

The Least Squares Algorithm, Parametric System Identification and Bounded Noise*

HÜSEYİN AKÇAY† and PRAMOD P. KHARGONEKAR‡

Key Words—Error analysis; identification; least squares estimation.

Abstract—The least squares parametric system identification algorithm is analyzed assuming that the noise is a bounded signal. A bound on the worst-case parameter estimation error is derived. This bound shows that the worst-case parameter estimation error decreases to zero as the bound on the noise is decreased to zero.

1. Introduction

THE LEAST SQUARES ALGORITHM, due to Gauss, is one of the most widely used algorithms in science. It has been extensively studied and used for parametric system identification—see, for example, the book by Ljung (1987). It is very well known that the least squares algorithm enjoys certain optimality properties under suitable stochastic assumptions about exogenous noise. In contrast, some recent papers have taken a worst-case deterministic approach to identification. See the recent paper by Khargonekar (1993) for a discussion of this general area and an extensive list of references. In particular, our work is most closely related to the work on time-domain worst-case identification problems (Chen *et al.*, 1992; Dahleh *et al.*, 1992; Kacewicz and Milanese, 1992; Mäkilä, 1991; Poolla and Tikku, 1992; Tse *et al.*, 1991; and the references cited in these papers).

Our work has grown out of a need to make connections between the classical identification theory and the more recent work in the area of robust identification. Towards this goal, in this paper we have investigated the performance of the least squares algorithm in the presence of worst-case bounded noise. In the result of this paper, we derive a bound on the worst-case parameter estimation error using the least squares algorithm in the presence of arbitrary bounded noise. This error bound shows that if the input is chosen to be a pseudorandom binary sequence, the worst-case parameter estimation error decreases to zero as the noise bound decreases to zero. In the terminology introduced by Helmicki *et al.* (1991), the least squares algorithm is robustly convergent. [We note that there is an important technical difference between the notion of robust convergence used in this paper and that in Helmicki *et al.* (1991).] While the problem formulation is motivated from the deterministic worst-case identification theory, the techniques employed in this paper draw upon the results in classical identification theory (Ljung, 1987). These results are similar in spirit to those in Wahlberg and Ljung (1992).

* Received 27 April 1992; revised 14 October 1992; received in final form 12 February 1993. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor P. van den Hof under the direction of Editor T. Söderström. Corresponding author P. P. Khargonekar. Tel. (1) 313 764 4328; Fax (1) 313 763 8041; email/pramod@vangogh. ccscs.u.mich.edu.

† Dept. of Mechanical Engineering and Applied Mechanics, The University of Michigan, Ann Arbor, MI 48109-4315, U.S.A.

‡ Dept. of Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, MI 48109-2122, U.S.A.

1.1. *Notation.* We end this introduction with some remarks on notation. Let \mathcal{H}_∞ denote the Hardy space of bounded analytic functions in the open unit disk \mathbf{D} of the complex plane \mathbf{C} . Let $G(z)$ denote the transfer function of a single-input/single-output (SISO), finite-dimensional linear time-invariant (FDLTI), discrete-time system. A transfer function G is called stable if all its poles have magnitude strictly greater than one. Thus the transfer function of a stable LTI system belongs to \mathcal{H}_∞ . For any stable FDLTI discrete-time transfer function G , its \mathcal{H}_∞ norm is defined as

$$\|G\|_\infty := \max_{-\pi \leq \omega \leq \pi} |G(e^{j\omega})|. \quad (1.1)$$

Let $\mathbf{D}_\rho := \{z : |z| < \rho\}$, and let $\mathcal{H}_\infty(\mathbf{D}_\rho, M)$ denote a subset of \mathcal{H}_∞ , defined as

$$\mathcal{H}_\infty(\mathbf{D}_\rho, M) := \{G : G \text{ analytic in } \mathbf{D}_\rho, |G(z)| \leq M, \forall z \in \mathbf{D}_\rho\}. \quad (1.2)$$

Thus if $\rho > 1$, the systems in $\mathcal{H}_\infty(\mathbf{D}_\rho, M)$ are exponentially stable with a decay rate greater than ρ , and a gain bounded by M . Let $\mathcal{RH}_\infty(n)$ denote the set of n th order proper, stable, real-rational transfer functions. Define the class of the systems

$$\begin{aligned} \mathcal{RH}_\infty(n, m, \rho, M) &:= \{G : G(z) \\ &= \frac{\sum_{k=1}^m b_k z^k}{1 + \sum_{k=1}^n a_k z^k}, G \in \mathcal{H}_\infty(\mathbf{D}_\rho, M) \cap \mathcal{RH}_\infty(n)\}, \end{aligned} \quad (1.3)$$

where $m \leq n$. Let g be the impulse response of G . Then l_1, l_2, l_∞ norms for g are defined as

$$\begin{aligned} \|g\|_1 &:= \sum_{k=0}^{\infty} |g(k)|, \quad \|g\|_2 := \left(\sum_{k=0}^{\infty} |g(k)|^2 \right)^{1/2}, \\ \|g\|_\infty &:= \sup_{0 \leq k < \infty} |g(k)|. \end{aligned} \quad (1.4)$$

