12

Identification of Hammerstein Systems with Quantized Observations

This chapter concerns the identification of Hammerstein systems whose outputs are measured by quantized sensors. The system consists of a memoryless nonlinearity that is polynomial and possibly noninvertible, followed by a linear subsystem. The parameters of linear and nonlinear parts are unknown but have known orders. We present input design, identification algorithms, and their essential properties under the assumptions that the distribution function of the noise and the quantization thresholds are known. Also introduced is the concept of strongly scaled full-rank signals to capture the essential conditions under which the Hammerstein system can be identified with quantized observations. Then under strongly scaled full-rank conditions, we construct an algorithm and demonstrate its consistency and asymptotic efficiency.

The structure of Hammerstein models using quantized observations is formulated in Section 12.1. The concepts of strongly full-rank signals and their essential properties are introduced in Section 12.2. Under strongly full rank inputs, estimates of unknown parameters based on individual thresholds are constructed in Section 12.3. Estimation errors for these estimates are established. The estimates are integrated in an optimal quasi-convex combination estimator (QCCE) in Section 12.4. The resulting estimates are shown to be strongly convergent. Their efficiency is also investigated. The algorithms are expanded in Section 12.5 to derive identification algorithms for both the linear and nonlinear parts. Illustrative examples are presented in Section 12.6 on input design and convergence properties of the methodologies and algorithms.

L.Y. Wang et al., *System Identification with Quantized Observations*, Systems & Control: Foundations & Applications, DOI 10.1007/978-0-8176-4956-2_12,

12.1 Problem Formulation

Consider the system in Figure 12.1, in which

$$\begin{cases} y_k = \sum_{i=0}^{n_0-1} a_i x_{k-i} + d_k, \\ x_k = b_0 + \sum_{j=1}^{q_0} b_j u_k^j, \quad b_{q_0} = 1, \end{cases}$$

where u_k is the input, x_k the intermediate variable, and d_k the measurement noise. Both n_0 and q_0 are known.

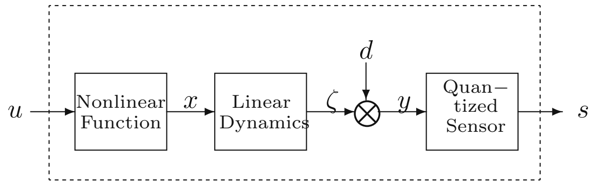


FIGURE 12.1. Hammerstein systems with quantized observations

The output y_k is measured by a sensor, which is represented by the indicator functions

$$s_k^{\{i\}} = I_{\{y_k \le C_i\}}, \quad i = 1, \dots, m_0,$$

where C_i for $i = 1, \dots, m_0$ are the thresholds. Denote $\theta = [a_0, \dots, a_{n_0-1}]'$, $\phi_k^0 = [1, \dots, 1]'$, and $\phi_k^j = [u_k^j, \dots, u_{k-n_0+1}^j]'$, $j = 1, \dots, q_0$. Then

$$\begin{aligned} y_k &= \sum_{i=0}^{n_0-1} a_i (b_0 + \sum_{j=1}^{q_0} b_j u_{k-i}^j) + d_k \\ &= b_0 \sum_{i=0}^{n_0-1} a_i + \sum_{j=1}^{q_0} b_j \sum_{i=0}^{n_0-1} a_i u_{k-i}^j + d_k \\ &= \sum_{j=0}^{q_0} b_j (\phi_k^j)' \theta + d_k. \end{aligned} \quad (12.1)$$

By using the vector notation, for $k = 1, 2, \ldots$,

$$\begin{aligned}
Y_l &= [y_{2(l-1)n_0(q_0+1)+n_0}, \ldots, y_{2ln_0(q_0+1)+n_0-1}]' \in \mathbb{R}^{2n_0(q_0+1)}, \\
\Phi_l^j &= [\phi^j_{2(l-1)n_0(q_0+1)+n_0}, \ldots, \phi^j_{2ln_0(q_0+1)+n_0-1}]' \in \mathbb{R}^{2n_0(q_0+1)\times n_0}, \\
&\qquad\qquad j = 0, \ldots, q_0, \\
D_l &= [d_{2(l-1)n_0(q_0+1)+n_0}, \ldots, d_{2ln_0(q_0+1)+n_0-1}]' \in \mathbb{R}^{2n_0(q_0+1)}, \\
S_l^{\{i\}} &= [s^{\{i\}}_{2(l-1)n_0(q_0+1)+n_0}, \ldots, s^{\{i\}}_{2ln_0(q_0+1)+n_0-1}]' \in \mathbb{R}^{2n_0(q_0+1)}, \\
&\qquad\qquad i = 1, \ldots, m_0, \quad l = 1, 2, \ldots,
\end{aligned} \tag{12.2}$$

we can rewrite (12.1) in block form as

$$Y_l = \sum_{j=0}^{q_0} b_j \Phi_l^j \theta + D_l, \quad l = 1, 2, \ldots \tag{12.3}$$

We proceed to develop identification algorithms of parameters θ and $\eta = [b_0, \ldots, b_{q_0-1}]'$ with the information of the input u_k and the output s_k of the quantized sensor.

The input signal, which will be used to identify the system, is a $2n_0(q_0+1)$-periodic signal u whose one-period values are

$$(v, v, \rho_1 v, \rho_1 v, \ldots, \rho_{q_0} v, \rho_{q_0} v),$$

where the base vector $v = (v_1, \ldots, v_{n_0})$ and the scaling factors are to be specified. The scaling factors 1, ρ_1, ..., ρ_{q_0} are assumed to be nonzero and distinct. Under $2n_0(q_0+1)$-periodic inputs, we have

$$\Phi_l^j = \Phi_1^j := \Phi^j, \quad l = 1, 2, \ldots$$

Thus, (12.3) can be written as

$$Y_l = \sum_{j=0}^{q_0} b_j \Phi^j \theta + D_l := \zeta + D_l. \tag{12.4}$$

The identification algorithm will be divided into two steps: (i) to estimate ζ (which can be reduced to estimation of gain systems), and (ii) to estimate θ from the estimated ζ.

12.2 Input Design and Strong-Full-Rank Signals

This section is to introduce a class of input signals, called strongly full-rank signals, which will play an important role in what follows. First, some basic properties of periodic signals will be derived.

Recall that an $n_0 \times n_0$ generalized circulant matrix

$$T = \begin{bmatrix} vn_0 & v_{n_0-1} & v_{n_0-2} & \cdots & v_1 \\ \lambda v_1 & vn_0 & v_{n_0-1} & \ddots & v_2 \\ \lambda v_2 & \lambda v_1 & vn_0 & \ddots & v_3 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \lambda v_{n_0-1} & \lambda v_{n_0-2} & \lambda v_{n_0-3} & \cdots & vn_0 \end{bmatrix} \tag{12.5}$$

is completely determined by its first row $[vn_0, \ldots, v_1]$ and λ, which will be denoted by $T(\lambda, [vn_0, \ldots, v_1])$. In the special case of $\lambda = 1$, the matrix in (12.5) is called a circulant matrix and will be denoted by $T([vn_0, \ldots, v_1])$.

