Robust Identification of Systems With Both Bias and Variance Disturbances*

GUO Lei (郭 雷) and WEI Chen (魏 晨)
(*Institute of Systems Science, Academia Sinica, Beijing* 100080, *PRC*)

Received January 11, 1994; revised April 14, 1994.

Keywords: robust identification, least-squares, model reduction.

In recent years, robust identification has become an area of growing interest, mainly because of its relevance to robust control design. For various reasons, low order nominal models are preferred in controller or filter designs, yet the true plants are usually of high or infinite order, with unmodeled dynamics or random/deterministic disturbances.

There are basically two ways of identifying infinite dimensional linear systems in literature: One (direct identification) is to obtain a low-order nominal model directly from the observation data by, for example, the prediction error methods[1] or the set-membership methods[2]. This way usually involves certain global optimization problems, and the computation is difficult except for some special cases. The other (indirect identification) is to first get a possibly high-order estimated model based on the measurements by, for example, the methods of least-squares-based approximation[3—6] or frequency response-based identification[7,8], and then to get a low-order nominal model through model reduction techniques.

The current note follows the latter way described above, and is based on some preliminary results in Refs. [4] and [6]. The main contribution of this note is as follows.

For a large class of linear systems, the bounded deterministic disturbances are shown to have no influence on L^∞-strong consistency of the identification algorithm.

This is a rather surprising result, as in the bounded noise case almost all of the previous studies do not show the strong consistency of the estimates, and only end up with the conclusion that the identification error is bounded by the magnitude of the deterministic disturbances or by functions thereof[7,8].

1 Main Results

Consider the following discrete-time SISO system:

* Project supported by the National Natural Science Foundation of China.

$$y_t = G(z) u_t + w_t + v_t, \ t \geqslant 0, \tag{1}$$

where y_t and u_t are respectively the system output and input sequences, w_t and v_t are respectively random and deterministic disturbances, $y_t = u_t = w_t = v_t = 0$, $\forall t < 0$, and $G(z)$ is the transfer function of the system

$$G(z) = \sum_{i=1}^{\infty} g_i z^i, \ \sum_{i=1}^{\infty} |g_i| < \infty, \tag{2}$$

where z is the backward-shift operator. We assume that

(i) $\{w_t, \mathscr{F}_t\}$ is a martingale difference sequence with respect to a sequence $\{\mathscr{F}_t\}$ of non-decreasing σ-algebras, and for some $\delta > 0$,

$$\sup_t E[w_{t+1}^2 | \mathscr{F}_t] < \infty, \ \sup_t E|w_t|^{4+\delta} < \infty. \tag{3}$$

(ii) $\{v_t\}$ is a bounded deterministic sequence, i.e.

$$|v_t| \leqslant c_v, \ \forall t, \text{ for some } c_v > 0. \tag{4}$$

(iii) The system input $\{u_t, \mathscr{F}_t\}$ is an adapted ARMA process of the form

$$A(z) u_t = B(z) e_t, \tag{5}$$

where $A(z)$ and $B(z)$ are polynomials of z with $A(z) \neq 0$, $|z| \leqslant 1$; $B(z) \neq 0$, $|z| = 1$, and $\{e_t\}$ is an independent and identically distributed random sequence with

$$Ee_t = \mu, \ \mathrm{Var}\{e_t\} = \sigma^2 > 0, \ E\{|e_t|^{4+\varepsilon}\} < \infty, \tag{6}$$

for some $\varepsilon > 0$.

The sequence $\{w_t\}$ is referred to as "variance disturbance" because it has zero mean but nonzero variance, while the sequence $\{v_t\}$ is called "bias disturbance" because it is nonzero in any averaging sense. Also, it is worth noting that condition (iii) is a standard one in the literature of system identification[1].

1.1 *Transfer Function Approximation*

Let $\{h_n\}$ be a non-decreasing sequence of positive integers with $h_n \to \infty$ and $h_n = O((\log n)^\alpha)$, $(\alpha > 1)$. Set

$$\theta(n) = [g_1 \cdots, \ g_{h_n}]^\tau, \tag{7}$$

$$\Phi_i(n) = [u_i, \cdots, \ u_{i-h_n+1}]^\tau, \ 1 \leqslant i \leqslant n. \tag{8}$$

The least-squares estimate $\hat{\theta}(n)$ for $\theta(n)$ at time n is given by

$$\hat{\theta}(n) = \left[\sum_{i=0}^{n-1} \Phi_i(n) \Phi_i^\tau(n) + \gamma I\right]^{-1} \sum_{i=0}^{n-1} \Phi_i(n) y_{i+1}, \tag{9}$$

with $\gamma > 0$ arbitrarily chosen.