Given an r by s matrix R , its $\|\cdot\|_1$ norm is defined as

$$\|R\|_1 := \max_{1 \leq k \leq r} \sum_{l=1}^s |R_{kl}|. \quad (1.5)$$

For vectors, $\|\cdot\|_2$ will denote the Euclidean norm. Finally, q^{-1} will denote the delay operator, *i.e.*

$$(q^{-1}u)(t) := u(t-1). \quad (1.6)$$

2. Parametric system identification

We will consider the identification of an SISO, LTI, discrete-time system represented by the ARX model

$$A(q^{-1})y(t) = B(q^{-1})u(t) + v(t), \quad (2.1)$$

where

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + \dots + a_n q^{-n}; \\ B(q^{-1}) &= b_1 q^{-1} + \dots + b_n q^{-m}. \end{aligned} \quad (2.2)$$

Let

$$\theta^T = (a_1, a_2, \dots, a_n, b_1, \dots, b_m) \quad (2.3)$$

denote the vector of unknown parameters. Let

$$\varphi^T(t) = (-y(t-1), -y(t-2), \dots, -y(t-n), u(t-1), \dots, u(t-m)) \quad (2.4)$$

denote the *regression vector* of previous output and input signals. The model (2.1) can also be written

$$y(t) = \theta^T \varphi(t) + v(t). \quad (2.5)$$

Define the transfer function

$$G(z) := \frac{b_1 z + \dots + b_m z^m}{1 + a_1 z^{-1} + \dots + a_n z^{-n}} = \frac{B(z)}{A(z)}. \quad (2.6)$$

We will assume that $G \in \mathcal{RH}_\infty(n, m, \rho, M)$.

The signal $v(t)$ will be assumed to be a bounded disturbance

$$|v(t)| \leq \epsilon, \quad \forall t. \quad (2.7)$$

The past inputs $\{u(t), t \leq 0\}$ will be assumed to be bounded, but otherwise unknown, *i.e.*

$$\sup_{t \leq 0} |u(t)| \leq C_1. \quad (2.8)$$

This bound corresponds to an upper bound on the size of the initial state of the unknown system. For example, if it is known that the system is at rest at $t = 0$, we may take $C_1 = 0$.

The input sequence $\{u(t), t > 0\}$ will be chosen by the experimenter and will only be required to satisfy an upper bound

$$\sup_{t > 0} |u(t)| \leq C_2. \quad (2.9)$$

The choice of the input sequence will be discussed in greater detail in the following sections.

Suppose that $N + n$ data have been collected

$$Z^N := \{y(i), u(i) : i = 1, \dots, N + n\}. \quad (2.10)$$

An identification algorithm produces an estimate $\hat{\theta}_N$ of θ by using the prior information and the set of measured data (2.10). The *performance* of the identification algorithm is measured by the worst-case identification error defined as

$$e_N(G, \epsilon) := \sup_{\|v\|_\infty \leq \epsilon} \|\theta - \hat{\theta}_N\|_2. \quad (2.11)$$

Note that in the definition of the worst-case identification error, we have not taken a supremum over the set of unknown systems as is done in Helmicki *et al.* (1991). The reason for this will become clear after the statement of the result. Following the terminology introduced in Helmicki *et al.* (1991), we will call an identification algorithm *convergent* if

$$\lim_{\substack{\epsilon \rightarrow 0 \\ N \rightarrow \infty}} e_N(G, \epsilon) = 0. \quad (2.12)$$

Moreover, the identification algorithm is said to be 'robustly convergent (and untuned)', if in addition the algorithm does not depend on the *a priori* information ρ and M .

3. Pseudorandom binary sequences and the least squares identification algorithm

Given the above problem formulation, the well known least squares algorithm is a potential choice for the identification algorithm. It is well known that the least squares algorithm enjoys many optimality and convergence properties in stochastic formulations of the identification problem. However, relatively little is known about the performance of the least squares algorithm in the presence of bounded but otherwise arbitrary noise. In other words, relatively little seems to be known about the performance of the least squares algorithm in the worst-case deterministic problem setting described in the previous section. In this section, we will demonstrate that the *least squares method* is

robustly convergent for the identification problem posed above for a special input design using pseudorandom binary sequences. This result follows from a bound on the worst-case identification error which is the main technical contribution of this paper.

First, we will briefly review some important properties of the Galois sequences. See Schroeder (1986) and references therein for more details.

3.1. *Galois sequences.* Let $\pi_p(x) = \sum_{i=0}^{p-1} c_i x^i$ be an irreducible polynomial over the binary field $\{0, 1\}$ of degree p such that x is a primitive element. Define

$$e_{k+p} = \sum_{i=0}^{p-1} c_i e_{k+i} \pmod{2}, \quad k = 1, 2, \dots, \quad (3.1)$$

and

$$e_1 = 1, \quad e_i = 0, \quad i = 2, 3, \dots, p.$$

(It is not necessary to choose $e_i, i = 2, 3, \dots, p$ to be all zero; these can be chosen to be either zero or one.) The resulting sequence $e_k, k = 1, 2, \dots$ is called a *Galois sequence*. A Galois sequence generated by an irreducible polynomial of degree p as above is periodic with period $2^p - 1$. Using this periodicity property, the Galois sequence $\{e_k\}$ can be extended for $k = -1, -2, \dots$. Irreducible polynomials π_p over the binary field $\{0, 1\}$ up to degree $p = 168$, have been published (Stahnke, 1973). Using $p = 168$, we can generate a Galois sequence with period $2^{168} - 1$ which is of the order of 10^{51} . Each polynomial has the minimum number of terms possible (at most five) so that the linear recursion formula (3.1) is easy to implement.

