Explicit expression of the parameter bias in identification of Laguerre models from step responses*

Andrea Lecchini[†,§] **and Michel Gevers**[‡]

[†] Department of Engineering - University of Cambridge
Trumpington Street, Cambridge, CB2 1PZ, U.K.
`al394@cam.ac.uk`

[‡]CESAME - Université Catholique de Louvain
Bâtiment Euler, 4 Av. Georges Lemaître,
B-1348 Louvain-la-Neuve, Belgium
`gevers@csam.ucl.ac.be`

Revised version: 12 September 2003

Abstract

This paper delivers an analysis of the least-squares estimation of the Laguerre coefficients of a linear discrete-time system from step response data. The original contribution consists in an explicit formula for the bias error on the estimated coefficients due to the under-modelling of the system. The formula, jointly with some a-priori information on the neglected dynamics, can be used to construct bounds on this error. The results presented in this paper are illustrated with a simulation example.

*This paper presents research results of the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The research is also funded by the European Research Network on System Identification (ERNSI) funded by the European Union. The scientific responsibility rests with its authors.

§This work was carried out while the first author was a post doc researcher at the Centre for Systems Engineering and Applied Mechanics (CESAME) Université Catholique de Louvain - Belgium.

1 Introduction

Consider a linear discrete-time plant with transfer function $T(z)$. If $T(z)$ is unknown, a model of the plant can be estimated from experimental data ([5]). A broadly used class of models in identification of linear systems consists in the finite Laguerre expansion ([4, 10, 8, 12]). A model in this class takes the form:

$$M(z,\eta) = \sum_{k=1}^{n} \eta_k L_k(z,a) \qquad L_k(z,a) = \frac{K}{z-a}\left(\frac{1-a\,z}{z-a}\right)^{k-1} \qquad K = \sqrt{1-a^2} \tag{1}$$

where $L_k(z,a)$ is a Laguerre function with poles in a, and $\eta = [\eta_1,\, \eta_2, \ldots\, \eta_n]^T$ is the parameter that has to be tuned according to the data. The parameter a is a fixed parameter chosen by the user. For $a = 0$, $M(z,\eta)$ is the usual FIR model. In general, one should choose the parameter a in accordance with the available a-priori information on $T(z)$. In fact, it is well known that, as long as $T(z)$ is stable, it can be expressed as infinite expansion of Laguerre functions, i.e:

$$T(z) = \sum_{k=1}^{\infty} g_k\, L_k(z,a)\,, \tag{2}$$

the convergence rate of the expansion depending on a. For this reason, a good choice of a improves the error committed when $T(z)$ is approximated through a truncated expansion as in (1). For results on the approximation of transfer functions with finite Laguerre expansions see also ([1, 3, 7, 9, 11]).

In this paper, we will consider the particular case in which the parameter vector η of a model $M(z,\eta)$ - in the form given in (1) where a is assumed to be already chosen - is estimated from a finite sequence of step response data, and where a least-squares criterion is used. We will give an analysis of the error existing between the estimated model parameter vector $\hat{\eta} = [\hat{\eta}_1\, \ldots\, , \hat{\eta}_n]$ and the corresponding vector $\eta_0 = [g_1\, \ldots\, , g_n]$ of Laguerre coefficients of the true system (2). We will focus on the deterministic component of the error (bias) due to the under-modelling of $T(z)$. We will not consider the stochastic component arising from the possible presence of noise affecting the data. This bias error on the coefficients depends of course on the unknown transfer function $T(z)$. It will be shown that our explicit bias error expression, jointly with some prior information on the decay rate of the Laguerre coefficients of $T(z)$, can be used to derive useful bounds on this error.

It is well known that the step signal is not persistently exciting (see [5]). This means that the condition number of the identification problem gets worse as N (i.e. the number of data) increases. Since increasing the number of data is needed to eliminate the effect of the noise, our analysis is of practical use only in a situation with a low noise level. On the other hand, the results in this paper are of theoretical importance. They can be the basis for further developments in the analysis of the bias error committed when square-wave or PRBS signals are used. Indeed, these signals, which are broadly used in identification, are superpositions of steps.

The paper is organized as follows. The problem statement and the notation are introduced in Section 2. The explicit expression of the bias error on the estimated coefficients is derived in Section 3. Section 4 is devoted to a brief discussion of the particular case in which the static gain is known. The expression of the bias error is used in Section 5 to calculate a worst-case bound on this error. An illustrative simulation example is given in Section 6. The conclusions, in Section 7, end the paper. The Appendix contains all the technical proofs.

2 Problem statement

Let $M(z,\eta)$ be as in (1); in order to simplify the notation, we will not always express the dependency on a, which is considered as a given fixed parameter. Let $[y_s(1), \ldots, y_s(N)]$ be the (noise-free) response of the plant $T(z)$ - up to the instant N - to a unitary step change in the input applied at time 0 when the plant is at rest. This step signal is denoted step(t). We denote by $\hat{\eta}(N)$ the parameter vector that minimizes the least-squares cost function $J(\eta, N)$ constructed on the basis of $[y_s(1), \ldots, y_s(N)]$:

$$J(\eta, N) = \sum_{t=1}^{N} \Big(y_s(t) - M(z,\eta)\, \mathrm{step}(t) \Big)^2 . \tag{3}$$

We denote by η_0 the vector of the first n coefficients of the Laguerre expansion of $T(z)$ (see (2)):

$$\eta_0 = [g_1\, g_2\, \ldots\, g_n]^T .$$

In the following, we will derive an explicit expression for the parameter bias-error $D(N)$ defined as the difference between the estimated coefficients of the model (i.e. $\hat{\eta}(N)$) and the first n coefficients of the Laguerre expansion of the true plant (i.e. η_0):

$$D(N) = \hat{\eta}(N) - \eta_0 .$$

The unknown transfer function $T(z)$ will enter the expression of $D(N)$, but only through the "tail" of the expansion of $T(z)$ (i.e. $[g_{n+1}\, g_{n+2}\, \ldots]$). In the event that some prior information on the tail is available, it will be natural to use such information to derive explicit bounds for $D(N)$.

3 Explicit expression of the parameter bias

The parameter vector $\hat{\eta}(N)$ is given by the well-known normal equations

$$\hat{\eta}(N) = \left[\sum_{t=1}^{N} \phi(t)\phi(t)^T\right]^{-1} \sum_{t=1}^{N} \phi(t) y_s(t) \tag{4}$$

in which $\phi(t)$ is the regression vector defined as:

$$\begin{aligned} \phi(t) &= [l_1(t),\, l_2(t), \dots, l_n(t)]^T & (5) \\ l_k(t) &= L_k(z,a)\, \text{step}(t)\,. & (6) \end{aligned}$$

Let $\sigma_{hk}(N)$ denote the (h,k)-element of the matrix $\left[\sum_{t=1}^{N} \phi(t)\phi(t)^T\right]$: it is given by $\sigma_{hk}(N) = \sum_{t=1}^{N} l_h(t) l_k(t)$. For $a = 0$ the $l_k(t)'s$ are just delayed steps. In Figures 1 and 2, the $l_k(t)'s$ are illustrated for some other values of a. Notice that the time scales are different in the two plots.

