Automatica 96 (2018) 342–353

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Finite time identification in unstable linear systems[☆]

Mohamad Kazem Shirani Faradonbeh [a,*], Ambuj Tewari [b], George Michailidis [a]

[a] Informatics Institute, University of Florida, Gainesville, FL, 32611, USA,
[b] Department of Statistics, University of Michigan, Ann Arbor, MI, 48109, USA

ARTICLE INFO

Article history:
Received 4 October 2017
Received in revised form 13 February 2018
Accepted 5 June 2018

Keywords:
Linear dynamics
Finite time stabilization
Stochastic control
Adaptive control
Autoregressive process
Non-asymptotic estimation

ABSTRACT

Identification of the parameters of stable linear dynamical systems is a well-studied problem in the literature, both in the low and high-dimensional settings. However, there are hardly any results for the unstable case, especially regarding *finite time bounds*. For this setting, classical results on least-squares estimation of the dynamics parameters are not applicable and therefore new concepts and technical approaches need to be developed to address the issue. Unstable linear systems arise in key real applications in control theory, econometrics, and finance.

This study establishes finite time bounds for the identification error of the least-squares estimates for a fairly large class of heavy-tailed noise distributions, and transition matrices of such systems. The results relate the time length (samples) required for estimation to a function of the problem dimension and key characteristics of the true underlying transition matrix and the noise distribution. To establish them, appropriate concentration inequalities for random matrices and for sequences of martingale differences are leveraged.

1. Introduction

Identification of the transition matrix in linear dynamical systems has been extensively studied in the literature for the stable case (Ljung, 1999; Lütkepohl, 2005; Söderström & Stoica, 1989). Further, new work has also addressed this topic under a high-dimensional scaling, with additional assumptions on sparsity of the parameters imposed on it (Basu & Michailidis, 2015; Zorzi & Chiuso, 2017; Zorzi & Sepulchre, 2016). However, in settings where the underlying dynamics are *not stable*, this problem has *not* been adequately studied. A key issue that arises in this case is that the magnitude of the state vector explodes with high probability, exponentially over time (Lai & Wei, 1985). Nevertheless, identification of the dynamics in the non-stable case is of interest due to a number of applications that give rise to such dynamics. In addition to adaptive control (Bertsekas, 1995; Kailath, Sayed, Hassibi, & Linear estimation. Prentice Hall, 2000; Kumar & Varaiya, 2015; Söderström, 2012), these applications include a class of identification problems involving asset bubbles and high inflation episodes (Alogoskoufis & Smith, 1991; Engsted, 2006; Garcia, Perron, et al., 1991; Giacomini & White, 2006; Juselius & Mladenovic, 2002; Nielsen, 2010, 2008; Pesaran, 2010; Pesaran & Timmermann, 2002, 2005; Stock & Watson, 1996, 1998).

Most existing work on the topic provides *asymptotic* results on the convergence (Lai & Wei, 1985), as well as the limit distribution (Buchmann & Chan, 2013; Buchmann, Chan, et al., 2007) of the model parameters. Specifically, early work investigated the limit distribution of the state vector under a set of restrictive assumptions on the dynamics matrix (Anderson, 1959). Ensuing work dealt with the accuracy of identification in infinite time, for a class of structured transition matrices (Lai & Wei, 1983a). Further extensions to more general classes were established by Nielsen (2005, 2006). Finally, additional asymptotic results together with the important concept of *irregularity* of the transition matrix which leads to inconsistency, are presented in the literature (Nielsen, 2009). However, finite time (i.e. non-asymptotic) results are not currently available.