Let us convert the Galois sequence $\{e_k\}$ into another sequence $\{s_k\}$ of integers according to the rule

$$s_k := (-1)^{e_k}, \quad k = \pm 1, \pm 2, \dots \quad (3.2)$$

A key property of this sequence is as follows. Define the correlation function of $\{s_k\}$ as

$$R_n := \sum_{k=0}^{2^p-2} s_k s_{k+n}. \quad (3.3)$$

It turns out that (Schroeder, 1986)

$$R_n = 2^p - 1, \quad n \equiv 0 \pmod{2^p - 1}; \quad R_n = -1 \text{ otherwise.} \quad (3.4)$$

Thus, for p sufficiently large, the ratio R_n/R_0 is approximately zero for $n \not\equiv 0 \pmod{2^p - 1}$. Hence the sequence $\{s_k\}$ is approximately uncorrelated. (Except for the p -tuple of all 1s, this sequence contains all possible p -tuples of ± 1 .) The sequence s_k will be called a *pseudorandom sequence*. It should be noted that the sequence s_k depends on p as well as π_p . For simplicity of notation, we will not show this dependence explicitly.

Galois sequences have found wide applications in communications and signal processing (Schroeder, 1986). Mäkilä (1991) has constructed a robustly convergent identification algorithm in l_1 . Tse *et al.* (1991) have used pseudorandom binary signals in their identification studies. Pseudorandom binary signals have also been used in the study of time complexity of identification (Dahleh *et al.*, 1992; Kacwicz and Milanese, 1992; Poolla and Tikku, 1992). In this paper, we will exploit the approximate pseudorandomness property for the analysis of worst-case performance of the least algorithm.

3.2. *Main result—the worst-case performance of the least squares identification.* In this subsection, we derive an upper bound on the worst-case identification error of the least squares algorithm for a special input. The length of the identification experiment will be taken to be $N + n$, where $N = 2^p - 1$ for some positive integer p which will be fixed later. We choose the inputs $\{u(t), t = 1, 2, \dots, N + n\}$, to be a scaled version of the pseudorandom sequence

$$u(t) := C_2 s_t, \quad t = 1, 2, \dots, N + n. \quad (3.5)$$

From (2.1), we can write the system of equations

$$Y_N = \theta^T \Phi_N + V_N, \tag{3.6}$$

where

$$Y_N := [y(n+1)y(n+2) \cdots y(n+N)], \tag{3.7}$$

$$\Phi_N := [\varphi(n+1)\varphi(n+2) \cdots \varphi(n+N)], \tag{3.8}$$

$$V_N := [v(n+1)v(n+2) \cdots v(n+N)]. \tag{3.9}$$

The usual *least squares estimate* of θ given Y_N, Φ_N is

$$\hat{\theta}_N := (\Phi_N \Phi_N^T)^{-1} (\Phi_N Y_N^T), \tag{3.10}$$

assuming that $\Phi_N \Phi_N^T$ is invertible.

For the model (2.1) and the pseudorandom input $u(t)$ described above, we obtain an upper bound on the worst-case error (2.11). This bound holds when the noise bound ϵ is small and the duration of the experiment $N+n$ is large. The derivation is long but straightforward and is included in the Appendix.

First, we need some definitions. Let $G \in \mathcal{H}_\infty$. Let n, m be fixed. For a vector $x \in \mathbf{R}^{n+m}$, let

$$X := \sum_{i=1}^n x_i e^{j\omega(i-1)}, \quad Y := \sum_{i=1}^m x_{n+i} e^{j\omega(i-1)}. \tag{3.11}$$

Now define the quadratic form

$$Q(z) := \frac{C_2^2}{2\pi} \int_0^{2\pi} |GX - Y|^2 d\omega, \tag{3.12}$$

and set $\lambda(G, C_2, n, m)$ by

$$\lambda(G, C_2, n, m) := \min \{Q(x) : x \in \mathbf{R}^{n+m}, \|x\|_2 = 1\}. \tag{3.13}$$

In other words, $\lambda(G, C_2, n, m)$ is the smallest eigenvalue of the quadratic form Q . Define constants ϵ_0, T, K_1, K_2 and K_3 as follows

$$\epsilon_0 := (M+1)C_2(1-\rho^{-1})^n, \tag{3.14}$$

$$T := \max \left\{ \frac{8M^2 C_1^2}{C_2^2(1-\rho^{-1})^3}, 2nM \right\}, \tag{3.15}$$

$$K_1 := \frac{nC_2^2(2M+1)^2}{(1-\rho^{-1})^3}, \quad K_2 := \frac{3C_2n(M+1)}{(1-\rho^{-1})^n}, \tag{3.16}$$

$$K_3 := (M+1)(1+\rho^{-1})^n + n^{-1/2},$$

where $C := C_1 + C_2$.