Definition 12.1. An n_0-periodic signal generated from its one-period values $(v_1, \ldots, vn_0)$ is said to be *strongly full rank with order* m_0 if the circulant matrices $T([v^i n_0, \ldots, v_1^i])$ are all full rank for $i = 1, \ldots, m_0$.

Obviously, an n_0-periodic signal generated from $v = (v_1, \ldots, vn_0)$ is strongly full rank with order m_0 if it is strongly $m_0 + 1$ full rank. An important property of circulant matrices is the following frequency-domain criterion. By Lemma 2.2, we have the following theorem.

Lemma 12.2. *An n_0-periodic signal generated from $v = (v_1, \ldots, vn_0)$ is strongly full rank with order m_0 if and only if for $l = 1, 2, \ldots, m_0$,*

$$\gamma_{k,l} = \sum_{j=1}^{n_0} v_j^l e^{-\mathrm{i}\omega_k j}$$

are nonzero at $\omega_k = 2\pi k/n_0$, $k = 1, \ldots, n_0$.

Proposition 12.3. *For $n_0 = 1, 2$, an n_0-periodic signal u generated from $v = (v_1, \ldots, vn_0)$ is strongly full rank with order m_0 if and only if it is full rank.*

Proof. For $n_0 = 1$, by Corollary 2.3, u is full rank if and only if $\gamma_1 = v_1 \neq 0$. By Lemma 12.2, u is strongly full rank with order m_0 if and only if $\gamma_{1,l} = v_1^l \neq 0, \forall l$. So, $\gamma_1 \neq 0$ is equivalent to $\gamma_{1,l} \neq 0$.

For $n_0 = 2$, by Corollary 2.3, u is full rank if and only if $\gamma_1 = v_2 - v_1 \neq 0$ and $\gamma_2 = v_2 + v_1 \neq 0$, that is, $v_2 \neq \pm v_1$. By Lemma 12.2, u is strongly full rank with order m_0 if and only if

$$\gamma_{1,l}\gamma_{2,l} = (v_1^l e^{-i\pi} + v_2^l e^{-i2\pi})(v_1^l e^{-i2\pi} + v_2^l e^{-i4\pi}) = v_2^{2l} - v_1^{2l}.$$

Thus, we have $\gamma_1\gamma_2 = v_2^2 - v_1^2 \neq 0$ if and only if u is full rank. □

Remark 12.4. For $n_0 > 2$, the conditions of strongly full rank with order m_0 may be different from the conditions of full rank. For example, for $n_0 = 3$ and $l = 1, \ldots, m_0$,

$$\gamma_{1,l}\gamma_{2,l}\gamma_{3,l} = (v_3^l + v_2^l + v_1^l)\Big[(v_2^l - \frac{1}{2}(v_3^l + v_1^l))^2 + \frac{3}{4}(v_3^l - v_1^l)^2\Big] \neq 0$$

is not equivalent to

$$\gamma_1\gamma_2\gamma_3 = (v_3 + v_2 + v_1)\Big[(v_2 - \frac{1}{2}(v_3 + v_1))^2 + \frac{3}{4}(v_3 - v_1)^2\Big] \neq 0$$

except $m_0 = 1$.

Definition 12.5. A $2n_0(m_0 + 1)$-periodic signal u is strongly scaled m_0 full rank if its one-period values are $(v, v, \rho_1 v, \rho_1 v, \ldots, \rho_{m_0} v, \rho_{m_0} v)$, where $v = (v_1, \ldots, v n_0)$ is strongly full rank with order m_0, i.e., $0 \notin \mathcal{F}[v]$; $\rho_j \neq 0$, $\rho_j \neq 1$, $j = 1, \ldots, m_0$, and $\rho_i \neq \rho_j$, $i \neq j$. We use $\mathcal{U}(n_0, q_0)$ to denote the class of such signals.

Definition 12.6. An $n_0(m_0+1)$-periodic signal u is exponentially strongly scaled full rank with order m_0 signal if its one-period values are $(v, \lambda v, \ldots, \lambda^{m_0} v)$, where $\lambda \neq 0$ and $\lambda \neq 1$, and $T_j = T_j(\lambda^j, [v^j n_0, \ldots, v_1^j])$ are all full rank for $j = 1, \ldots, m_0$. We use $\mathcal{U}_\lambda(n_0, q_0)$ to denote this class of input signals.

By Definition 12.6 and Lemma 2.2, we have the following result.

Lemma 12.7. *An $n_0(q_0+1)$-periodic signal u with one-period values $(v, \lambda v, \ldots, \lambda^{m_0} v)$ is exponentially strongly scaled full rank with order m_0 if $\lambda \neq 0$, $\lambda \neq 1$, and for $l = 1, \ldots, m_0$,*

$$\gamma_{k,l} = \sum_{j=1}^{n_0} v_j^l \lambda^{-\frac{jl}{n_0}} e^{-\mathrm{i}\omega_k j}$$

are nonzero at $\omega_k = (2\pi k)/n_0$, $k = 1, \ldots, n_0$.

Remark 12.8. Definitions 12.5 and 12.6 require that $T(\lambda^i, [v^i n_0, \ldots, v_1^i])$, $i = 1, \ldots, m_0$, are all full rank for $\lambda = 1$ and $\lambda \neq 0, 1$, respectively. However, since the event of singular random matrices has probability zero, if v is chosen randomly, almost all v will satisfy the conditions in Definitions 12.5 and 12.6, which will be shown in the following example.

Example 12.9. For $n_0 = 4$, $m_0 = 4$, $\lambda = 0.9$, $v = (0.5997, 0.9357, 0.9841, 1.4559)$ is generated randomly by Matlab, v is strongly 4 full rank since

$$\det(T([v_4, v_3, v_2, v_1])) = 0.4041, \quad \det(T([v_4^2, v_3^2, v_2^2, v_1^2])) = 2.4823,$$
$$\det(T([v_4^3, v_3^3, v_2^3, v_1^3])) = 7.7467, \quad \det(T([v_4^4, v_3^4, v_2^4, v_1^4])) = 19.8312.$$

Furthermore, for $\lambda = 0.9$,

$$\det(T(\lambda, [v_4, v_3, v_2, v_1])) = 0.3796, \quad \det(T(\lambda^2, [v_4^2, v_3^2, v_2^2, v_1^2])) = 1.7872,$$
$$\det(T(\lambda^3, [v_4^3, v_3^3, v_2^3, v_1^3])) = 4.2853, \quad \det(T(\lambda^4, [v_4^4, v_3^4, v_2^4, v_1^4])) = 8.5037.$$

v is generated randomly 10000 times, it is shown that all $T([v_4^i, v_3^i, v_2^i, v_1^i])$ and $T(\lambda^i, [v_4^i, v_3^i, v_2^i, v_1^i])$, $i = 1, \ldots, 4$, are nonsingular.