Let us define the vectors $\bar{\phi}(t)$ and $\bar{g}$ as:

$$\begin{aligned} \bar{\phi}(t) &= [l_{n+1}(t), l_{n+2}(t), \dots]^T \\ \bar{g} &= [g_{n+1}, g_{n+2}, \dots]^T\,. \end{aligned}$$

The vector $\bar{g}$ is referred to as the tail of $T(z)$. We can then write

$$\begin{aligned} \hat{\eta}(N) &= \left[\sum_{t=1}^{N} \phi(t)\phi(t)^T\right]^{-1} \sum_{t=1}^{N} \phi(t)[\phi(t)^T \;\bar{\phi}(t)^T] \begin{bmatrix} \eta_0 \\ \bar{g} \end{bmatrix} \\ &= \left[\sum_{t=1}^{N} \phi(t)\phi(t)^T\right]^{-1} \left[\sum_{t=1}^{N} \phi(t)\phi(t)^T\right] \eta_0 + \left[\sum_{t=1}^{N} \phi(t)\phi(t)^T\right]^{-1} \left[\sum_{t=1}^{N} \phi(t)\bar{\phi}(t)^T\right] \bar{g} \end{aligned}$$

from which we obtain:

$$D(N) = \left[\sum_{t=1}^{N} \phi(t)\phi(t)^T\right]^{-1} \left[\sum_{t=1}^{N} \phi(t)\bar{\phi}(t)^T\right] \bar{g}\,. \tag{7}$$

Therefore, $D(N)$ depends only on the tail $\bar{g}$.

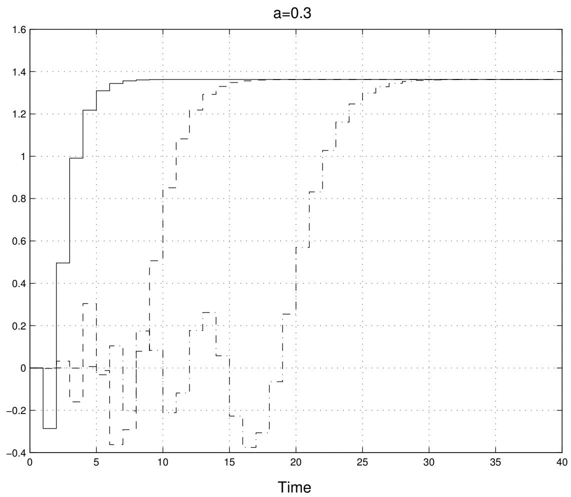

Figure 1: the $l_2(t)$ (continuous), $l_6(t)$ (dashed), $l_{12}(t)$ (dash-dotted) for $a = 0.3$.

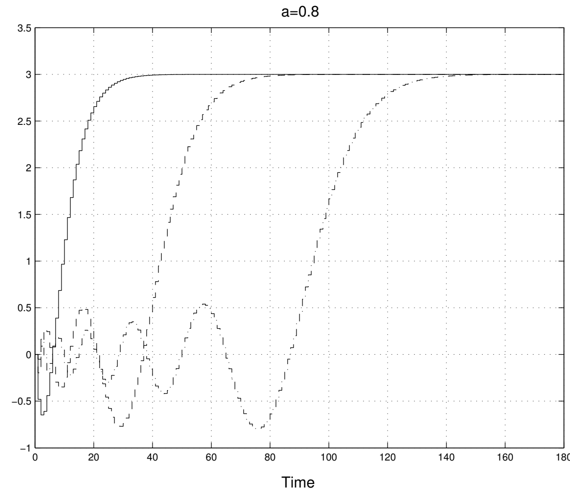

Figure 2: the $l_2(t)$ (continuous), $l_6(t)$ (dashed), $l_{12}(t)$ (dash-dotted) for $a = 0.8$.

3.1 The case of an FIR model

Let us start by considering the FIR case (i.e. $a = 0$). In this case we have

$$\sum_{t=1}^{N} l_h(t) l_k(t) \,\Big|_{a=0} = \begin{cases} N - \max(h,k) + 1 & \max(h,k) \leq N \\ 0 & \max(h,k) > N \end{cases} \tag{8}$$

from which we can easily calculate the right hand side of equation (7).
We obtain:

$$D(N)\,\Big|_{a=0} = \frac{\displaystyle\sum_{k=n+1}^{N} \left(1 - \frac{k}{N+1}\right) g_k}{1 - \dfrac{n}{N+1}} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \tag{9}$$

We observe from expressions (9) that the "shape" of the parameter bias-error is independent of the plant. The first $n-1$ coefficients are unbiased: the error is entirely confined to the last coefficient. The tail of the Laguerre expansion of $T(z)$ (i.e. $\bar{g}$) enters the error vector only through the proportionality coefficient $\sum_{k=n+1}^{N} \left(1 - \frac{k}{N+1}\right) g_k$.

3.2 The case of general Laguerre models

Let us now consider the general case (i.e. $a \neq 0$). We will show that the bias presents similar features to the case $a = 0$. On the other hand, in order to make the problem tractable, we need to introduce some approximation.

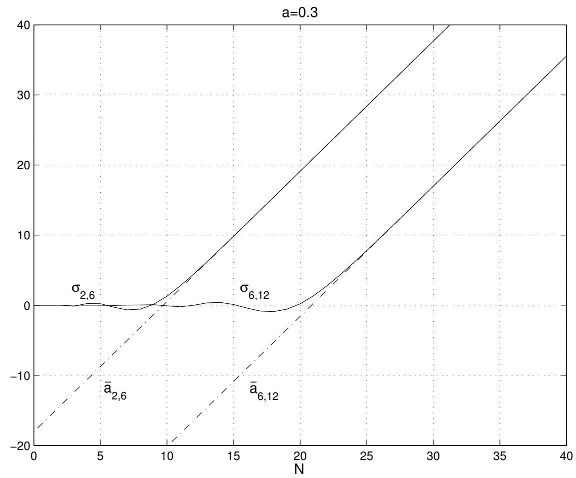

Figure 3: some $\sigma_{hk}(N)$ (continuous) compared to the corresponding $\bar{a}_{hk}(N)$ (dash-dotted) for $a = 0.3$.

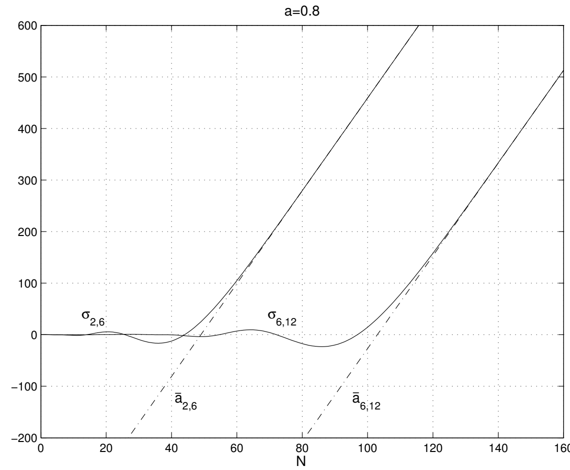

Figure 4: some $\sigma_{hk}(N)$ (continuous) compared to the corresponding $\bar{a}_{hk}(N)$ (dash-dotted) for $a = 0.8$.

To start with, we have the following result on the matrix $\left[\sum_{t=1}^{N} \phi(t)\phi(t)^T\right]$.