Theorem 3.1. Let $G \in \mathcal{RH}_\infty(n, m, \rho, M)$. Let $\lambda(G, C_2, n, m), \epsilon_0, T, K_1, K_2, K_3$ be as defined in (3.13)–(3.16), respectively. Let $N \geq T$. Suppose the input $u(t), t = 1, 2, \dots, N+n$ is a pseudorandom binary sequence as given by (3.5). For all $\epsilon \leq \epsilon_0$, if $\lambda > K_1 N^{-1} + K_2 \epsilon$, the worst-case identification error (2.11) of the least squares algorithm (3.10) is bounded as

$$e_N(G, \epsilon) \leq \frac{K_1 K_3}{\lambda - (K_1 N^{-1} + K_2 \epsilon) N} + \frac{K_2 K_3}{\lambda - (K_1 N^{-1} + K_2 \epsilon)} \epsilon. \tag{3.17}$$

Consequently, the least squares is robustly convergent.

In order to get a better appreciation of the various aspects of this result, we offer the following remarks.

- It turns out that λ is the limit of $\lambda_{\min}(\Phi_N \Phi_N^T / N)$, the smallest eigenvalue of $\Phi_N \Phi_N^T / N$, as $N \rightarrow \infty$ and $\epsilon \rightarrow 0$. If $m = n$, it can be shown that $\lambda \leq C_2^2 \sigma_n^2$, where σ_n is the smallest Hankel singular value of G (Adamyam *et al.*, 1971).

- From information based complexity theory, it is known that $e_N(G, \epsilon) \geq \epsilon$. Moreover, it has been shown (Dahleh *et al.*, 1992; Kacewicz and Milanese, 1992; Poolla and Tikku, 1992) that if G is an n th order finite-impulse response (FIR) system, then $e_N(G, \epsilon) = \mathcal{O}(\epsilon)$ implies that $N/2^n = \mathcal{O}(1)$. Thus, if n is large, as might be the case if G is taken to be an FIR system, then the minimal input length may be very long.

In contrast, using an ARX representation, the input length can possibly be reduced quite significantly.

- The upper bound on $e_N(G, \epsilon)$ depends on the quantity λ . This dependence prevents us taking the supremum in 2.11 with respect to $\mathcal{RH}_\infty(n, m, \rho, M)$. If an *a priori* lower bound for λ is available then this can be done and appropriate robust convergence results can be obtained. A proof of this result is given in the Appendix.

4. Conclusions

In this paper, we have analyzed the worst-case performance of the least squares algorithm. We have obtained an upper bound on the worst-case parameter identification error which is used to establish appropriate convergence properties.

There are many different directions in which further research is necessary. We feel that *suitable modifications* of the least squares algorithm for the problem of parametric system identification in the presence of arbitrary but bounded noise may be very useful. Also, it is necessary to develop good methods that allow for undermodeling in this problem context.

References

- Adamyam, V. M., D. Z. Arov and M. G. Krein (1971). Analytic properties of Schmidt pairs for a Hankel operator and the generalized Schur–Takagi problem. *Math. USSR Sbornik*, **15**, 31–73.
- Chen, J., C. N. Nett and M. K. H. Fan (1992). Optimal non-parametric system identification from arbitrary corrupt finite time series: a worst case/deterministic approach. *Proc. American Control Conference*, pp. 279–285.
- Dahleh, M. A., T. Theodosopoulos and J. N. Tsitsiklis (1992). The sample complexity of worst-case identification of FIR linear systems. Preprint, LIDS, MIT, Cambridge, MA, U.S.A.
- Grenander, U. and G. Szegö (1958). *Toeplitz Forms and Their Applications*. University of California Press, Berkeley, CA, U.S.A.
- Helmicki, A. J., C. A. Jacobson and C. N. Nett (1991). Control-oriented system identification: A worst-case/deterministic approach in \mathcal{H}_∞ . *IEEE Trans. Aut. Control*, **AC-36**, 1163–1176.
- Khargonekar, P. P. (1993). Identification and Robust Control. *Proceedings of the Workshop on the Modeling of Uncertainty in Control Systems*, Santa Barbara, U.S.A. Springer, Berlin, to appear.
- Kacewicz, B. and M. Milanese (1992). On the optimal experiment design in the worst-case l_1 system identification. Preprint.
- Ljung, L. (1987). *System Identification, Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ.
- Mäkilä, P. M. (1991). Robust identification and Galois sequences. *Int. J. Control*, **54**, 1189–1200.
- Poolla, K. and A. Tikku (1992). Time complexity of worst-case identification. Preprint, University of California, Berkeley, CA, U.S.A.
- Schroeder, M. R. (1986). *Number Theory in Science and Communication*. Springer, Berlin.
- Stahnke, W. (1973). Primitive binary polynomials. *Math. Comput.*, **27**, no. 124, pp. 977–980.
- Tse, D. N. C., M. A. Dahleh and J. N. Tsitsiklis (1991). Optimal asymptotic identification under bounded disturbances. Preprint, LIDS, MIT, Cambridge, MA. An abridged version is in the *Proc. 1991 American Control Conference*, pp. 1786–1787.
- Wahlberg, B. and L. Ljung (1992). Hard frequency-domain model error bounds from least-squares like identification techniques *IEEE Trans. Aut. Control*, **AC-37**, 900–912.