12.3 Estimates of ζ with Individual Thresholds

Based on strongly scaled full-rank signals, we now derive the estimation algorithms for ζ and analyze their convergence. To this end, estimation algorithms based on the information of individual thresholds are first investigated.

(A12.1) The noise $\{d_k\}$ is a sequence of i.i.d. random variables whose distribution function $F(\cdot)$ and its inverse $F^{-1}(\cdot)$ are twice continuously differentiable and known.

(A12.2) The prior information on $\theta = [a_0, \ldots, a_{n_0-1}]'$ and $\eta = [b_0, \ldots, b_{q_0-1}]'$ is that $\sum_{i=0}^{n_0-1} a_i \neq 0$, $b_{q_0} = 1$, $\eta \neq 0$, $\theta \in \Omega_\theta$, and $\eta \in \Omega_\eta$, where Ω_θ and Ω_η are known compact sets.

The input is a scaled $2n_0(q_0+1)$-periodic signal with one-period values

$$(v, v, \rho_1 v, \rho_1 v, \ldots, \rho_{q_0} v, \rho_{q_0} v),$$

where $v = (v_1, \ldots, v_{n_0})$ is strongly q_0 full rank.

By periodicity, $\Phi_l^j = \Phi^j$ for $j = 0, \ldots, n_0$, and Φ^j can be decomposed into $2(q_0+1)$ submatrices $\Phi^j(i)$, $i = 1, \ldots, 2(q_0+1)$, of dimension $n_0 \times n_0$: $\Phi^j = [(\Phi^j(1))', (\Phi^j(2))', \ldots, (\Phi^j(2(q_0+1)))']'$. Actually, for $k = 1, \ldots, 2(q_0+1)$,

$$\Phi^j(k) = \left[\ \phi_{kn_0}^j, \phi_{kn_0+1}^j, \ldots, \phi_{kn_0+n_0-1}^j\ \right]'.$$

Denote the $n_0 \times n_0$ circulant matrices

$$V^0 = T([1, \ldots, 1]), \quad \text{and} \quad V^j = T([v_{n_0}^j, \ldots, v_1^j]), \quad j = 1, \ldots, q_0.$$

Then, for $j = 0, \ldots, q_0$, the odd-indexed block matrices satisfy the simple scaling relationship

$$\Phi^j(1) = V^j, \quad \Phi^j(3) = \rho_1^j V^j, \quad \ldots, \quad \Phi^j(2q_0+1) = \rho_{q_0}^j V^j, \tag{12.6}$$

and the even-indexed block matrices are

$$\Phi^j(2l) = \rho_{l-1}^j T((\rho_l/\rho_{l-1})^j, [v_{n_0}, v_{n_0-1}, \ldots, v_1]), \quad l = 1, \ldots, q_0+1,$$

where $\rho_0 = \rho_{q_0+1} = 1$. Denote

$$\tau^{\{j\}} = [\tau^{\{j,1\}}, \ldots, \tau^{\{j,n_0\}}]' = V^j \theta,\ j = 0, \ldots, q_0. \tag{12.7}$$

Then, we have

$$\Phi^j(1)\theta = \tau^{\{j\}}, \quad \Phi^j(3)\theta = \rho_1^j \tau^{\{j\}}, \quad \ldots, \quad \Phi^j(2q_0+1)\theta = \rho_{q_0}^j \tau^{\{j\}}. \tag{12.8}$$

Let

$$\Psi_\theta = [\Phi^0\theta, \Phi^1\theta, \ldots, \Phi^{q_0}\theta].$$

Then, from (12.4), we have

$$Y_l = \Psi_\theta[\eta', 1]' + D_l = \zeta + D_l. \tag{12.9}$$

Remark 12.10. In $(v, v, \rho_1 v, \rho_1 v, \ldots, \rho_{q_0} v, \rho_{q_0} v)$, there are always two identical subsequences $\rho_i v, i = 1, \ldots, q_0$, appearing consecutively. The main reason for this input structure is to generate block matrices that satisfy the above scaling relationship (12.6).

For (12.9) and $i = 1, \ldots, m_0$, let

$$\begin{aligned} \mu_N^{\{i\}} &= [\mu_N^{\{i,1\}}, \ldots, \mu_N^{\{i,2n_0(q_0+1)\}}]' \\ &= \frac{1}{N}\sum_{k=1}^N S_k^{\{i\}} = \frac{1}{N}\sum_{k=1}^N I\{D_k \le C_i \mathbb{1}_{2n_0(q_0+1)} - \Psi_\theta[\eta', 1]'\}, \end{aligned}$$

which is the empirical distribution of D_l at

$$C_i \mathbb{1}_{2n_0(q_0+1)} - \zeta = C_i \mathbb{1}_{2n_0(q_0+1)} - \Psi_\theta[\eta', 1]'.$$

Then, by the strong law of large numbers,

$$\mu_N^{\{i\}} \to p^{\{i\}} = F(C_i \mathbb{1}_{2n_0(q_0+1)} - \Psi_\theta[\eta', 1]') \text{ w.p.1.}$$

Denote $S_N^{\{i\}} = [S_N^{\{i,1\}}, \ldots, S_N^{\{i,2n_0(q_0+1)\}}]'$, where $S_N^{\{i\}}$ is as defined in (12.2) and $S_N^{\{ij\}}$ denotes its jth component. By Assumption (A12.1), for each $i = 1, \ldots, m_0$, $\{S_k^{\{i\}}\}$ is an i.i.d. sequence. Since $j = 1, \ldots, 2n_0(q_0+1)$, $ES_k^{\{i,j\}} = p^{\{i,j\}} = F(C_i - \zeta_j)$ and

$$E(S_k^{\{i,j\}} - p^{\{i,j\}})^2 = p^{\{i,j\}}(1 - p^{\{i,j\}}) := \Delta_{i,j}^2.$$

Define $z_N^{\{ij\}} = \sum_{k=1}^N S_k^{\{ij\}}/N$. Then,

$$Ez_N^{\{i,j\}} = \frac{1}{N}\sum_{k=1}^N ES_k^{\{i,j\}} = p^{\{i,j\}},$$

$$E(\mu_N^{\{i,j\}} - p^{\{i,j\}})^2 = \frac{\Delta_{i,j}^2}{N}. \tag{12.10}$$

Note that F is a monotone function by Assumption (A12.1), and Ω_θ and Ω_η are bounded by Assumption (A12.2). Then, there exists $z > 0$ such that

$$z \le p^{\{i,j\}} = F(C_i - \zeta_j) \le 1 - z, \quad i = 1, \ldots, m_0, \quad j = 1, \ldots, 2n_0(q_0 + 1).$$