Proposition 1

Define the matrix $\bar{A}(N)$ with elements $\bar{a}_{hk}(N)$ $h, k = 1, \ldots, n$ as:

$$\begin{aligned} \bar{a}_{kk}(N) &= \frac{K^2}{(1-a)^2}\left(N + \frac{1}{1-a} - k\,\frac{1+a}{1-a}\right) - \frac{K^2}{(1-a)^2}\frac{a}{1-a^2} \\ \bar{a}_{hk}(N) &= \frac{K^2}{(1-a)^2}\left(N + \frac{1}{1-a} - m\,\frac{1+a}{1-a}\right) \quad m = \max(h,k) \qquad h \neq k\,. \end{aligned}$$

Then the following holds:

$$\lim_{N\to\infty}\left(\bar{A}(N) - \left[\sum_{t=1}^{N} \phi(t)\phi(t)^T\right]\right) = \mathbf{0}\,. \tag{10}$$

□

Proof: see Appendix.

The result in Proposition 1 says that the matrix $\left[\sum_{t=1}^{N} \phi(t)\phi(t)^T\right]$ is approximated by $\bar{A}(N)$ as N goes to infinity. Let us give a graphical illustration. In Figures 3 and 4, the elements $\sigma_{hk}(N)$ of the matrix $\left[\sum_{t=1}^{N} \phi(t)\phi(t)^T\right]$ - corresponding to the $l_k(t)$'s displayed in Figures 1 and 2 - are compared to the corresponding elements $\bar{a}_{hk}(N)$. It is apparent, from a direct comparison of Figures 1 and 2 with Figures 3 and 4 respectively, that each couple of elements $\sigma_{hk}(N)$ and $\bar{a}_{hk}(N)$ becomes indistinguishable as long as N is greater than the maximum settling time of the corresponding $l_h(t)$ and $l_k(t)$. In general, it is reasonable to assume that $\bar{A}(N)$ is a good approximation of $\left[\sum_{t=1}^{N} \phi(t)\phi(t)^T\right]$ as long as N is greater than the settling time of $l_n(t)$. Moreover, under this assumption, we are allowed to consider $\bar{A}(N)$ as nonnegative definite (notice that $\left[\sum_{t=1}^{N} \phi(t)\phi(t)^T\right]$ is nonnegative definite by construction, whereas $\bar{A}(N)$ is not so for small N).

Let us now construct a suitable approximation for the vector $\left[\sum_{t=1}^{N} \phi(t)\bar{\phi}(t)^T\right] \bar{g}$ which also appears in (7) and (23). It is based on the following result.

Proposition 2
For any given positive integer d define the vector $\bar{\phi}_d(t)$ as:

$$\bar{\phi}_d(t) = [l_{n+1}(t),\ \ldots\ , l_{n+d}(t)]^T \, .$$

Define the matrix $\bar{B}_d(N)$ with elements $\bar{b}_{hk}(N)$ $h = 1, \ldots, n, k = 1, \ldots, d$ as:

$$\bar{b}_{hk}(N) = \frac{K^2}{(1-a)^2} \left(N + \frac{1}{1-a} - (k+n)\,\frac{1+a}{1-a} \right) \, .$$

Then the following holds:

$$\lim_{N \to \infty} \left(\bar{B}_d(N) - \left[\sum_{t=1}^{N} \phi(t)\bar{\phi}_d(t)^T \right] \right) = \mathbf{0} \, . \tag{11}$$

□

Proof: the proof follows directly form the proof of Proposition 1.

In the above proposition, we have introduced the matrix $\left[\sum_{t=1}^{N} \phi(t)\bar{\phi}_d(t)^T\right]$ and its asymptotic approximation $\bar{B}_d(N)$. Similarly to the case of Proposition 1, we have established that $\bar{B}_d(N)$ is a good approximation of $\left[\sum_{t=1}^{N} \phi(t)\bar{\phi}_d(t)^T\right]$ as long as N is greater than the settling time of $l_{n+d}(t)$.

At this point, using the results in Propositions 1 and 2, we can construct the asymptotic approximation (as $N \to \infty$) of the bias error obtained when the system under identification is a truncated version of the system $T(z)$ (i.e. a system formed by the expansion of $T(z)$ up to the element $n+d$). Indeed, if the true system $T(z)$ were replaced by a truncation (up to element $n+d$) of the Laguerre expansion of $T(z)$, then the quantity $\left[\sum_{t=1}^{N} \phi(t)\bar{\phi}(t)^T\right]$ in the bias expression (7) would be replaced by $\left[\sum_{t=1}^{N} \phi(t)\bar{\phi}_d(t)^T\right]$, for which Proposition 2 delivers a convergent approximate. Let us denote $D_d(N)$ the bias error in the identification of the truncated system. Clearly, the vector $D_d(N)$ is given by:

$$D_d(N) = \left[\sum_{t=1}^{N} \phi(t)\phi(t)^T \right]^{-1} \left[\sum_{t=1}^{N} \phi(t)\bar{\phi}_d(t)^T \right] \bar{g}_d \tag{12}$$

where $\bar{g}_d = [g_{n+1}, \ldots, g_{n+d}]^T$. An asymptotic (in N) approximation of $D_d(N)$ is then given by the vector $\bar{D}_d(N)$ defined as:

$$\bar{D}_d(N) = \bar{A}(N)^{-1}\bar{B}_d(N)\bar{g}_d\,. \tag{13}$$

Moreover, we have:

$$D_d(N) = \bar{D}_d(N) + o\,(1/N) \quad \text{as} \quad N \to \infty\,. \tag{14}$$

This relation can easily be checked using equations (10) and (11) with some algebraic manipulations involving the definitions of $\bar{A}(N)$ and $\bar{B}_d(N)$.
An explicit expression of $\bar{D}_d(N)$ is given in the following proposition.

Proposition 3
Assume that $\bar{A}(N)$ is non-singular, then the vector $\bar{D}_d(N)$ is given by:

$$\bar{D}_d(N) = \left(\frac{\left(N + \dfrac{1}{1-a}\right) \displaystyle\sum_{k=n+1}^{n+d} g_k - \dfrac{1+a}{1-a} \displaystyle\sum_{k=n+1}^{n+d} k\, g_k}{\dfrac{1-a^{2n}}{1+a} N + \dfrac{1}{1+a} - n\, \dfrac{1-a^{2n}}{1-a}} \right) \bar{x}(a)\,, \tag{15}$$

where

$$\bar{x}(a) = \begin{bmatrix} (-a)^{n-1} \\ \vdots \\ (-a)^1 \\ (-a)^0 \end{bmatrix} + (-a)^n \begin{bmatrix} (-a)^0 \\ (-a)^1 \\ \vdots \\ (-a)^{n-1} \end{bmatrix}. \tag{16}$$

□

Proof: see Appendix.

Notice that, even if $\bar{A}(N)$ becomes singular as $N \to \infty$, the expression $\bar{D}_d(N)$ remains well defined and finite.

We are now in the position to find an asymptotic approximation for the vector $D(N)$, defined in (7), by letting d go to infinity in eq. (13). However, in order to assure that (13) is still an asymptotic (in N) approximation of (12) as $d \to \infty$, we need the uniform convergence of $\lim_{N\to\infty} \left[\sum_{t=1}^{N} \phi(t)\bar{\phi}_d(t)^T\right] \bar{g}_d$. In the following Proposition we prove uniform convergence under the assumption that $\sum_{k=n+1}^{d} k|g_k|$ converges as $d \to \infty$. This is a standard assumption; notice that all stable rational transfer functions have this property - see e.g. [5].