A. Appendix—Proofs

We will prove the theorem in several steps. First, we need some definitions. Given two signals $\{u(t) : -\infty < t \leq N+n\}$ and $\{y(t) : -\infty < t \leq N+n\}$, we define the autocorrelation

and cross correlation functions as follows

$$R_{uu}(k, l; n, N) := \frac{1}{N} \sum_{t=n+1}^{N+n} u(t-k)u(t-l), \quad 0 \leq k, l < \infty \quad (\text{A.1})$$

$$R_{yy}(k, l; n, N) := \frac{1}{N} \sum_{t=n+1}^{N+n} y(t-k)y(t-l), \quad 0 \leq k, l < \infty \quad (\text{A.2})$$

$$R_{yu}(k, l; n, N) := \frac{1}{N} \sum_{t=n+1}^{N+n} y(t-k)u(t-l), \quad 0 \leq k, l < \infty. \quad (\text{A.3})$$

Let us define three matrices R_{yy} , R_{uu} , R_{yu} and two vectors R_y , R_u as follows

$$\begin{aligned} R_{yy}(k, l) &:= R_{yy}(k, l; n, N), \quad 1 \leq k, l \leq n; \\ R_{uu}(k, l) &:= R_{uu}(k, l; n, N), \quad 1 \leq k, l \leq m; \\ R_{yu}(k, l) &:= R_{yu}(k, l; n, N), \quad 1 \leq k \leq n, 1 \leq l \leq m; \\ R_y(l) &:= R_{yy}(0, l; n, N), \quad 1 \leq l \leq n; \\ R_u(l) &:= R_{yu}(0, l; n, N), \quad 1 \leq l \leq m. \end{aligned} \quad (\text{A.4})$$

Notice that

$$\begin{aligned} R_N &:= \begin{pmatrix} R_{yy} & -R_{yu} \\ -R_{yu}^T & R_{uu} \end{pmatrix} = \frac{1}{N} \Phi_N \Phi_N^T; \\ \Psi_N &:= \begin{pmatrix} -R_y \\ R_u \end{pmatrix} = \frac{1}{N} \Phi_N Y_N^T. \end{aligned} \quad (\text{A.5})$$

Lemma A.1. Let the future inputs $\{u(t): 1 \leq t < N+n\}$ to a system $G \in \mathcal{H}_\infty(\mathbf{D}_\rho, M)$ be as in (3.5). Let the past inputs $\{u(t): -\infty < t \leq 1\}$ to G be bounded by C_1 . Let y denote the output of G in response to the input u . Then for $1 \leq k, l \leq n$

$$\begin{aligned} R_{uu}(k, l; n, N) &= C_2^2 + \Delta R_{uu}(k, l; n, N) \\ &\quad \text{if } k = l; \Delta R_{uu}(k, l; n, N) \text{ otherwise} \\ R_{yu}(k, l; n, N) &= C_2^2 g(l-k) + \Delta R_{yu}(k, l; n, N), \\ R_{yy}(k, l; n, N) &= C_2^2 \sum_{s=1}^{\infty} g(s)g(|k-l|+s) \\ &\quad + \Delta R_{yy}(k, l; n, N), \end{aligned} \quad (\text{A.6})$$

where

$$\begin{aligned} |\Delta R_{uu}(k, l; n, N)| &\leq \frac{C_2^2}{N}, \\ |\Delta R_{yu}(k, l; n, N)| &\leq \frac{MCC_2(2\rho-1)}{N(\rho-1)^2}, \\ |\Delta R_{yy}(k, l; n, N)| &\leq \frac{M^2C_2^2(4\rho-1)}{N(\rho-1)^3}. \end{aligned} \quad (\text{A.7})$$

Proof. Let \bar{u} be periodic extension of the inputs in (3.5). Let $\hat{u} = u - \bar{u}$. Notice that $\hat{u}(t) = 0$, $t \geq 1$ and $\|\hat{u}(t)\|_\infty \leq C_1 + C_2$. Then, we have for $1 \leq k, l \leq n$

$$R_{uu}(k+s, l; n, N) = R_{\bar{u}\bar{u}}(k, l; n, N) + \Delta R_{uu}(k+s, l; n, N), \quad 0 \leq s, \quad (\text{A.8})$$

where

$$\Delta R_{uu}(k+s, l; n, N) := \frac{1}{N} \sum_{t=n+1}^{N+n} \hat{u}(t-k-s)\hat{u}(t-l). \quad (\text{A.9})$$

Since \bar{u} is a pseudorandom sequence, $R_{\bar{u}\bar{u}}(k+s, l; N)$ is C_2^2 if $s = l-k$ and $-C_2^2/N$ otherwise. Furthermore

$$\begin{aligned} |\Delta R_{uu}(k+s, l; n, N)| &\leq \left| \frac{1}{N} \sum_{t=n+1}^{N+n} \hat{u}(t-k-s)\hat{u}(t-l) \right| \\ &\leq \frac{s}{N} C_2 C, \quad s \geq 1, \end{aligned} \quad (\text{A.10})$$

which is also true for $s = 0$. Thus

$$\begin{aligned} R_{yu}(k, l; n, N) &= \sum_{s=1}^{\infty} g(s)R_{uu}(k+s, l; n, N) \\ &= C_2^2 g(l-k) + \Delta R_{yu}(k, l; n, N), \end{aligned} \quad (\text{A.11})$$

where

$$\begin{aligned} \Delta R_{yu}(k, l; n, N) &:= -\frac{C_2^2}{N} \sum_{s \neq l-k} g(s) \\ &\quad + \sum_{s=1}^{\infty} g(s) \Delta R_{uu}(k+s, l; n, N). \end{aligned} \quad (\text{A.12})$$