In this work, we consider a linear dynamical system $x(t) \in \mathbb{R}^p, t = 0, 1, \ldots$ that evolves according to the following Vector Autoregressive (VAR) model

$$x(t+1) = A_0 x(t) + w(t+1), \tag{1}$$

starting from an arbitrary initial state $x(0)$, which can be either deterministic or stochastic. Note that systems of longer but finite memory can also be written in the above form (Kailath et al., 2000; Söderström, 2012). We examine the general case where

☆ The work of A. T. was partially supported by a Sloan Research Fellowship. A. T. also acknowledges the support of NSF via grant DMS-1612549. The work of G. M. was supported in part by NSF grants CCF-1442493, IIS-1632730, and CNS-1422078. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Juan C Aguero under the direction of Editor Torsten Söderström.

* Corresponding author.
E-mail addresses: mfaradonbeh@ufl.edu (M.K. Shirani Faradonbeh), tewaria@umich.edu (A. Tewari), gmichail@ufl.edu (G. Michailidis).

https://doi.org/10.1016/j.automatica.2018.07.008

the system is not necessarily stable. The key contributions are: (i) establishing finite time identification bounds for the ℓ_2 error of the least-squares estimates of the transition matrix A_0, (ii) under a fairly general heavy tailed noise (disturbance) process $\{w(t)\}_{t=1}^{\infty}$. In addition, the results due to the presence of a heavy-tailed noise term are of independent interest for the stable case as well. The novel results established provide insights on how the time length required for identification scales both with the dimension of the system, as well as with the characteristics of the transition matrix and the noise process.

In order to establish results for accurate finite time identification of A_0, one needs to address the following set of technical issues. Note that as long as A_0 has eigenvalues outside of the unit circle in the complex plane, the behavior of the Gram matrix of the state vector is governed by a random matrix. However, when A_0 has eigenvalues both inside and outside of the unit circle, the smallest eigenvalue of the Gram matrix scales linearly over time, while its largest eigenvalue grows exponentially, which in turn leads to the failure of the classical approaches to establish accurate identification. These issues are addressed in Sections 3.2 and 3.3, respectively. In the proofs, we leverage selected concentration inequalities for random matrices (Tropp, 2012), as well as an anti-concentration property of martingale difference sequences (Lai & Wei, 1983b).

The problem of fast accurate identification in unstable systems has a number of interesting applications. For example, in stochastic control, this includes the canonical problems of both stabilization, as well as design of an efficient adaptive policy for linear systems. First, since the dynamics are governed by unknown transition matrices, the control action can destabilize the system. Moreover, the user first needs to have an approximation of the dynamics, to be able to design a suitable control policy. Therefore, *accurate* identification of the dynamics of the transition matrices is necessary, even if they happen to lead to instability of the underlying system. More importantly, the identification result needs to be provided within a relative *short* time period for the user to be able to design the adaptive policy accordingly. More details are discussed in Example 1.

Applications of this setting in econometrics and finance also create the need to obtain finite time theoretic results. For example, in macroeconomics, the outstanding performance of the linear models marked them as a benchmark of forecasting the market (Giacomini & White, 2006; Pesaran & Timmermann, 2005; Stock & Watson, 1998). Their applications to the analysis of inflationary episodes in a number of OECD,[1] countries (Pesaran & Timmermann, 2005) as well as US stock prices (Engsted, 2006; Lin & Michailidis, 2017) are available in the literature. The former study establishes the structural non-stationarity of the process, where the latter verifies the explosive behavior of speculative bubbles. In particular, if a technology market is capable of important innovations with uncertain outcomes, it has been argued (Pesaran, 2010) that a bubble is very likely to emerge.

Another application involving unstable dynamics deals with episodes of hyperinflation. For example, Juselius and Mladenovic (2002) consider the case of (former) Yugoslavia and use data on various economic indicators to gain insights into the dynamics of the late 1990s episode. The analysis identifies wages, price level expectations, and currency depreciation as the key factors. In follow-up work, infinite time analysis techniques were used (Nielsen, 2010), but as emphasized in the original work (Juselius & Mladenovic, 2002) "hyperinflation episodes almost by definition are *short*". Therefore, the small sample size available can easily lead to incorrect inference, while finite time guarantees are informative about the sample size needed to make precise statements about the effects of different macroeconomic factors. Another hyperinflation episode from Germany in the early 1920s is studied by Nielsen (2008).