Proposition 4
Assume that $\sum_{k=n+1}^{\infty} k|g_k|$ is finite, then the following holds:

$$\lim_{N\to\infty} \left(\bar{B}_d(N)\, \bar{g}_d - \left[\sum_{t=1}^{N} \phi(t) \bar{\phi}_d(t)^T \right] \bar{g}_d \right) = \mathbf{0} \quad \text{uniformly in } d\,. \tag{17}$$

□

Proof: see Appendix.

Now, let us define

$$\bar{D}(N) = \lim_{d\to\infty} \bar{D}_d(N)\,. \tag{18}$$

Then, from the result in Proposition 4, we finally obtain:

$$\lim_{N\to\infty} \left(D(N) - \bar{D}(N) \right) = \mathbf{0}\,. \tag{19}$$

The asymptotic approximation, introduced in (19), can be considered valid as long as N is larger than the settling time of the step response $y_s(t)$. This can be easily deduced from the proof of Proposition 4 considering that $y_s(t) = \sum_{k=1}^{\infty} g_k l_k(t)$. Moreover, if we substitute the vector $\bar{B}_d(N)\, \bar{g}_d$ in (13) by its limit expression (see eq. (46) in the proof of Proposition 3), then - similarly to (14) - we obtain:

$$D(N) = \bar{D}(N) + o\,(1/N) \quad \text{as} \quad N \to \infty\,. \tag{20}$$

Collecting all previous results, we have now proved the main result of this paper.

Proposition 5
Assume that $\sum_{k=n+1}^{\infty} k|g_k|$ is finite and define $\bar{D}(N)$ as:

$$\bar{D}(N) = \left(\frac{\left(N + \dfrac{1}{1-a}\right) \displaystyle\sum_{k=n+1}^{\infty} g_k - \dfrac{1+a}{1-a} \displaystyle\sum_{k=n+1}^{\infty} k\, g_k}{\dfrac{1-a^{2n}}{1+a} N + \dfrac{1}{1+a} - n\, \dfrac{1-a^{2n}}{1-a}} \right) \bar{x}(a)\,, \tag{21}$$

where $\bar{x}(a)$ is defined in (16). Let $D(N)$ be the parameter bias error of the unconstrained least-squares identification problem defined in (7). Then, the following holds:

$$D(N) = \bar{D}(N) + o\,(1/N) \quad \text{as} \quad N \to \infty\,. \tag{22}$$

□

Similarly to what we obtained in the FIR case, the "shape" of the (asymptotic) parameter bias vectors (21) and (25) is independent of the plant. In the present case of a general Laguerre expansion, we observe from the definition of $\bar{x}(a)$ that the bias error turns out to be smaller for the first coefficients and it then increases as k increases to n. The plant $T(z)$ enters the vector only through some proportionality factor.

4 A particular case: known static gain

In this section we consider the particular case in which the static gain of $T(z)$ is known. For example it could be known from specific physical reasons or it could be the case that $T(z)$ is a closed-loop transfer function of a feedback loop that includes a controller with a fixed integral action in such a way that the static gain is unitary. The user can include this information and constrain the identified system to have the same static gain of $T(z)$. Interestingly enough, it will be shown that, by including this constraint in the identification, one obtains the same asymptotic bias as in the unconstrained case but with faster convergence.

We denote $\hat{\eta}_c(N)$ the minimizer of $J(\eta, N)$ subject to the constraint that the model $M(\hat{\eta}_c(N))$ must have the same static gain as the plant: i.e. $M(1, \hat{\eta}_c(N)) = T(1)$. The parameter bias error in the constrained case is denoted:

$$D_c(N) = \hat{\eta}_c(N) - \eta_0 \,.$$

Without loss of generality we assume that the static gain of $T(z)$ is 1. Since each element $L_k(z, a)$ of the Laguerre expansion has gain $\frac{K}{1-a}$, the vector $\hat{\eta}_c(N)$ is defined as:

$$\hat{\eta}_c(N) = \arg\min_{\eta} \ J(\eta, N) \quad \text{subject to} \quad V^T \eta = \frac{1-a}{K} \tag{23}$$

in which $V = [1 \quad 1 \ldots 1]^T$. This is a quadratic problem with a linear constraint therefore the solution can be easily obtained.

In the FIR case, the exact explicit solution can be calculated as usual by using the property (8). We obtain the following expression for $D_c(N)$:

$$D_c(N) \,\big|_{a=0} = \left(\sum_{k=n+1}^{\infty} g_k \right) \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} . \tag{24}$$

Notice that it does not depend on N.

In the general case, an asymptotic expression for $D_c(N)$ can also be obtained. The result is stated in the proposition below. The proof of the proposition is omitted since it is similar to the proof of Proposition 5.

Proposition 6

Assume that $\sum_{k=n+1}^{\infty} k|g_k|$ is finite and define $\bar{D}_c$ as:

$$\bar{D}_c = \left(\frac{1+a}{1-a^{2n}} \sum_{k=n+1}^{\infty} g_k \right) \bar{x}(a) \,, \tag{25}$$

where $\bar{x}(a)$ is defined in (16). Let $D_c(N)$ be the parameter bias error of the constrained least-squares identification problem defined in (23). Then the following holds:

$$D_c(N) = \bar{D}_c + o\,(1/N) \quad \text{as} \quad N \to \infty\,. \tag{26}$$

□

The following comments are in order.
The bias in the constrained case is the limit (as $N \to \infty$) of the bias in the unconstrained case. This is true both for the exact FIR case (compare (9) and (24)) and for the general case (compare (21) and (25)). Obviously, we have that:

$$\bar{D}(N) = \bar{D}_c + o(1) \quad \text{as} \quad N \to \infty\,. \tag{27}$$

Moreover, if we substitute (27) in (22), we obtain:

$$D(N) = \bar{D}_c + o(1) \quad \text{as} \quad N \to \infty \tag{28}$$

Therefore, by comparing (26) and (28), we conclude that the bias, in the constrained case, converges faster than the bias in the unconstrained case.

5 Explicit bounds on the parameter bias

If prior information on the tail of $T(z)$ is available, the results contained in the previous section can be used to derive bounds on the bias of the estimated coefficients.
A typical a-priori assumption, for finite dimensional linear systems, is that there exist λ and ρ such that:

$$|g_k| < \lambda\,\rho^k \qquad k \geq n+1\,. \tag{29}$$

Under this assumption the worst-case bias-error can be calculated.

In particular, on the basis of equations (21) and (25), we have the following results.

Proposition 7
Let $\bar{k}$ be:

$$\bar{k} = \left\lfloor \frac{1-a}{1+a}\,N + \frac{1}{1-a} \right\rfloor\,.$$

where $\lfloor x \rfloor$ denotes the greatest integer smaller than x. Then, for the unconstrained identification, the worst-case (asymptotic) bias over all possible $\{g_k, k \geq n+1\}$ satisfying (29) is given by $\bar{D}(N) = \pm\bar{D}^{wc}(N)$ where:

$$\bar{D}^{wc}(N) = \frac{\left(N + \dfrac{1}{1-a}\right) c_1 - \left(\dfrac{1+a}{1-a}\right) c_2}{\dfrac{1-a^{2n}}{1+a}\,N + \dfrac{1}{1+a} - n\,\dfrac{1-a^{2n}}{1-a}}\,\bar{x}(a) \tag{30}$$

in which c_1 and c_2 are given by

$$c_1 = \begin{cases} \lambda \dfrac{\rho^{n+1}}{1-\rho} - 2\lambda \dfrac{\rho^{\bar{k}+1}}{1-\rho} & \text{if} \quad n < \bar{k} \\[2ex] -\lambda \dfrac{\rho^{n+1}}{1-\rho} & \text{if} \quad n \geq \bar{k} \end{cases}$$

$$c_2 = \begin{cases} \lambda \dfrac{(n+1)\rho^{n+1} - n\rho^{n+2}}{(1-\rho)^2} - 2\lambda \dfrac{(\bar{k}+1)\,\rho^{\bar{k}+1} - \bar{k}\rho^{\bar{k}+2}}{(1-\rho)^2} & \text{if} \quad n < \bar{k} \\[2ex] -\lambda \dfrac{(n+1)\rho^{n+1} - n\rho^{n+2}}{(1-\rho)^2} & \text{if} \quad n \geq \bar{k} \end{cases}$$

and $\bar{x}(a)$ is defined in (16). □

Proof: see Appendix.