Write $\hat{\theta}(n)$ in its component form

$$\hat{\theta}(n)=[g_1(n),\cdots,\ g_{h_n}(n)]^{\tau}, \tag{10}$$

and denote the estimate of $G(z)$ at time n by

$$\hat{G}_n(z)=\sum_{i=1}^{h_n} g_i(n)z^i. \tag{11}$$

Theorem 1. *Consider systems* (1) *and* (2) *and the transfer function estimate defined by* (11). *If Assumptions* (i)—(iii) *hold, then*

$$\|\hat{G}_n(z)-G(z)\|_{\infty}$$
$$=O\left(\sqrt{h_n\delta_n}+(1+c_v|a|h_n)h_n\sqrt{\frac{\log\log n}{n}}\right)+\frac{r_1c_v|a|h_n}{r_0(a^2h_n+r_1)}\quad a.s., \tag{12}$$

where c_v *is defined by* (4) *and*

$$\delta_n=\left(\sum_{i=h_n+1}^{\infty}|g_i|\right)^2,\ a=\left[\frac{B(1)}{A(1)}\right]\mu, \tag{13}$$

$$r_0=\min_{\lambda\in[0,2\pi]}\left|\frac{B(e^{i\lambda})}{A(e^{i\lambda})}\right|^2\sigma^2,\ r_1=\max_{\lambda\in[0,2\pi]}\left|\frac{B(e^{i\lambda})}{A(e^{i\lambda})}\right|^2\sigma^2. \tag{14}$$

Remark 1. If we take $\mu=0$ (i.e. the input sequence is of zero mean), then $a=0$ and Theorem 1 reads

$$\|\hat{G}_n(z)-G(z)\|_{\infty}=O\left(\sqrt{h_n\delta_n}+h_n\sqrt{\frac{\log\log n}{n}}\right)\text{ a.s.}, \tag{15}$$

which tends to zero under some mild conditions on the decaying rate of δ_n (or g_n), for example, $|g_n|=O(n^{-\beta})$, $\beta>\frac{3}{2}$. This result is rather surprising since, contrary to the recent results in H^{∞}-identification[7, 8], the bias disturbance $\{v_i\}$ does not upset the L^{∞}-strong consistency of the identification algorithm.

1.2 *Model Reduction*

Because the degree of the estimated transfer function $\hat{G}_n(z)$ increases with n, it may not be desirable in some applications, and the model reduction will be considered now.

Here we adopt an input normal realization truncation which has properties similar to balanced realization truncation, but is easier to compute[9, 10]. A simple realization $(\Phi_n,\ \Gamma_n,\ C_n)$ for $\hat{G}_n(z)$ is given by

$$\Phi_n=\begin{bmatrix} 0\cdots 0 & 0 \\ 1 & \\ \ddots & \\ & 1 \quad 0 \end{bmatrix}, \quad \Gamma_n=\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad C_n^{\tau}=\begin{bmatrix} g_1(n) \\ \vdots \\ g_{h_n}(n) \end{bmatrix}. \tag{16}$$

Define the controllability and observability Gramians respectively by P and Q, i.e.

$$P=\sum_{i=0}^{\infty} \Phi_n^i \Gamma_n \Gamma_n^{\tau} (\Phi_n^{\tau})^i \text{ and } Q=\sum_{i=0}^{\infty} (\Phi_n^{\tau})^i C_n^{\tau} C_n \Phi_n^i. \tag{17}$$

Then P and Q are the unique solutions to the following Lyapunov equations:

$$P-\Phi_n P \Phi_n^{\tau}=\Gamma_n \Gamma_n^{\tau} \text{ and } Q-\Phi_n^{\tau} Q \Phi_n=C_n^{\tau} C_n. \tag{18}$$

Substituting (16) into (18), we get $P=I_{h_n}$. Let the singular value decomposition of Q be $Q=T^{\tau}\Sigma^2 T$, where Σ is a diagonal matrix whose diagonal elements are Hankel singular values of $\hat{G}_n(z)$ in descending order and T is an orthogonal matrix. Define the input normal realization by

$$[\overline{\Phi}_n, \overline{\Gamma}_n, \overline{C}_n]=[T\Phi_n T^{\tau}, T\Gamma_n, C_n T^{\tau}].$$