Since $F(\cdot)$ is not invertible at 0 and 1, we modify $\mu_N^{\{i,j\}}$ to avoid this "singularity." Let

$$\xi_N^{\{i,j\}} = \begin{cases} \mu_N^{\{i,j\}}, & \text{if} \quad z \le \mu_N^{\{i,j\}} \le 1 - z, \\ z, & \text{if} \quad \mu_N^{\{i,j\}} < z, \\ 1 - z, & \text{if} \quad \mu_N^{\{i,j\}} > 1 - z. \end{cases} \tag{12.11}$$

Since $\mu_N^{\{i,j\}} \to p^{\{i,j\}}$, w.p.1 and $z < p^{\{i,j\}} < 1 - z$, we have $\xi_N^{\{i,j\}} \to p^{\{i,j\}}$, w.p.1. Denote

$$\xi_N^{\{i\}} = [\xi_N^{\{i,1\}}, \ldots, \xi_N^{\{i,2n_0(q_0+1)\}}]'. \tag{12.12}$$

By Assumption (A12.1), F has a continuous inverse. Hence, for each $i = 1, \ldots, m_0$,

$$\begin{aligned} \zeta_N^{\{i\}} &= [\zeta_N^{\{i,1\}}, \ldots, \zeta_N^{\{i,2n_0(q_0+1)\}}]' \\ &:= C_i \mathbb{1}_{2n_0(q_0+1)} - F^{-1}(\xi_N^{\{i\}}) \\ &\to C_i \mathbb{1}_{2n_0(q_0+1)} - F^{-1}(p_i) = \Psi_\theta[\eta', 1]' \quad \text{as} \quad N \to \infty \\ &= \zeta = [\zeta_1, \ldots, \zeta_{2n_0(q_0+1)}]' \quad \text{w.p.1.} \end{aligned} \tag{12.13}$$

12.4 Quasi-Convex Combination Estimators of ζ

Since $\zeta_N^{\{i\}}$ is constructed from each individual threshold C_i, this enables us to treat the coefficients of the quasi-convex combination as design variables such that the resulting estimate has the minimal variance. This resulting estimate is exactly the optimal QCCE in Chapter 6.

For $j = 1, \ldots, 2n_0(q_0 + 1)$, define $\zeta_N(j) = [\zeta_N^{\{1,j\}}, \ldots, \zeta_N^{\{m_0,j\}}]'$ and $c_N(j) = [c_N(j,1), \ldots, c_N(j, m_0)]'$ with $c_N(j,1) + \cdots + c_N(j, m_0) = 1$. Construct an estimate of ζ_j by defining

$$\widehat{\zeta}_N(j) = c_N'(j)\zeta_N(j) = \sum_{k=1}^{m_0} c_N(j,k)\zeta_N^{\{k,j\}}.$$

Denote $c(j) = [c(j,1), \ldots, c(j, m_0)]'$ such that $c_N(j) \to c(j)$. Then $c(j,1) + \cdots + c(j, m_0) = 1$, and by (12.13),

$$\widehat{\zeta}_N(j) = \sum_{k=1}^{m_0} c_N(j,k)\zeta_N^{\{k,j\}} \to \zeta_j \sum_{k=1}^{m_0} c(j,k) = \zeta_j.$$

Denote the estimation errors

$$e_N(j) = \widehat{\zeta}_N(j) - \zeta_j,$$
$$\varepsilon_N(j) = \zeta_N(j) - \zeta_j \mathbb{1}_{m_0},$$

and their covariances

$$\sigma_N^2(j) = E e_N(j) e_N'(j), \quad Q_N(j) = E\varepsilon_N(j)\varepsilon_N'(j),$$

respectively. Then the covariance of estimation error is

$$\begin{aligned}\sigma_N^2(j) &:= E\left(\widehat{\zeta}_N(j) - \zeta_j\right)^2 = E\left(\sum_{k=1}^{m_0} c_N(j,k)(\zeta_N^{\{k,j\}} - \zeta_j)\right)^2 \\ &= c_N'(j) E\varepsilon_N(j)\varepsilon_N'(j) c_N(j) = c_N'(j) Q_N(j) c_N(j).\end{aligned}$$

That is, the variance is a quadratic form with respect to the variable $c(j)$. To obtain the quasi-convex combination estimate, we choose $c(j)$ to

$$\text{minimize } \sigma_N^2(j), \text{ subject to the constraint } c_N'(j)\mathbb{1}_{m_0} = 1.$$

Theorem 12.11. *Under Assumptions* (A12.1) *and* (A12.2)*, suppose* $u \in \mathcal{U}_{q_0}$ *and* $R_N(j) = NQ_N(j) = NE\varepsilon_N(j)\varepsilon_N'(j)$ *for* $j = 1, \ldots, 2n_0(q_0+1)$ *is positive definite. Then, the quasi-convex combination estimate can be obtained by choosing*

$$c_N^*(j) = \frac{R_N^{-1}(j)\mathbb{1}_{m_0}}{\mathbb{1}_{m_0}' R_N^{-1}(j)\mathbb{1}_{m_0}}, \quad \widehat{\zeta}_N(j) = \sum_{i=1}^{m_0} c^*(j,i)\zeta_N^{\{i,j\}}, \tag{12.14}$$

and the minimal variance satisfies

$$N\sigma_N^{2*}(j) = \frac{1}{\mathbb{1}_{m_0}' R_N^{-1}(j)\mathbb{1}_{m_0}}. \tag{12.15}$$

Consistency and Efficiency

From (12.14), $\widehat{\zeta}(j)$ can be regarded as an estimate of ζ_j. In this subsection, consistency and efficiency properties of this estimate will be analyzed.

By Assumption (A12.1), $G(x) = F^{-1}(x)$ is continuous on $(0,1)$. As a result, $G(x)$ is bounded on the compact set $[z, 1-z]$. Since $\zeta_N^{\{i,j\}} = C_i - G(\xi_N^{\{i,j\}}) \to \zeta^{\{i,j\}}$ w.p.1, we have $\zeta_N^{\{i,j\}} \to \zeta^{\{i,j\}}$ in probability. Furthermore, by the Lebesgue dominated convergence theorem [19, p. 100], $E\zeta_N^{\{i,j\}} \to \zeta^{\{i,j\}}$. Hence,

$$E\widehat{\zeta}_N(j) = E\sum_{k=1}^{m_0} c_N(j,k)\zeta_N^{\{k,j\}} \to \zeta_j \quad \text{as} \quad N \to \infty,$$

which means the estimate of ζ_j is asymptotically unbiased.

Subsequently, the efficiency of the estimate will be studied. To this end, the properties of $\xi_N^{\{i,j\}}$ in (12.11) will be introduced first.