Proposition 8

The worst-case (asymptotic) bias over all possible $\{g_k, k \geq n+1\}$ satisfying (29), for the case of constrained identification, is given by $\bar{D}_c = \pm \bar{D}_c^{wc}$ where:

$$\bar{D}_c^{wc} = \lambda \frac{1+a}{1-a^{2n}} \frac{\rho^{n+1}}{1-\rho} \bar{x}(a) \tag{31}$$

and $\bar{x}(a)$ is defined in (16). □

Proof: the result is obtained by substituting $\lambda\,\rho^k$ to g_k in (25).

The worst-case bounds are then obtained by adding and subtracting the worst-case bias to the estimated parameter vector $\hat{\eta}(N)$ (or $\hat{\eta}_c(N)$).

It is well-known that worst-case bounds can be very conservative. This is not a weakness of the analysis contained in this paper but a feature of the worst-case approach. The expressions (21) and (25) can be used to derive bounds under other approaches to uncertainty estimation (see e.g. [2]). In addition, if the a-priori bound (29) is known to hold for all $k \geq 1$, then such information can also be used to constrain the parameter estimation problem: see e.g. [6] for a discussion of the role of prior knowledge versus data in system identification.

6 Simulation example

In this section we illustrate the results presented in this paper. Let $T(z)$ be:

$$T(z) = \frac{0.4\,(z-0.5)^2}{(z-0.6)(z^2-1.4z+0.65)}\,.$$

The coefficients of its Laguerre expansion with $a = 0.7$ are displayed in Figure 5. The step response is in Figure 6. The model $M(z;\eta)$ is chosen to be:

$$M(\eta) = \sum_{k=1}^{n=6} \eta_k L_k(a)\,.$$

The coefficients of $M(\eta)$ are estimated using noise-free step response data.
In Figure 7 the elements of the bias error $D(N)$ as a function of N are compared with the elements of the vector $\bar{D}(N)$ defined in (21). In Figure 8 the elements of the bias error for the constrained identification $D_c(N)$ are compared with the elements of the vector $\bar{D}_c$ defined in (25). Notice, by comparing the two figures, the faster convergence of the bias in the constrained case.

In the following, we consider the unconstrained identification with $N = 80$.
The parameter $\hat{\eta}(80)$ is given by:

$$\hat{\eta}(80) = [0.611098 \quad -0.123120 \quad -0.210881 \quad 0.219969 \quad -0.082229 \quad 0.004293].$$

As for the difference between the true and estimated parameters, we have:

$$\begin{array}{rcl} \eta_0 & = & [\;\; 0.614731 \quad -0.134486 \quad -0.190322 \quad 0.187573 \quad -0.033830 \quad -0.066331] \\ D(80) & = & [-0.003633 \quad\;\; 0.011366 \quad -0.020559 \quad 0.032396 \quad -0.048399 \quad\;\; 0.070624]\,. \end{array}$$

The true and estimated parameters are compared in Figures 9 and 10. The difference between the obtained $D(80)$ and the theoretical $\bar{D}(80)$ - given by (21) - is of order 10^{-10}.
For completeness, we illustrate also the construction of the worst-case bounds. Let the a-priori bounds on the tail be (see Figure 11):

$$|g_k| < 0.35 \cdot 0.8^k \qquad k \geq 7\,. \tag{32}$$

Then, the worst-case bounds obtained from (30) are displayed in Figure 12. It can be seen that such bounds are quite conservative with respect to the actual error. Nevertheless, they represent exactly the worst-case bias over all possible $\{g_k, k \geq 7\}$ satisfying (32).

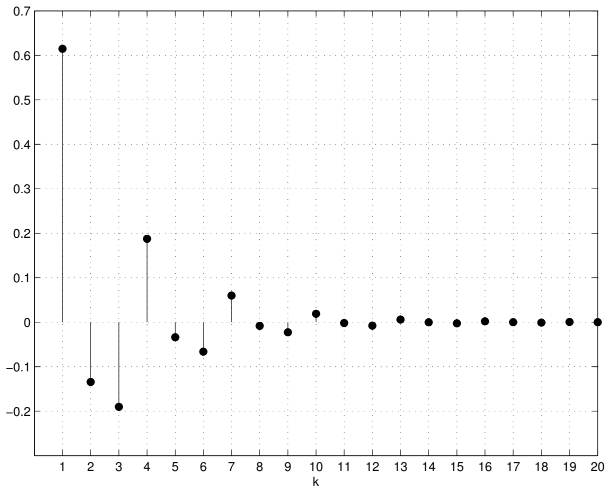

Figure 5: the coefficients of the Laguerre expansion of $T(z)$.

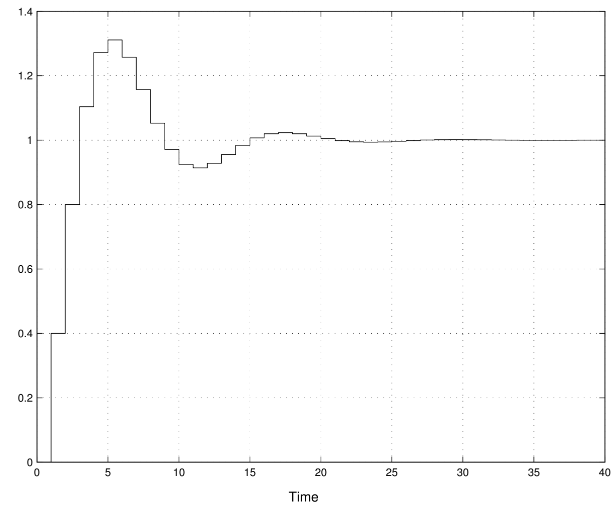

Figure 6: the step response of $T(z)$.

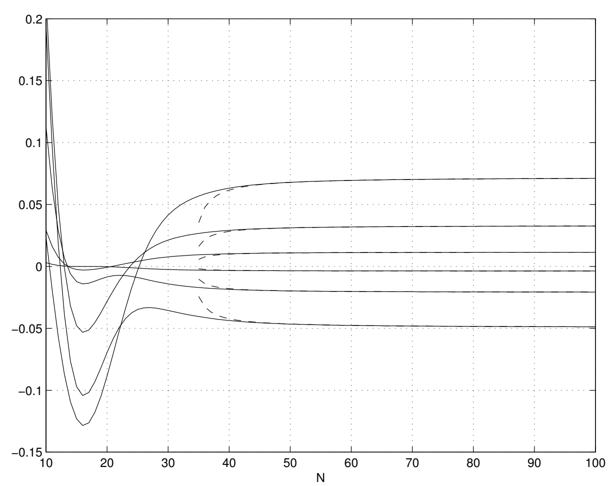

Figure 7: evolution with N (number of data) of the 6 elements of $D(N)$ (continuous) and the 6 elements of $\bar{D}(N)$ (dashed).