Therefore

$$|\Delta R_{yu}(k, l; n, N)| \leq \frac{C_2^2}{N} \|g\|_1 + \frac{C_2 C}{N} \sum_{s=1}^{\infty} s |g(s)|. \quad (\text{A.13})$$

From

$$\begin{aligned} \Delta R_{uu}(k+s_2, l+s_1; n, N) &:= \frac{1}{N} \sum_{t=n+1}^{N+n} \hat{u}(t-k-s_2)\hat{u}(t-l-s_1) \\ &\quad + \frac{1}{N} \sum_{t=n+1}^{N+n} \bar{u}(t-k-s_2)\bar{u}(t-l-s_1) \\ &\quad + \frac{1}{N} \sum_{t=n+1}^{N+n} \hat{u}(t-k-s_2)\bar{u}(t-l-s_1), \quad 1 \leq s_1, s_2 < \infty. \end{aligned} \quad (\text{A.14})$$

We have for $1 \leq k, l \leq n$

$$\begin{aligned} |\Delta R_{uu}(k+s_2, l+s_1; n, N)| &\leq \frac{s_2}{N} C_2 C + \frac{s_1}{N} C_2 C \\ &\quad + \frac{\min\{s_1, s_2\}}{N} C^2 \leq \frac{3(s_1+s_2)}{2N} C^2. \end{aligned} \quad (\text{A.15})$$

Hence

$$\begin{aligned} R_{yy}(k, l; n, N) &= C_2^2 \sum_{s=1}^{\infty} g(s)g(k-l+s) \\ &\quad + \Delta R_{yy}(k, l; n, N), \end{aligned} \quad (\text{A.16})$$

where

$$\begin{aligned} \Delta R_{yy}(k, l; n, N) &:= -\frac{C_2^2}{N} \sum_{s_2=1}^{\infty} \sum_{s_1 \neq l-k+s_2} g(s_1)g(s_2) \\ &\quad + \sum_{s_1=1}^{\infty} \sum_{s_2=1}^{\infty} g(s_1)g(s_2) \Delta R_{uu} \\ &\quad (k+s_2, l+s_1; n, N). \end{aligned} \quad (\text{A.17})$$

Moreover

$$|\Delta R_{yy}(k, l; n, N)| \leq \frac{C_2^2}{N} \|g\|_1^2 + \frac{3C^2}{N} \|g\|_1 \sum_{s=1}^{\infty} s |g(s)|. \quad (\text{A.18})$$

Since $G \in \mathcal{H}_\infty(\mathbf{D}_\rho, M)$, by an application of Cauchy's formula $G' \in \mathcal{H}_\infty(\mathbf{D}_\rho, M/(\rho-1))$. Hence

$$\|g\|_1 \leq \frac{M}{\rho-1}, \quad (\text{A.19})$$

$$\sum_{s=1}^{\infty} s |g(s)| = \|G'\|_1 \leq \frac{M\rho}{(\rho-1)^2}. \quad (\text{A.20})$$

Therefore for $1 \leq k, l \leq n$

$$\begin{aligned} |\Delta R_{yu}(k, l; n, N)| &\leq \frac{MC_2^2}{N(\rho-1)} + \frac{M\rho C_2 C}{N(\rho-1)^2} \\ &\leq \frac{MCC_2(2\rho-1)}{N(\rho-1)^2}, \end{aligned} \quad (\text{A.21})$$

$$\begin{aligned} |\Delta R_{yy}(k, l; n, N)| &\leq \frac{M^2C_2^2}{N(\rho-1)^2} + \frac{3C^2M^2\rho}{N(\rho-1)^3} \\ &\leq \frac{M^2C_2^2(4\rho-1)}{N(\rho-1)^3}. \end{aligned} \quad (\text{A.22})$$

Corollary A.2. Consider the system given by (2.1) and let $G \in \mathcal{R}_\infty(\mathbf{D}_\rho, M)$. Let the future inputs $\{u(t): 1 \leq t < N+n\}$ to the system be as in (3.5). Let the past inputs $\{u(t): -\infty < t \leq 0\}$ to G be bounded by C_1 . Let the noise v be bounded by ϵ . Let y denote the output of the system, in response to the input u . Then for $1 \leq k, l < n$; R_{uu} , R_{yu} , R_{yy} are given by (A.6), where

$$|\Delta R_{uu}(k, l; n, N)| \leq \frac{C_2^2}{N}, \quad (\text{A.23})$$

$$\begin{aligned} |\Delta R_{yu}(k, l; n, N)| &\leq \frac{MCC_2(2\rho-1)}{N(\rho-1)^2} + \frac{C_2\epsilon}{(1-\rho^{-1})^n}, \\ |\Delta R_{yy}(k, l; n, N)| &\leq \frac{M^2C^2(4\rho-1)}{N(\rho-1)^3} + \frac{\epsilon^2}{(1-\rho^{-1})^{2n}} \\ &\quad + \frac{2M\epsilon}{(1-\rho^{-1})^n} \\ &\quad \times \left(\frac{C_1\rho}{\sqrt{2N}(\rho-1)^{3/2}} + C_2 \left(1 + \frac{n}{2N}\right) \right). \end{aligned} \quad (\text{A.24})$$