Recently, the problem of forecasting non-stationary mixing (Kuznetsov & Mohri, 2014, 2017), and non-mixing (Kuznetsov & Mohri, 2015) time series has received attention, assuming the loss function employed is bounded. Unstable VAR models are a special, yet interesting, case of non-stationary time series. However, the problem of estimation/identification is not still addressed in the existing literature. Moreover, the results on forecasting are not applicable to the identification problem, since the least-squares loss function used in that study is not bounded. On the other hand, the obtained results on identification are applicable to forecasting.

The remainder of the paper is organized as follows. In Section 2 we provide a rigorous formulation of the problem, introduce the identification procedure, and outline examples that require accurate identification but the system cannot assumed to be stable. The contributions are discussed in Section 3, where we study different scenarios. First, we provide identification results on (non-stationary) stable linear systems in Section 3.1, followed by the explosive case (Section 3.2). Finally, we study the accurate identification of the dynamics for general systems in Section 3.3.

1.1. Notations

The following notation is used throughout this paper. For a matrix $A \in \mathbb{C}^{p\times q}$, A' denotes its transpose. When $p = q$, the smallest (respectively largest) eigenvalue of A (in magnitude) is denoted by $\lambda_{\min}(A)$ (respectively $\lambda_{\max}(A)$). For $\gamma \in \mathbb{R}, \gamma > 0, x \in \mathbb{C}^q$, define the norm $\|x\|_\gamma = \left(\sum_{i=1}^{q} |x_i|^\gamma\right)^{1/\gamma}$. For $\gamma = \infty$, define the norm $\|x\|_\infty = \max_{1\le i\le q} |x_i|$.

We also use the following notation for the operator norm of matrices. For $\beta, \gamma \in (0, \infty]$, $A \in \mathbb{C}^{p\times q}$ let,

$$|||A|||_{\gamma\to\beta} = \sup_{v\in\mathbb{C}^q\setminus\{0\}} \frac{\|Av\|_\beta}{\|v\|_\gamma}.$$

Whenever $\gamma = \beta$, we simply write $|||A|||_\beta$. To show the dimension of manifold M over the field F, we use $\dim_F(M)$. The sigma-field generated by random vectors $X_1, \ldots, X_n$ is denoted by $\sigma(X_1, \ldots, X_n)$. Finally, the symbol $\vee$ denotes the maximum of two or more quantities.

2. Problem formulation and preliminaries

The system $\{x(t)\}_{t=0}^{\infty}$ evolves according to (1), while the unknown transition matrix $A_0 \in \mathbb{R}^{p\times p}$ is not assumed to be stable, i.e. the eigenvalues of A_0 do not necessarily lie inside the unit circle. Further, $\{w(t)\}_{t=1}^{\infty}$ is the sequence of independent mean-zero noise vectors with covariance matrix C, i.e. $\mathbb{E}[w(t)] = 0$, and $\mathbb{E}\left[w(t)w(t)'\right] = C$.

Remark 1. The results established also hold if the noise vectors are martingale difference sequences. Further, the generalization to heteroscedastic noise, where the covariance matrix C is time varying, is rather straightforward.

The objective is to identify A_0, using the least-squares estimator. One observes the state vector during a finite time interval, $\{x(t)\}_{t=0}^{n}$, and defines the sum-of-squares loss function

$$\mathcal{L}_n(A) = \sum_{t=0}^{n-1} \|x(t+1) - Ax(t)\|_2^2.$$

Then, A_0 is estimated by $\hat{A}^{(n)}$, which is the minimizer of the above sum-of-squares; $\mathcal{L}_n\left(\hat{A}^{(n)}\right) = \min_{A\in\mathbb{R}^{p\times p}} \mathcal{L}_n(A)$.

[1] Organization for economic co-operation and development.