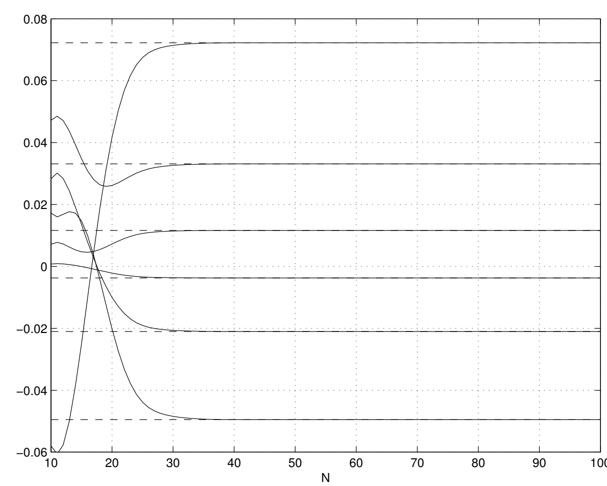

Figure 8: evolution with N (number of data) of the 6 elements of $D_c(N)$ (continuous) and the 6 elements of $\bar{D}_c$ (dashed).

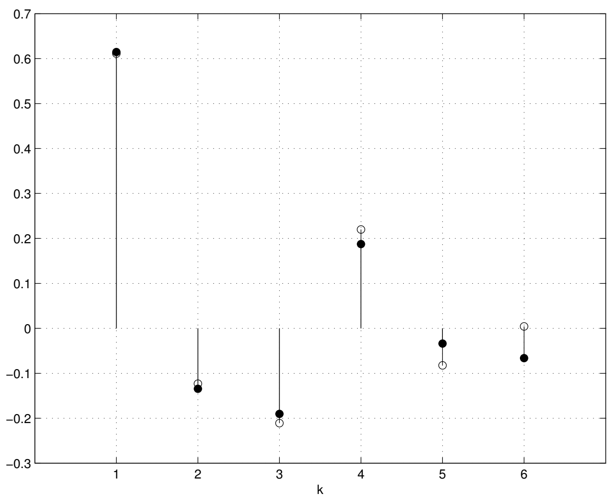

Figure 9: the true coefficients η_0 (•) and the estimated coefficients $\hat{\eta}(80)$ (o).

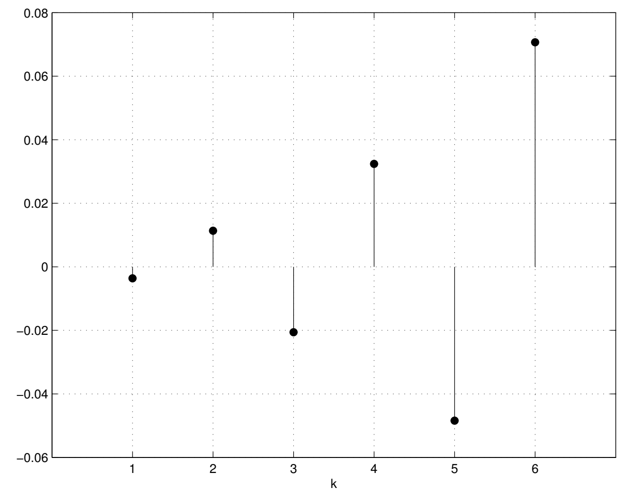

Figure 10: the elements of the parameter bias-error $D(80)$.

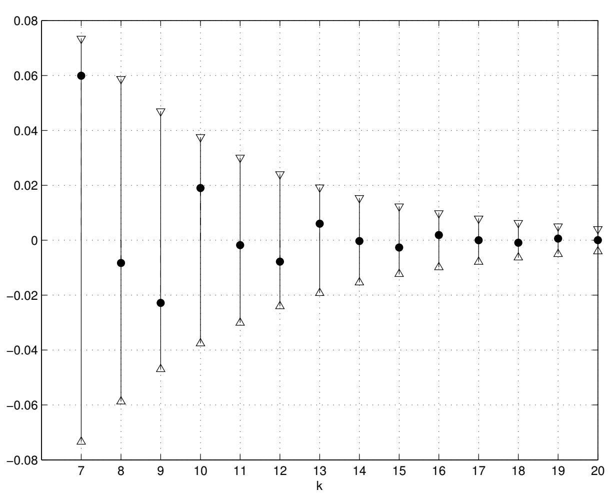

Figure 11: the coefficients of the tail of the Laguerre expansion of $T(z)$ (•) and the a-priori bounds.

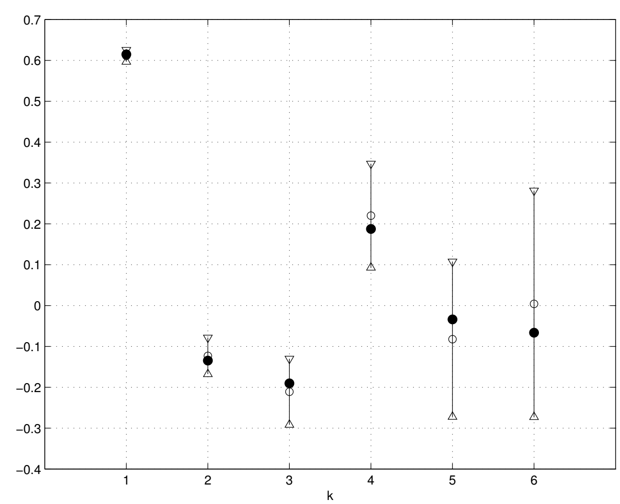

Figure 12: the true coefficients η_0 (•), the estimated coefficients $\hat{\eta}(80)$ (o) and the worst-case bounds.

7 Conclusions

In this paper, we have derived explicit expressions for the parameter bias in the identification of Laguerre coefficients from step response data. It has been shown that these expressions can be used to calculate bounds for this error. The object of further work will be the extension of the results of this paper to square-wave signals.

References

[1] Y. Fu and G.A. Dumont. An Optimum Time Scale for Discrete Laguerre Network. *IEEE Trans. Automatic Control*, 38(6):934–938, 1993.

[2] G.C. Goodwin, M. Gevers, and B. Ninness. Quantifying the Error in Estimated Transfer Functions with Application to Model Order Selection. *IEEE Trans. Automatic Control*, 37(7):913–927, 1992.

[3] P.S.C. Heuberger, P.M.J. Van den Hof, and O. K. Bosgra. A Generalized Orthonormal Basis for Linear Dynamical System. *IEEE Trans. Automatic Control*, 40(3):451–465, 1995.

[4] R. E. King and P. N. Paraskevopoulos. Parametric identification of SISO systems. *Inr. J. Control*, 30(6):1023–1029, 1979.

[5] L. Ljung. *System Identification: theory for the user*. Prentice Hall, 1999.

[6] L. Ljung, B. Whalberg, and H. Hjalmarsson. Model quality: the roles of prior knowledge and data information. In *30th IEEE Conf. on Decision and Control*, pages 273–278, Brighton, 1991.

[7] P. M. Makila. Approximation of Stable Systems by Laguerre Filters. *Automatica*, 26(2):333–345, 1990.