Lemma 12.12. *Suppose* $u \in \mathcal{U}(n_0, q_0)$, *where* $\mathcal{U}(n_0, q_0)$ *is defined in Definition* 12.5. *Under Assumptions* (A12.1) *and* (A12.2), *there exist* $K_{i,j} \in (0, \infty)$ *and* $L_{i,j} \in (0, \infty)$, $i = 1, \ldots, m_0$, $j = 1, \ldots, 2n_0(q_0 + 1)$, *such that*

$$P\{\xi_N^{\{i,j\}} \neq \mu_N^{\{i,j\}}\} \leq K_{i,j} e^{-L_{i,j} N}. \tag{12.16}$$

Proof. Denote $X^{\{i,j\}} = (S_1^{\{i,j\}} - p^{\{i,j\}})/\Delta_{i,j}$. Note that $EX^{\{i,j\}} = 0$ and $E(X^{\{i,j\}})^2 = 1$. By the i.i.d. assumption, taking a Taylor expansion of $M_N^{\{i,j\}}(h) = [E \exp(hX^{\{i,j\}}/\sqrt{N})]^N$, the moment generating function of $\sqrt{N}(\mu_N^{\{i,j\}} - p^{\{i,j\}})/\Delta_{i,j}$, we obtain

$$\begin{aligned} M_N^{\{i,j\}}(h) &= \Big[E[1 + \frac{hX^{\{i,j\}}}{\sqrt{N}} + \frac{h^2 (X^{\{i,j\}})^2}{2N} + O(N^{-3/2})]\Big]^N \\ &= \Big[1 + \frac{h^2}{2N} + O(N^{-(3/2)})\Big]^N. \end{aligned}$$

Consequently, for any $t \in \mathbb{R}$,

$$\inf_h e^{-ht} M_N^{\{i,j\}}(h) = \inf_h e^{-ht}\Big[1 + \frac{h^2}{2N} + O(N^{-(3/2)})\Big]^N \leq K e^{-\frac{t^2}{2}}, \tag{12.17}$$

where $K > 0$ is a positive constant.

By means of the Chernoff bound [83, p. 326], for any $t \in (-\infty, p^{\{i,j\}}]$,

$$\begin{aligned} P\left\{\mu_N^{\{i,j\}} \leq t\right\} &= P\left\{\sum_{k=1}^{N} (S_k^{\{i,j\}} - p_{i,j}) \leq N \frac{(t - p_{i,j})}{\Delta_{i,j}}\right\} \\ &\leq \left\{\inf_h \left[e^{-\frac{h(t-p_{i,j})}{\Delta_{i,j}}} M_N^{\{i,j\}}(h)\right]\right\}^N \end{aligned} \tag{12.18}$$

and for any $p_{i,j} \leq t < \infty$,

$$P\{\mu_N^{\{i,j\}} \geq t\} \leq \left\{\inf_h \left[e^{-\frac{h(t-p_{i,j})}{\Delta_{i,j}}} M_N^{\{i,j\}}(h)\right]\right\}^N. \tag{12.19}$$

Considering

$$P\{\xi_{i,j}(N) \neq \mu_N^{\{i,j\}}\} = P(\mu_N^{\{i,j\}} \leq z) + P(\mu_N^{\{i,j\}} \geq 1 - z)$$

and (12.17)–(12.19), (12.16) is true. □

Theorem 12.13. *Under the conditions of Lemma* 12.12, *we have*

$$NE(\xi_N^{\{i,j\}} - p^{\{i,j\}})^2 \to \Delta_{i,j}^2 \quad as \quad N \to \infty, \tag{12.20}$$

and

$$NE|(\xi_N^{\{i,j\}} - p^{\{i,j\}})|^{q_0} \to 0 \quad as \quad N \to \infty, \; q_0 = 3, 4, \ldots \tag{12.21}$$

Proof. (i) By Theorem 12.12, there exist $K_{i,j} \in (0,\infty)$ and $L_{i,j} \in (0,\infty)$ such that

$$\begin{aligned} EN(\xi_N^{\{i,j\}} - \mu_N^{\{i,j\}})^2 &\le NzP\{\xi_N^{\{i,j\}} \ne \mu_N^{\{i,j\}}\} \\ &\le zK_{i,j}Ne^{-L_{i,j}N} \to 0. \end{aligned}$$

This together with

$$\begin{aligned} &EN(\mu_N^{\{i,j\}} - p^{\{i,j\}})(\xi_N^{\{i,j\}} - \mu_N^{\{i,j\}}) \\ &\quad \le \sqrt{EN(\mu_N^{\{i,j\}} - p^{\{i,j\}})^2 EN(\xi_N^{\{i,j\}} - \mu_N^{\{i,j\}})^2} \\ &\quad = \Delta_{i,j}\sqrt{EN(\xi_N^{\{i,j\}} - \mu_N^{\{i,j\}})^2} \end{aligned}$$

implies that

$$\begin{aligned} &EN(\xi_N^{\{i,j\}} - p^{\{i,j\}})^2 - EN(\mu_N^{\{i,j\}} - p^{\{i,j\}})^2 \\ &\quad = 2EN(\mu_N^{\{i,j\}} - p^{\{i,j\}})(\xi_N^{\{i,j\}} - \mu_N^{\{i,j\}}) \\ &\qquad + EN(\xi_N^{\{i,j\}} - \mu_N^{\{i,j\}})^2 \\ &\quad \to 0 \quad \text{as} \quad N \to \infty. \end{aligned} \tag{12.22}$$

Thus, by (12.10), we obtain (12.20).

(ii) Similarly, for $q_0 = 3, 4, \ldots$, one obtains

$$NE|(\xi_N^{\{i,j\}} - p^{\{i,j\}})|^{q_0} - NE|(\mu_N^{\{i,j\}} - p^{\{i,j\}})|^{q_0} \to 0.$$

By Hölder's inequality,

$$NE|\mu_N^{\{i,j\}} - p^{\{i,j\}}|^{q_0} \le \Delta_{i,j}\sqrt{NE(\mu_N^{\{i,j\}} - p^{\{i,j\}})^{2(q_0-1)}}. \tag{12.23}$$

Notice that for each i, j, $S_k^{\{i,j\}}$ is i.i.d. Then, we have

$$\begin{aligned} NE(\mu_N^{\{i,j\}} - p^{\{i,j\}})^{2(q_0-1)} &= NE\Big[\frac{1}{N}\sum_{k=1}^{N}(S_k^{\{i,j\}} - p^{\{i,j\}})\Big]^{2(q_0-1)} \\ &= N^{-2(m_0-2)}E(S_1^{\{i,j\}} - p^{\{i,j\}})^{2(q_0-1)} \\ &\le N^{-2(q_0-2)}, \end{aligned}$$

which together with (12.23) results in

$$NE|\mu_N^{\{i,j\}} - p^{\{i,j\}}|^{q_0} \le \Delta_{i,j}N^{-(q_0-2)} \to 0.$$

Hence, (12.21) is obtained. □

From (12.15), the covariance of the estimation $\widehat{\zeta}_N(j)$ is decided by $R_N(j)$.