Proof. (2.1) can be rewritten as

$$y = \bar{y} + \eta, \quad (\text{A.25})$$

where

$$\eta := A^{-1}v, \quad \eta(t) = 0, \quad t \leq 0; \quad \text{and } \bar{y} = Gu. \quad (\text{A.26})$$

Let $u = u_1 + u_2$ be a decomposition of u into past inputs and future inputs, i.e.

$$\begin{aligned} u_1(t) &:= u(t), \quad t \leq 0 \quad \text{and } u_1(t) := 0, \quad t > 0; \\ u_2(t) &:= 0, \quad t \leq 0 \quad \text{and } u_2(t) := u(t), \quad t > 0. \end{aligned} \quad (\text{A.27})$$

Let $y_p := Gu_1$ and $y_f := Gu_2$. Then easy calculations show that

$$\begin{aligned} \left(\sum_{t=n+1}^{N+n} y_p^2(t-k) \right)^{1/2} &\leq \frac{MC_1}{(\rho-1)^{3/2}} \frac{\rho}{\sqrt{\rho+1}} \rho^{-(n-k)} \\ &\leq \frac{MC_1\rho}{\sqrt{2}(\rho-1)^{3/2}}, \quad 1 \leq k \leq n, \\ \left(\sum_{t=n+1}^{N+n} y_f^2(t-k) \right)^{1/2} &\leq \left(\sum_{t=1}^{N+n} y_f^2(t-k) \right)^{1/2} \\ &\leq \|G\|_\infty C_2 \sqrt{N+n} \\ &\leq \sqrt{N} \left(1 + \frac{n}{2N}\right) MC_2, \quad 1 \leq k \leq n. \end{aligned} \quad (\text{A.28})$$

Thus

$$\begin{aligned} \left(\sum_{t=n+1}^{N+n} \bar{y}(t-k)^2 \right)^{1/2} &\leq \frac{MC_1\rho}{\sqrt{2}(\rho-1)^{3/2}} \\ &\quad + \sqrt{N} \left(1 + \frac{n}{2N}\right) MC_2, \quad 1 \leq k \leq n, \end{aligned} \quad (\text{A.29})$$

$$\left(\sum_{t=n+1}^{N+n} \eta(t-k)^2 \right)^{1/2} \leq \sqrt{N} \|A^{-1}\|_\infty \epsilon, \quad 1 \leq k \leq n. \quad (\text{A.30})$$

We can write

$$\begin{aligned} R_{yu}(k, l; n, N) &= R_{\bar{y}u}(k, l; n, N) + R_{\eta u}(k, l; n, N), \\ R_{yy}(k, l; n, N) &= R_{\bar{y}\bar{y}}(k, l; n, N) + R_{\eta\eta}(k, l; n, N) \\ &\quad + R_{\bar{y}\eta}(k, l; n, N) + R_{\eta\bar{y}}(k, l; n, N), \end{aligned} \quad (\text{A.31})$$

where $R_{\bar{y}u}$ and $R_{\bar{y}\bar{y}}$ are given by (A.6). Now an application of Schwarz inequality results in

$$|R_{\eta u}(k, l; n, N)| \leq \|A^{-1}\|_\infty C_2 \epsilon, \quad 1 \leq k, l \leq n, \quad (\text{A.32})$$

$$\begin{aligned} |R_{\bar{y}\eta}(k, l; n, N)| &\leq \frac{1}{\sqrt{N}} \left(\sum_{t=n+1}^{N+n} \bar{y}(t-k)^2 \right)^{1/2} \\ &\quad \times \|A^{-1}\|_\infty \epsilon, \quad 1 \leq k, l \leq n, \end{aligned} \quad (\text{A.33})$$

$$|R_{\eta\eta}(k, l; n, N)| \leq \|A^{-1}\|_\infty^2 \epsilon^2, \quad 1 \leq k, l \leq n. \quad (\text{A.34})$$

Finally, since G is analytic in \mathbf{D}_ρ , $A(z)$ does not have zeros

in \mathbf{D}_ρ . Therefore $A(z)$ can be factored as

$$A(z) = \prod_{i=1}^n (1 + p_i^{-1}z), \quad (\text{A.35})$$

where $p_i, i = 1, \dots, n$ are the poles of G . These observations lead to

$$(1 - \rho^{-1})^n \leq \|A\|_\infty \leq (1 + \rho^{-1})^n. \quad (\text{A.36})$$

■

As a consequence of Corollary A.2, we may define a matrix R and a vector Ψ

$$R := \lim_{\epsilon \rightarrow 0, N \rightarrow \infty} R_N; \quad \Psi := \lim_{\epsilon \rightarrow 0, N \rightarrow \infty} \Psi_N. \quad (\text{A.37})$$

Let $\Delta R_N := R_N - R$ and $\Delta \Psi_N := \Psi_N - \Psi$. The frequency-domain expression of R allows us to show that the smallest eigenvalue of R is $\lambda(G, C_2, n, m)$ and the following fact.

Fact A.3. Let R be as in A.37 for a system $G \in \bigcup_{i=n}^{\infty} \mathcal{R}\mathcal{H}_\infty(i)$.

Then R is invertible. If $G \in \mathcal{R}\mathcal{H}_\infty(n, m)$, then $\lim_{\epsilon \rightarrow 0, N \rightarrow \infty} \hat{\theta}_N = \theta$.