[8] B. Ninness, H. Hjalmarsson, and F. Gustafsson. The Fundamental Role of General Orthonormal Bases in System Identification. *IEEE Trans. Automatic Control*, 44(7):1384–1406, 1999.

[9] P.M.J. Van den Hof, P.S.C. Heuberger, and J. Bokor. System Identification with Generalized Orthonormal Basis Functions. *Automatica*, 31(12):1821–1834, 1995.

[10] B. Wahlberg. System Identification using Laguerre models. *IEEE Trans. Automatic Control*, 36(5):551–562, 1991.

[11] B. Wahlberg and P. M. Makila. On Approximation of Stable Linear Dynamical Systems using Laguerre and Kautz Functions. *Automatica*, 32(5):693–708, 1996.

[12] C. Zervos, P. R. Belanger, and Dumont G. A. On PID Controller Tuning using Orthonormal Series Identification. *Automatica*, 24(2):165–175, 1988.

A Proofs

The following Lemma will be used in the proof of Proposition 1.

Lemma 1

Consider $l_k(t)$ defined in (6) and define $Q(z,a)$ as

$$Q(z,a) = \frac{1-az}{z-a}\,,$$

then the following expression for $l_k(t)$ holds:

$$l_k(t) = l_1(t) - \sum_{h=1}^{k-1} q_h(t) \tag{33}$$

$$q_h(t) = Q(z,a)^{h-1}\left(K\,\frac{1-a^2}{1-a}\,\frac{z}{(z-a)^2}\,\delta(t)\right)\,. \tag{34}$$

□

Proof

From the definition of $L_k(z,a)$ we have that $l_k(t)$ satisfies the following recursive relation:

$$l_1(t) = \frac{K}{1-a}\left(1-a^t\right) \tag{35}$$

$$l_k(t) = Q(z,a)\,l_{k-1}(t)\,. \tag{36}$$

The impulse response of $Q(z,a)$ is

$$\begin{aligned} h_Q(0) &= -a \\ h_Q(t) &= a^{t-1} - a^{t+1} \qquad t \geq 1 \end{aligned}$$

from which we have:

$$\begin{aligned} l_2(t) &= \sum_{m=0}^{t} h_Q(t-m)\,l_1(m) \\ &= \sum_{m=0}^{t-1} \frac{K}{1-a}\left(a^{t-m-1} - a^{t-m+1}\right)(1-a^m) - K\,\frac{a}{1-a}\left(1-a^t\right) \\ &= K\,\frac{1-a^2}{1-a}\left(\frac{1-a^t}{1-a} - ta^{t-1}\right) - K\,\frac{a}{1-a}\left(1-a^t\right) \\ &= \frac{K}{1-a}\left(1-a^t\right) - K\,\frac{1-a^2}{1-a}\,t\,a^{t-1}. \end{aligned}$$

Since $\frac{z}{(z-a)^2}\,\delta(t) = t\,a^{t-1}$, the statement of the lemma is easily obtained by applying the recursive formula (35-36) to the above expression of $l_2(t)$. □

Proof of Proposition 1

Using Lemma 1 we can write:

$$\sum_{t=1}^{N} l_h(t) l_k(t) = \sum_{t=1}^{N}\left(l_1(t)^2 - l_1(t)\sum_{i=1}^{h-1} q_i(t) - l_1(t)\sum_{i=1}^{k-1} q_i(t) + \left(\sum_{i=1}^{h-1} q_i(t)\right)\left(\sum_{i=1}^{k-1} q_i(t)\right)\right). \quad (37)$$

In the following we will find the limit expression for each term in (37).

As for the first term, we have:

$$\sum_{t=1}^{N} l_1(t)^2 = \frac{K^2}{(1-a)^2}\sum_{t=1}^{N}\left(1-a^t\right)^2 = \frac{K^2}{(1-a)^2}\left(N - 2a\,\frac{1-a^N}{1-a} + a^2\,\frac{1-a^{2N}}{1-a^2}\right),$$

from which we obtain:

$$\lim_{N\to\infty}\left(\sum_{t=1}^{N} l_1(t)^2 - \frac{K^2}{(1-a)^2}\,N\right) = \frac{K^2}{(1-a)^2}\left(\frac{a^2}{1-a^2} - \frac{2a}{1-a}\right). \quad (38)$$

As for the remaining terms, we will make use of the following equations:

$$l_1(t) = \frac{K}{1-a} - \frac{z}{1-a}\,L_1(z,a)\,\delta(t) \quad (39)$$

$$q_i(t) = \frac{a\,z}{1-a}\,L_i(z,a)\,\delta(t) + \frac{z}{1-a}\,L_{i+1}(z,a)\,\delta(t)\,. \quad (40)$$

By using (39)-(40) and the orthogonality of the Laguerrre basis, after some simple calculations we obtain

$$\lim_{N\to\infty}\sum_{t=1}^{N} l_1(t) q_i(t) = \begin{cases} \dfrac{K^2}{(1-a)^2}\dfrac{1+a}{1-a} - \dfrac{K^2}{(1-a)^2}\dfrac{a}{1-a^2} & i=1 \\[2ex] \dfrac{K^2}{(1-a)^2}\dfrac{1+a}{1-a} & i>1 \end{cases}$$

from which we can derive the limit of the second (and the third) term in (37):

$$\lim_{N\to\infty}\sum_{t=1}^{N}\left(l_1(t)\sum_{i=1}^{h-1} q_i(t)\right) = (h-1)\,\frac{K^2}{(1-a)^2}\frac{1+a}{1-a} - \frac{K^2}{(1-a)^2}\frac{a}{1-a^2}\,. \quad (41)$$

In a similar way, we can calculate

$$\lim_{N\to\infty}\sum_{t=1}^{N} q_i(t) q_{i+d}(t) = \begin{cases} \dfrac{K^2}{(1-a)^2}\dfrac{1+a^2}{1-a^2} & d=0 \\[2ex] \dfrac{K^2}{(1-a)^2}\dfrac{a}{1-a^2} & d=1 \\[2ex] 0 & d>1 \end{cases}$$

from which we obtain the limit of the last term in (37):

$$\lim_{N\to\infty}\sum_{t=1}^{N}\left(\sum_{i=1}^{h-1}q_i(t)\sum_{i=1}^{k-1}q_i(t)\right)=\begin{cases}\dfrac{K^2}{(1-a)^2}\left((h-1)\,\dfrac{1+a^2}{1-a^2}+(2h-4)\,\dfrac{a}{1-a^2}\right) & h=k\\[2ex] \dfrac{K^2}{(1-a)^2}\left((l-1)\,\dfrac{1+a^2}{1-a^2}+(2l-3)\,\dfrac{a}{1-a^2}\right)\quad l=\min(h,k) & h\neq k\end{cases}\tag{42}$$

By summing up the equations (38) (41) (42) according to equation (37), the statement of the proposition is finally achieved. □

Proof of Proposition 3

The proof is obtained by simple verification of eq. (13). We have to show that:

$$\bar{A}(N)\bar{D}_d(N)=\bar{B}_d(N)\bar{g}_d$$

when $\bar{D}_d(N)$ is given by (15). In the following we sketch the calculations.
Let us denote $\bar{y}$ the vector $\bar{A}(N)\bar{x}(a)$. Moreover, define:

$$\alpha=N+\frac{1}{1-a}\,,\quad \beta=\frac{1+a}{1-a}\,,\quad \gamma=\frac{a}{1-a^2}\,.$$