Theorem 12.14. *Suppose $u \in \mathcal{U}(n_0, q_0)$. If, in addition to Assumptions* (A12.1) *and* (A12.2), *the density function $f(x)$ is continuously differentiable, then as $N \to \infty$,*

$$R_N(j) := NQ_N(j) = NE\varepsilon_N(j)\varepsilon'_N(j) \to \Lambda(j)W(j)\Lambda(j) := R(j), \quad (12.24)$$

where

$$\varepsilon_N(j) = \zeta_N(j) - \zeta_j \mathbb{1}_{m_0},$$
$$\Lambda(j) = \operatorname{diag}^{-1}\{f(C_1 - \zeta_j), \ldots, f(C_{m_0} - \zeta_j)\},$$

and

$$W(j) = \begin{bmatrix} p^{\{1,j\}}(1-p^{\{1,j\}}) & \ldots & p^{\{1,j\}}(1-p^{\{m_0,j\}}) \\ \vdots & \ddots & \vdots \\ p^{\{1,j\}}(1-p^{\{m_0,j\}}) & \ldots & p^{\{m_0,j\}}(1-p^{\{m_0,j\}}) \end{bmatrix}. \quad (12.25)$$

Proof. Denote $\varepsilon_N(j,i)$ as the ith component of $\varepsilon_N(j)$, $\dot{G}(x) = dG(x)/dx$, and $\ddot{G}(x) = d\dot{G}(x)/dx$. Then

$$\begin{aligned} \dot{G}(x) &= \frac{dG(x)}{dx} = \frac{dG(x)}{dF(G(x))} = \frac{1}{f(G(x))}, \\ \ddot{G}(x) &= \frac{d\dot{G}(x)}{dx} = -\frac{1}{f^2(G(x))}\dot{f}(G(x))\dot{G}(x). \end{aligned}$$

Since $\dot{f}(x)$ is continuous, by Assumption (A12.1), both $\dot{G}(x)$ and $\ddot{G}(x)$ are continuous, and hence bounded in $[z, 1-z]$. Let

$$\beta^{\{i,j\}} = \sup_{x\in[z,1-z]} \{|\dot{G}(x)|\} \text{ and } \gamma^{\{i,j\}} = \sup_{x\in[z,1-z]} \{|\ddot{G}(x)|\}.$$

Then, there exists a number $\lambda_N^{\{i,j\}}$ between $p^{\{i,j\}}$ and $\xi_N^{\{i,j\}}$ such that

$$\begin{aligned} \varepsilon_N(j,i) &= \zeta_N^{\{i,j\}} - \zeta_j = G(\xi_N^{\{i,j\}}) - G(p^{\{i,j\}}) \\ &= \dot{G}(p^{\{i,j\}})(\xi_N^{\{i,j\}} - p^{\{i,j\}}) + \frac{1}{2}\ddot{G}(\lambda_N^{\{i,j\}})(\xi_N^{\{i,j\}} - p^{\{i,j\}})^2. \end{aligned}$$

This implies that for $i, k = 1, \ldots, m_0$,

$$\begin{aligned} &NE\varepsilon_N(j,i)\varepsilon_N(j,k) \\ &= NE(\zeta_N^{\{i,j\}} - \zeta_j)(\zeta_N^{\{k,j\}} - \zeta_j) \\ &= N\dot{G}(p^{\{i,j\}})\dot{G}(p^{\{k,j\}})E(\xi_N^{\{i,j\}} - p^{\{i,j\}})(\xi_N^{\{k,j\}} - p^{\{k,j\}}) \\ &\quad + NE\dot{G}(p^{\{i,j\}})(\xi_N^{\{i,j\}} - p^{\{i,j\}})(\xi_N^{\{k,j\}} - p^{\{k,j\}})^2\ddot{G}(\lambda_N^{\{k,j\}}) \\ &\quad + NE\ddot{G}(\lambda_N^{\{i,j\}})(\xi_N^{\{i,j\}} - p^{\{i,j\}})^2(\xi_N^{\{k,j\}} - p^{\{k,j\}})\dot{G}(p^{\{k,j\}}) \\ &\quad + NE\ddot{G}(\lambda_N^{\{i,j\}})(\xi_N^{\{i,j\}} - p^{\{i,j\}})^2(\xi_N^{\{k,j\}} - p^{\{k,j\}})^2\dot{G}(\lambda_N^{\{k,j\}}). \end{aligned} \quad (12.26)$$

By Hölder's inequality and Theorem 12.13, we have

$$\begin{aligned}
&|NE\dot{G}(p^{\{i,j\}})(\xi_N^{\{i,j\}} - p^{\{i,j\}})(\xi_N^{\{k,j\}} - p^{\{k,j\}})^2 \ddot{G}(\lambda_N^{\{k,j\}})| \\
&\quad \le \beta^{\{i,j\}}\gamma^{\{i,j\}}\sqrt{NE(\xi_N^{\{i,j\}} - p^{\{i,j\}})^2 NE(\xi_N^{\{k,j\}} - p^{\{k,j\}})^4|} \\
&\quad \le \beta^{\{i,j\}}\gamma^{\{i,j\}}\Delta_{i,j}\sqrt{NE(\xi_N^{\{k,j\}} - p^{\{k,j\}})^4|} \to 0.
\end{aligned} \tag{12.27}$$

Similarly,

$$|NE\ddot{G}(\lambda_N^{\{i,j\}})(\xi_N^{\{i,j\}} - p^{\{i,j\}})^2(\xi_N^{\{k,j\}} - p^{\{k,j\}})\dot{G}(p^{\{k,j\}})| \to 0, \tag{12.28}$$

$$|NE\ddot{G}(\lambda_N^{\{i,j\}})(\xi_N^{\{i,j\}} - p^{\{i,j\}})^2(\xi_N^{\{k,j\}} - p^{\{k,j\}})^2\dot{G}(\lambda_N^{\{k,j\}})| \to 0. \tag{12.29}$$

Thus, similarly to (12.22), we have

$$\begin{aligned}
N\Big[&E(\xi_N^{\{i,j\}} - p^{\{i,j\}})(\xi_N^{\{k,j\}} - p^{\{k,j\}}) \\
&- E(\mu_N^{\{i,j\}} - p^{\{i,j\}})(\mu_N^{\{k,j\}} - p^{\{k,j\}})\Big] \to 0.
\end{aligned} \tag{12.30}$$

Since d_k, $k = 1, 2, \ldots$, are i.i.d.,

$$\begin{aligned}
&NE(\mu_N^{\{i,j\}} - p^{\{i,j\}})(\mu_N^{\{k,j\}} - p^{\{k,j\}}) \\
&= \frac{1}{N}E\Big[\Big(\sum_{l_1=1}^{N} I\{d_{l_1} \le p^{\{i,j\}}\} - p^{\{i,j\}})(\sum_{l_2=1}^{N} I\{d_{l_2} \le p^{\{k,j\}}\} - p^{\{k,j\}}\Big)\Big] \\
&= \frac{1}{N}E\sum_{l_1=1}^{N} I\{d_{l_1} \le p^{\{i,j\}}\}I\{d_{l_1} \le p^{\{k,j\}}\} - p^{\{i,j\}}p^{\{k,j\}} \\
&= p^{\{\min\{i,k\},j\}} - p^{\{i,j\}}p^{\{k,j\}}
\end{aligned} \tag{12.31}$$

and

$$\dot{G}(p^{\{i,j\}}) = \frac{1}{f(G(p^{\{i,j\}}))} = \frac{1}{f(C_i - \zeta_j)}. \tag{12.32}$$

