M_{T-1}^{-1}\boldsymbol{x}_T\right).$$

and iterating recursively we obtain

$$\log\det\left(M_T^{-1}\right) = -\sum_{i=0}^{T} ln(1 + \frac{1}{\sigma^2}\boldsymbol{x}_i' M_{i-1}^{-1}\boldsymbol{x}_i)$$

□