Proof of Theorem 3.1. We only need to derive the error bounds. According to Theorem 1.18 in Grenander and Szego (1958), the largest eigenvalue $\lambda_{\max}(\Delta R_N)$ of ΔR_N is bounded above by $\|\Delta R_N\|_1$. Hence if $\lambda > \lambda_0 \geq \|\Delta R_N\|_1$ for some λ_0 , from the following decomposition

$$\begin{aligned} \hat{\theta}_N - \theta &= R_N^{-1} \Psi_N - R^{-1} \Psi = (R + \Delta R_N)^{-1} \Delta \Psi_N \\ &\quad - (R + \Delta R_N)^{-1} (\Delta R_N) \theta \end{aligned} \quad (\text{A.38})$$

we get

$$\|\hat{\theta}_N - \theta\|_2 \leq \frac{\|\Delta \Psi_N\|_2 + \|\theta\|_2 \|\Delta R_N\|_1}{\lambda - \|\Delta R_N\|_1}. \quad (\text{A.39})$$

From Corollary A.2, since $m \leq n$ and $C_2 \leq C$, we have

$$\begin{aligned} \|\Delta R_{yy}\|_1 + \|\Delta R_{yu}\|_1 &\leq \left(M \frac{4\rho-1}{\rho-1} + 2\rho-1 \right) \frac{nMC^2}{N(\rho-1)^2} \\ &\quad + \left((2M+1) + \frac{nM}{N} \right) C_2 \\ &\quad + \frac{2MC_1\rho}{\sqrt{2N}(\rho-1)^{3/2}} + \frac{\epsilon}{(1-\rho^{-1})^n} \\ &\quad \times \frac{n\epsilon}{(1-\rho^{-1})^n}, \end{aligned} \quad (\text{A.40})$$

$$\begin{aligned} \|\Delta R_{yu}\|_1 + \|\Delta R_{uu}\|_1 &\leq n \left(1 + M \frac{2\rho-1}{(\rho-1)^2} \right) \\ &\quad \times \frac{CC_2}{N} + \frac{nC_2\epsilon}{(1-\rho^{-1})^n}. \end{aligned} \quad (\text{A.41})$$

Let $T := \max \{8(MC_1/C_2)^2(1-\rho^{-1})^{-3}, 2nM\}$. If $N \geq T$, then $\sqrt{2}N^{-1/2}MC_1\rho(\rho-1)^{-3/2} + nMC_2N^{-1} \leq C_2$. Let $\epsilon_0 := (M+1)C_2(1-\rho^{-1})^n$. Then, for $N \geq T$ and $\epsilon \leq \epsilon_0$, we have

$$\begin{aligned} \left((2M+1) + \frac{nM}{N} \right) C_2 + \frac{2MC_1\rho}{\sqrt{2N}(\rho-1)^{3/2}} + \frac{\epsilon}{(1-\rho^{-1})^n} \\ \times \frac{n\epsilon}{(1-\rho^{-1})^n} \leq \frac{3(M+1)C_2n}{(1-\rho^{-1})^n} \epsilon. \end{aligned} \quad (\text{A.42})$$

Since $\rho > 1$, the following inequalities hold

$$\begin{aligned} \left(M \frac{4\rho-1}{\rho-1} + 2\rho-1 \right) \frac{nMC^2}{N(\rho-1)^2} &\leq \frac{2nMC^2(2M+1)}{N(1-\rho^{-1})^3} \\ &\leq \frac{nC^2(2M+1)^2}{N(1-\rho^{-1})^3}, \end{aligned} \quad (\text{A.43})$$

$$n\left(1 + M \frac{2\rho - 1}{(\rho - 1)^2}\right) \frac{CC_2}{N} \leq \frac{n(2M + 1)CC_2}{N(1 - \rho^{-1})^2} \leq \frac{nC^2(2M + 1)^2}{N(1 - \rho^{-1})^3}. \quad (\text{A.44})$$

Then, from the inequalities (A.40)–(A.44), an upper bound on $\|\Delta R_N\|_1$ is obtained as

$$\|\Delta R_N\|_1 \leq \lambda_0 := \frac{nC^2(2M + 1)^2}{N(1 - \rho^{-1})^3} + \frac{3(M + 1)C_2n}{(1 - \rho^{-1})^n} \epsilon. \quad (\text{A.45})$$

Since the upper bounds on $\Delta R_{uu}(k, l; n, N)$,

$\Delta R_{yu}(l, l; n, N)$, $\Delta R_{yy}(k, l; n, N)$ given in Corollary A.2 are uniform in k and l in the range of interest, an upper bound on $\|\Delta \Psi_N\|_2$ is obtained as follows

$$\|\Delta \Psi_N\|_2 \leq n^{-1/2} \left(\frac{nC^2(2M + 1)^2}{N(1 - \rho^{-1})^3} + \frac{3(M + 1)C_2n}{(1 - \rho^{-1})^n} \epsilon \right) \leq n^{-1/2} \lambda_0. \quad (\text{A.46})$$

Finally, we have the following estimate on $\|\theta\|_2$

$$\|\theta\|_2 \leq (\|A\|_\infty^2 - 1 + \|B\|_\infty^2)^{1/2} \leq (M + 1) \|A\|_\infty \leq (M + 1)(1 + \rho^{-1})^n. \quad (\text{A.47})$$

Using (A.45), (A.46), (A.47), the desired error bound follows. ■