Then the approximant $\hat{G}_n^k(z)$ of degree k can be obtained by a direct truncation of $[\overline{\Phi}_n, \overline{\Gamma}_n, \overline{C}_n]$ as follows[10]:

$$\Phi_n^k=[I_k, 0]\overline{\Phi}_n\begin{bmatrix} I_k \\ 0 \end{bmatrix}, \Gamma_n^k=[I_k, 0]\overline{\Gamma}_n, \ C_n^k=\overline{C}_n\begin{bmatrix} I_k \\ 0 \end{bmatrix},$$

$$\hat{G}_n^k(z) \triangleq C_n^k[z^{-1}I-\Phi_n^k]^{-1}\Gamma_n^k. \tag{19}$$

Theorem 2. *For the kth order estimated transfer function $\hat{G}_n^k(z)$ defined by* (19), *the identification/approximation error is bounded by*

$$\|G(z)-\hat{G}_n^k(z)\|_{\infty} \leqslant 2\sum_{i=k+1}^{h_n} \sigma_i(G(z))+[2(h_n-k)+1]\|\hat{G}_n(z)-G(z)\|_{\infty}, \tag{20}$$

where $\sigma_i(G(z))$ is the ith Hankel singular value of $G(z)$, $\sigma_i(G(z)) \geqslant \sigma_{i+1}(G(z))$, $\forall i>0$, and $\hat{G}_n(z)$ is the same as that in Theorem 1.

Remark 2. If $\hat{G}_n(z)$ has the following convergence rate

$$\|\hat{G}_n(z)-G(z)\|_{\infty}=o(h_n^{-1}) \text{ a.s.}, \tag{21}$$

then by Theorem 2 we have

$$\lim_{n\to\infty} \|\hat{G}_n^k(z)-G(z)\|_{\infty} \leqslant 2\sum_{i=k+1}^{h_n} \sigma_i(G(z)) \text{ a.s.} \tag{22}$$

Consequently, the estimated transfer function $\hat{G}_n^k(z)$ is strongly consistent in the sense that if the McMillan degree of $G(z)$ is k, then $\lim_{n\to\infty} \|\hat{G}_n^k(z)-G(z)\|_{\infty}=0$ a.s. This follows directly from (22) since in this case $\sigma_i(G(z))=0$, $\otimes i>k$.

The requirement (21) is fulfilled if, for instance, in addition to the conditions of Theorem 1, the input sequence is of zero mean and $\{g_n\}$ has a decaying rate $|g_n|=O(n^{-\beta})$ with $\beta>\frac{5}{2}$ (see Remark 1).

2 Proofs of the Main Results

We first present several lemmas which can be derived based on the results in Refs. [4— 6], the proof details are omitted here due to space limitations.

Lemma 1. *Let* $\{w_j, \mathscr{F}_j\}$ *satisfy condition* (i), *and let* $\{x_j, \mathscr{F}_j\}$ *be any adapted random sequence satisfying*

$$\sum_{i=1}^{n} x_i^2=O(n),\ \sup_n E|x_n|^{4+\delta}<\infty \ a.\,s.,$$

where $\delta>0$ *is a constant. Then as* $n\to\infty$,

$$\max_{1\leqslant t\leqslant(\log n)^\alpha}\ \max_{1\leqslant i\leqslant n}\left|\sum_{j=1}^{i} x_{j-t}w_j\right|=O(\sqrt{n\log\log n}\,)\ a.s.,\ \otimes\alpha>0.$$

Lemma 2. *Let conditions* (ii) *and* (iii) *hold with* $\mu=0$, *where* μ *is defined in* (6). *Then*

$$\max_{0\leqslant k\leqslant(\log n)^\alpha}\left|\sum_{i=0}^{n-1} u_{i-k}v_{i+1}\right|=O(\sqrt{n\log\log n}\,)\ a.s.,\ \otimes\alpha>0.$$

Lemma 3. *Let* $\{u_t\}$ *satisfy condition* (iii). *Then*

$$\lim_{0\leqslant k,\ t\leqslant(\log n)^\alpha}\left|\frac{1}{n}\sum_{i=0}^{n-1}[u_{i-k}u_{i-t}-E(u_{i-k}u_{i-t})]\right|=O\left(\sqrt{\frac{\log\log n}{n}}\right)\ a.s.,\ \otimes\alpha>0.$$

Now, we are in a position to prove our main results.