Therefore, (12.24) follows from (12.26)–(12.32). □

Proposition 12.15. *$R(j)$, $j = 1, \ldots, 2n_0(q_0 + 1)$, defined by (12.24), is positive definite, and*

$$1\!\!1'_{m_0} R^{-1}(j) 1\!\!1_{m_0} = \sum_{k=1}^{m_0+1} \frac{h^2(j,k)}{\widetilde{p}^{\{k,j\}}}, \tag{12.33}$$

where

$$\widetilde{p}^{\{i,j\}} = F(C_i - \zeta_j) - F(C_{i-1} - \zeta_j),$$

$$h(j,i) = f(C_{i-1} - \zeta_j) - f(C_i - \zeta_j),$$

with $C_0 = -\infty$ and $C_{l+1} = \infty$.

Proof. Since

$$R_N(j) = NE\varepsilon_N(j)\varepsilon'_N(j) \geq 0,$$

so is $R(j)$. Noting

$$R(j) = \Lambda(j)W(j)\Lambda(j),$$
$$\Lambda(j) = \mathrm{diag}^{-1}\{f(C_1 - \zeta_j), \ldots, f(C_{m_0} - \zeta_j)\},$$

and $f(C_i - \zeta_j) > 0$, $i = 1, \ldots, m_0$, we need only to show that $W(j)$ is positive definite.

From (12.25),

$$\begin{aligned}
\det(W(j)) &= \begin{vmatrix} p^{\{1,j\}}(1-p^{\{1,j\}}) & \ldots & p^{\{1,j\}}(1-p^{\{m_0,j\}}) \\ \vdots & \ddots & \vdots \\ p^{\{1,j\}}(1-p^{\{m_0,j\}}) & \ldots & p^{\{m_0,j\}}(1-p^{\{m_0,j\}}) \end{vmatrix} \\
&= p^{\{1,j\}} \begin{vmatrix} 1-p^{\{1,j\}} & p^{\{1,j\}}-p^{\{2,j\}} & \ldots & p^{\{1,j\}}-p^{\{m_0,j\}} \\ 1-p^{\{2,j\}} & 0 & & p^{\{2,j\}}-p^{\{m_0,j\}} \\ \vdots & & \ddots & \vdots \\ 1-p^{\{m_0,j\}} & 0 & \cdots & 0 \end{vmatrix} \\
&= p^{\{1,j\}}(p^{\{1,j\}}-p^{\{2,j\}})\ldots(p^{\{m_0,j\}}-p^{\{j\}}_{m_0-1})(1-p^{\{m_0,j\}}) \neq 0.
\end{aligned}$$

Thus, $R(j) > 0$. Furthermore, by Lemma 6.4,

$$\mathbb{1}'R^{-1}(j)\mathbb{1} = \sum_{k=1}^{m_0+1} \frac{h^2(j,k)}{\widetilde{p}^{\{k,j\}}}.$$

Thus, (12.33) is also true. □

Lemma 12.16. *The Cramér–Rao lower bound for estimating ζ_j based on $\{s_k\}$ is*

$$\sigma^2_{\mathrm{CR}}(N,j) = \left(N \sum_{j=1}^{m_0+1} \frac{h^2(j,i)}{\widetilde{p}^{\{i,j\}}} \right)^{-1}.$$

Next, we demonstrate that the aforementioned algorithms are asymptotically efficient based on the following theorem.

Theorem 12.17. *Under the conditions of Theorem* 12.14*, for* $j = 1, \ldots,$ $2n_0(q_0+1)$,

$$\lim_{N\to\infty} N\left(\sigma^{2*}_N(j) - \sigma^2_{\mathrm{CR}}(N,j)\right) = 0 \quad as \quad N \to \infty.$$

Proof. This theorem can be proved directly by Theorem 12.14, Proposition 12.15, and Lemma 12.16.

Recursive Quasi-Convex Combination Estimates

Since $\sigma_N^2(j) = E\varepsilon_N(j)\varepsilon_j'(N)$ contains an unknown parameter ζ_j, it cannot be directly computed. As a result, the quasi-convex combination estimate $\zeta_N(j)$ in (12.14) cannot be computed. In this section, we will derive computable estimates. The basic idea is to employ a recursive structure in which the unknown ζ_j is replaced by the current estimate $\widehat{\zeta}_N(j)$. Convergence of the algorithms will be established.

For $i = 1, \ldots, m_0$ and $j = 1, \ldots, 2n_0(q_0+1)$, let $\xi_0(i) = 0_{2n_0(q_0+1)}$, $\widehat{c}_0(j) = 0_{q_0}$, $\widehat{R}_0(j) = 0_{q_0 \times q_0}$, and $\widehat{\zeta}_0(j) = 0_{2n_0(q_0+1)}$. Suppose that at step $N-1$ ($N \geq 1$), $\xi_{N-1}(i)$, $c_{N-1}(j)$, and $\widehat{R}_{N-1}(j)$ have been obtained. Then the estimation algorithms can be constructed as follows.

(i) Calculate the sample distribution values

$$\xi_N^{\{i\}} = \frac{1}{N} S_N^{\{i\}} + \frac{N-1}{N} \xi_{N-1}^{\{i\}}.$$

(ii) Calculate the data points

$$\zeta_N^{\{i\}} = F^{-1}(\xi_N^{\{i\}}).$$

Let

$$\zeta_N(j) = [\zeta_N^{\{1,j\}}, \ldots, \zeta_N^{\{q_0,j\}}]', \ j = 1, \ldots, 2n_0(q_0+1).$$

(iii) Calculate each covariance estimate $R_N(j)$.

Let

$$p_N^{\{i,j\}} = F(C_i - \zeta_{N-1}^{\{i,j\}}),$$
$$\widehat{\Lambda}_N(j) = \operatorname{diag}^{-1}\{f(p_N^{\{1,j\}}), \ldots, f(p_N^{\{1,m_0\}})\},$$
$$W_N(j) = \begin{bmatrix} p_N^{\{1,j\}}(1-p_N^{\{1,j\}}) & \ldots & p_N^{\{1,j\}}(1-p_N^{\{1,m_0\}}) \\ \vdots & \ddots & \vdots \\ p_N^{\{1,j\}}(1-p_N^{\{1,m_0\}}) & \ldots & p_N^{\{1,m_0\}}(1-p_N^{\{1,m_0\}}) \end{bmatrix}.$$

Calculate $R_N(j)$ by

$$\widehat{R}_N(j) = \widehat{\Lambda}_N(j) W_N(j) \widehat{\Lambda}_N(j).$$

(iv) If $\widehat{R}_N(j)$ is nonsingular, then let

$$\widehat{c}_N(j) = \frac{\widehat{R}_j^{-1}(N)\mathbb{1}}{\mathbb{1}'\widehat{R}_N^{-1}(j)\mathbb{1}},$$

and compute

$$\widehat{\zeta}_N^{\{j\}} = \widehat{c}'_N(j)([C_1, \ldots, C_{m_0}]' - \zeta_N(j)).$$

Otherwise, $\widehat{\zeta}_N^{\{j\}} = \widehat{\zeta}_{N-1}^{\{j\}}$.