Then, the element $\bar{y}_i$ is given by:

$$\begin{aligned}\bar{y}_i &= \frac{K^2}{(1-a)^2}\left(\alpha\sum_{k=0}^{n-1}(-a)^k-i\beta\sum_{k=n-i}^{n-1}(-a)^k-\beta\sum_{k=0}^{n-i-1}(n-k)(-a)^k-\gamma\,(-a)^{n-i}\right)\\ &+ (-a)^n\frac{K^2}{(1-a)^2}\left(\alpha\sum_{k=0}^{n-1}(-a)^k-i\beta\sum_{k=0}^{i-1}(-a)^k-\beta\sum_{k=i}^{n-1}(k+1)(-a)^k-\gamma\,(-a)^{i-1}\right).\end{aligned}$$

The sums in the above expression can be made explicit by using:

$$\sum_{k=0}^{n-1}(-a)^k = \frac{1-(-a)^n}{1+a}\tag{43}$$

$$\begin{aligned}\sum_{k=0}^{n-1}k(-a)^k &= \sum_{k=0}^{n-1}(k+1)(-a)^k-\sum_{k=0}^{n-1}(-a)^k \qquad (44)\\ &= \frac{d}{da}\left(\sum_{k=1}^{n}(-a)^k\right)-\frac{1-(-a)^n}{1+a}\\ &= -\frac{(n+1)(-a)^n}{1+a}+\frac{1-(-a)^{n+1}}{(1+a)^2}-\frac{1-(-a)^n}{1+a}\,.\end{aligned}$$

After some calculations one obtains that $\bar{y}_i$ is independent of i and is given by:

$$\bar{y}_i=\frac{K^2}{(1-a)^2}\left(\frac{1-a^{2n}}{1+a}\,N+\frac{1}{1+a}-n\,\frac{1-a^{2n}}{1-a}\right).\tag{45}$$

Since the vector $\bar{B}_d(N)\,\bar{g}_d$ results to be

$$\bar{B}_d(N)\,\bar{g}_d = \left(\left(N + \frac{1}{1-a} \right) \sum_{k=n+1}^{n+d} g_k - \frac{1+a}{1-a} \sum_{k=n+1}^{n+d} k\, g_k \right) \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \tag{46}$$

one can be easily check that the equation (13) is satisfied. □

Proof of Proposition 4

Let $F_{j,d}(N)$ denote the j-th element of the vector $\left(\left[\sum_{t=1}^{N} \phi(t)\bar{\phi}_d(t)^T\right] \bar{g}_d - \bar{B}_d(N)\bar{g}_d\right)$:

$$F_{j,d}(N) = \sum_{k=1}^{d} \left[\left(\sum_{t=1}^{N} l_{n+k}(t) l_j(t) \right) - \frac{K^2}{(1-a)^2} \left(N + \frac{1}{1-a} - (k+n)\frac{1+a}{1-a} \right) \right] g_{n+k}\,.$$

We have to show that:

$$\forall \epsilon > 0 \quad \exists\, \bar{N} \;:\; |F_{j,d}(N)| < \epsilon \quad \forall N > \bar{N} \quad \forall d\,.$$

Let us assume $1 < j < n$ (the cases $j = 1$ and $j = n$ are similar), by using Lemma 1 and letting d go to infinity we can write:

$$\begin{aligned} |F_{j,d}(N)| \;\leq\; & \left| \left(\sum_{t=1}^{N} l_1(t) l_j(t) \right) - \frac{K^2}{(1-a)^2} \left(N + \frac{1}{1-a} - j\frac{1+a}{1-a} \right) \right| \sum_{k=1}^{\infty} |g_{n+k}| \\ + \; & \left| \left(\sum_{t=1}^{N} \left(\sum_{i=1}^{n-1} q_i(t) \right) l_j(t) \right) - \frac{K^2}{(1-a)^2}\frac{1+a}{1-a}(n-j) \right| \sum_{k=1}^{\infty} |g_{n+k}| \\ + \; & \sum_{k=1}^{\infty} \sum_{i=1}^{k} \left| \left(\sum_{t=1}^{N} q_{n-1+i}(t) l_j(t) \right) - \frac{K^2}{(1-a)^2}\frac{1+a}{1-a} \right| |g_{n+k}|\,. \end{aligned}$$

The three terms in the right-hand side of the above inequality are independent of d. From the results contained in the proof of Proposition 1 we have that the first two terms tend to zero as $N \to \infty$. Let us consider the third term (denoted by $\bar{F}_j^3(N)$):

$$\bar{F}_j^3(N) = \sum_{k=1}^{\infty} \sum_{i=1}^{k} \left| \left(\sum_{t=1}^{N} q_{n-1+i}(t) l_j(t) \right) - \frac{K^2}{(1-a)^2}\frac{1+a}{1-a} \right| |g_{n+k}|\,.$$

Each term $\left| \left(\sum_{t=1}^{N} q_{n-1+i}(t) l_j(t) \right) - \frac{K^2}{(1-a)^2}\frac{1+a}{1-a} \right|$ converges to zero as $N \to \infty$. Moreover, since $q_i(t)$ is bounded uniformly in i:

$$\exists\, M \;:\; \left| \left(\sum_{t=1}^{N} q_{n-1+i}(t) l_j(t) \right) - \frac{K^2}{(1-a)^2}\frac{1+a}{1-a} \right| < M \quad \forall N \quad \forall i\,.$$

Therefore, we can write:

$$\begin{aligned}\bar{F}_j^3(N) &\leq \sum_{k=1}^{\bar{d}} \sum_{i=1}^{k} \left| \left(\sum_{t=1}^{N} q_{n-1+i}(t) l_j(t) \right) - \frac{K^2}{(1-a)^2} \frac{1+a}{1-a} \right| |g_{n+k}| \\ &+ |M| \sum_{k=\bar{d}+1}^{\infty} k |g_{n+k}| \end{aligned}$$

and, with a suitable choice of $\bar{d}$, the right-hand side of the above inequality can be made arbitrarily small by choosing N sufficiently large. □

Proof of Proposition 7

From equation (21) we have that:

$$\bar{D}(N) = \frac{\sum_{k=n+1}^{\infty} \left(N + \frac{1}{1-a} - \frac{1+a}{1-a} k \right) g_k}{\frac{1-a^{2n}}{1+a} N + \frac{1}{1+a} - n \frac{1-a^{2n}}{1-a}} .$$

The coefficients of the g_k's in the sum are positive for $k \leq \bar{k}$ and negative for $k > \bar{k}$. Therefore, we have that, under the assumption (29), the sequence that maximize the sum is given by:

$$\begin{cases} g_k^{wc} = \lambda \rho^k & k \leq \bar{k} \\ g_k^{wc} = -\lambda \rho^k & k > \bar{k} \end{cases} .$$

The worst case bias error is then given by:

$$\bar{D}^{wc}(N) = \frac{\sum_{k=n+1}^{\bar{k}} \left(N + \frac{1}{1-a} - \frac{1+a}{1-a} k \right) \lambda \rho^k - \sum_{k=\bar{k}+1}^{\infty} \left(N + \frac{1}{1-a} - \frac{1+a}{1-a} k \right) \lambda \rho^k}{\frac{1-a^{2n}}{1+a} N + \frac{1}{1+a} - n \frac{1-a^{2n}}{1-a}} .$$

The calculation of the sums leads to the statement of the proposition. □