Proof of Theorem 1 (Outline). Set

$$L(z)=[z,\cdots,z^{h_n}],\ G_n(z)=\sum_{i=1}^{h_n} g_i z^i, \tag{23}$$

$$P_n=\sum_{i=0}^{n-1}\Phi_i(n)\Phi_i^\tau(n)+\gamma I,\ \varepsilon_i(\hat{n})=\sum_{j=h_n+1}^{\infty} g_j u_{i-j+1}. \tag{24}$$

By (13) and the triangle inequality

$$\|\hat{G}_n(z)-G(z)\|_\infty\leqslant\|\hat{G}_n(z)-G_n(z)\|_\infty+\sqrt{\delta_n}\,. \tag{25}$$

Furthermore, by (7)—(11) and (23)—(24),

$$\|\hat{G}_n(z)-G_n(z)\|_\infty=\|L(z)(\hat{\theta}(n)-\theta(n))\|_\infty$$
$$\leqslant\|L(z)\hat{P}_n^{-1}\sum_{i=0}^{n-1}\Phi_i(n)\varepsilon_i(n)\|_\infty+\|L(z)\hat{P}_n^{-1}\sum_{i=0}^{n-1}\Phi_i(n)w_{i+1}\|_\infty$$
$$+\|L(z)\hat{P}_n^{-1}\sum_{i=1}^{n-1}\Phi_i(n)v_{i+1}\|_\infty+O\left(\frac{\sqrt{h_n}}{n}\right)\text{ a.s.} \tag{26}$$

Now, by Lemmas 1— 3 and the matrix inverse formula, we have the following estimations for the first three terms on the RHS of (26):

$$\left\|L(z)\hat{P}_n^{-1}\sum_{i=0}^{n-1}\Phi_i(n)\varepsilon_i(n)\right\|_\infty=O(\sqrt{h_n\delta_n})\text{ a.s.,} \tag{27}$$

$$\left\|L(z)\hat{P}_n^{-1}\sum_{i=0}^{n-1}\Phi_i(n)w_{i+1}\right\|_\infty=O\left(h_n\sqrt{\frac{\log\log n}{n}}\right), \tag{28}$$

and

$$\left\|L(z)\hat{P}_n^{-1}\sum_{i=0}^{n-1}\Phi_i(n)v_{i+1}\right\|_\infty$$
$$=O\left((1+h_n|a|c_v)h_n\sqrt{\frac{\log\log n}{n}}\right)+\frac{r_1|a|c_vh_n}{r_0(a^2h_n+r_1)}\quad\text{a.s.,}$$

which in conjunction with (25)—(28) yields Theorem 1. #

Proof of Theorem 2. By the triangle inequality, it is known that

$$\|G(z)-\hat{G}_n^k(z)\|_\infty\leqslant\|G(z)-\hat{G}_n(z)\|_\infty+\|\hat{G}_n(z)-\hat{G}_n^k(z)\|_\infty. \tag{29}$$

By Theorem 2 in Ref. [11] and the fact that the truncation error using an input normal realization is no larger than that of a balanced realization[10], we get

$$\|\hat{G}_n(z)-\hat{G}_n^k(z)\|_\infty\leqslant 2\sum_{i=k+1}^{h_n}\sigma_i(\hat{G}_n(z)). \tag{30}$$

Also, by Ref. [12] and the fact that the Hankel-norm is not greater than the L^∞-norm, we have

$$|\sigma_i(\hat{G}_n(z))-\sigma_i(G(z))|\leqslant\|\hat{G}_n(z)-G(z)\|_\infty. \tag{31}$$

From (29)—(31) it is easy to see that Theorem 2 holds.

References

1 Ljung, L., *System Identification: Theory for the User*, Englewood Cliffs, Prentice-Hall, New Jersey, 1987.
2 Kosut, R. L., Lau, M. K. & Boyd, S. P., *IEEE Trans. Automatic Control*, 1992, **37**: 929.
3 Durbin, J., *Internat. Statist. Rev.*, 1961, **28**: 233.
4 Guo, L., Huang, D. W. & Hannan, E. J., *J. Multivariate Anal.*, 1990, **32**: 17.

5 Hannan, E. J. & Deistler, M., *The Statistical Theory of Linear Systems*, Wiley, New York, 1988.

6 Huang, D. W. & Guo, L., *The Annals of Statistics*, 1990, **18**: 1729.

7 Helmicki, A. J., Jacobson, C. N. & Nett, C. N., *IEEE Trans. Automatic Control*, 1991, **36**: 1163.

8 Partington, J. R., *International J. of Control*, 1991, **54**: 1281.

9 Glover, K., Curtain, R. F. & Partington, J. R. *SIAM J. Control Optim.*, 1988, **26**:863.

10 Gu, G., Khargonekar, P. P. & Lee, E. B., *IEEE Trans. Automatic Control* , 1989, **34**: 610.

11 Al-Saggaf, U. M. & Franklin, G. F., *IEEE Trans. Automatic Control*, 1987, **AC-32**: 815.

12 Gohberg, I. C. & Krein, M. G., *Trans. Math. Monographs*, 1969, **18**: 30.