(v) Let $\widehat{\zeta}_N = [\widehat{\zeta}_N^{\{1\}}, \ldots, \widehat{\zeta}_N^{\{2n_0(q_0+1)\}}]'$. Go to step 1.

This algorithm depends only on sample paths. At each step, it minimizes the estimation variance based on the most recent information on the unknown parameter. In addition, the following asymptotic properties hold.

Theorem 12.18. *Under the conditions of Theorem* 12.14*, for* $j = 1, \ldots, 2n_0(q_0+1)$*, the above recursive algorithms have the following properties:*

$$\lim_{N\to\infty} \widehat{\zeta}_N(j) = \zeta_j \quad \textit{w.p.1}, \tag{12.34}$$

$$\lim_{N\to\infty} \widehat{R}_N(j) = R(j) \quad \textit{w.p.1}, \tag{12.35}$$

$$\lim_{N\to\infty} NE(\widehat{\zeta}_N(j) - \zeta_j)^2 = \frac{1}{\mathbb{1}' R^{-1}(j)\mathbb{1}} \quad \textit{w.p.1}. \tag{12.36}$$

Proof. Note that $\xi_N^{\{i\}} \to F(C_i \mathbb{1}_{2n_0(q_0+1)} - \zeta)$ w.p.1 and the convergence is uniform in $C_i \mathbb{1}_{2n_0(q_0+1)} - \zeta$. Since $F(\cdot)$ and $F^{-1}(\cdot)$ are both continuous,

$$\begin{aligned} \zeta_N^{\{i\}} &= C_i \mathbb{1}_{2n_0(q_0+1)} - F^{-1}(\xi_N^{\{i\}}) \\ &\to C_i \mathbb{1}_{2n_0(q_0+1)} - F^{-1}(F(C_i \mathbb{1}_{2n_0(q_0+1)} - \zeta) = \zeta \end{aligned}$$

w.p.1 as $N \to \infty$. Thus, the quasi-convex combination $\widehat{\zeta}_N(j)$ converges to ζ w.p.1. That is, (12.34) holds.

By Assumption (A12.1), $F(\cdot)$ and $f(\cdot)$ are both continuous. Hence,

$$\widehat{\Lambda}_N(j) \to \Lambda(j) \quad \text{and} \quad W_N(j) \to W_j.$$

As a result, (12.35) holds, and by (12.15),

$$E(\widehat{\zeta}_N(j) - \zeta_j)^2 = \frac{1}{\mathbb{1}'_{m_0} \widehat{R}_N^{-1}(j) \mathbb{1}_{m_0}} \to \frac{1}{\mathbb{1}'_{m_0} R^{-1}(j) \mathbb{1}_{m_0}},$$

which results in (12.36). □

12.5 Estimation of System Parameters

Identification algorithms of the system parameters will be constructed based on the estimate of ζ. The parameters of the linear part are first estimated, then the nonlinearity is identified.

Identifiability of the Unknown Parameters

Theorem 12.19. *Suppose* $u \in \mathcal{U}(n_0, q_0)$. *Then,*

$$\Psi_\theta[\eta', 1]' = \zeta$$

has a unique solution (θ^*, η^*).

Proof. (i) To obtain θ^*.

By the first component of (12.13), we have $\zeta = [\zeta_1, \ldots, \zeta_{2n_0(q_0+1)}]'$, and

$$b_0\tau^{\{0,1\}} + b_1\tau^{\{1,1\}} + \cdots + b_{q_0}\tau^{\{q_0,1\}} = \zeta_1.$$

From (12.8), the $2in_0 + 1$ ($i = 1, \ldots, q_0$) component of (12.13) turns out to be

$$b_0\tau^{\{0,1\}} + \rho_i b_1\tau^{\{1,1\}} + \cdots + \rho_i^{q_0} b_{q_0}\tau^{\{q_0,1\}} = \zeta_{2in_0+1},$$

or equivalently,

$$\Re \begin{bmatrix} b_0\tau^{\{0,1\}} \\ b_1\tau^{\{1,1\}} \\ \vdots \\ b_{q_0}\tau^{\{q_0,1\}} \end{bmatrix} = \begin{bmatrix} \zeta_1 \\ \zeta_{2n_0+1} \\ \vdots \\ \zeta_{2q_0n_0+1} \end{bmatrix}, \quad \text{where} \quad \Re = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 1 & \rho_1 & \ldots & \rho_1^{q_0} \\ \vdots & \ldots & \ldots & \vdots \\ 1 & \rho_{q_0} & \ldots & \rho_{q_0}^{q_0} \end{bmatrix}.$$

Since $\rho_j \neq 0$, $\rho_j \neq 1$, $j = 1, \ldots, q_0$, and $\rho_i \neq \rho_j$, the determinant of the Vandermonde matrix

$$\det \Re = \prod_{0 \le i < j \le q_0 - 1} (\rho_j - \rho_i) \neq 0 \quad \text{with} \quad \rho_0 = 1.$$

Hence, $b_j\tau^{\{j,1\}}$, $j = 0, \ldots, q_0$, can be solved by

$$\begin{bmatrix} b_0\tau^{\{0,1\}} \\ b_1\tau^{\{1,1\}} \\ \vdots \\ b_{q_0}\tau^{\{q_0,1\}} \end{bmatrix} = \Re^{-1} \begin{bmatrix} \zeta_1 \\ \zeta_{2n_0+1} \\ \vdots \\ \zeta_{2q_0n_0+1} \end{bmatrix}.$$

Similarly, we have

$$\Gamma = \Re^{-1}\Xi, \tag{12.37}$$

where

$$\Gamma = \begin{bmatrix} b_0\tau^{\{0,1\}} & b_0\tau^{\{0,2\}} & \dots & b_0\tau^{\{0,n_0\}} \\ b_1\tau^{\{1,1\}} & b_1\tau^{\{1,2\}} & \dots & b_1\tau^{\{1,n_0\}} \\ & & \vdots & \\ b_{q_0}\tau^{\{q_0,1\}} & b_{q_0}\tau^{\{q_0,2\}} & \dots & b_{q_0}\tau^{\{q_0,n_0\}} \end{bmatrix},$$

$$\Xi = \begin{bmatrix} \zeta_1 & \zeta_2 & \dots & \zeta_{n_0} \\ \zeta_{2n_0+1} & \zeta_{2n_0+2} & \dots & \zeta_{3n_0} \\ & & \vdots & \\ \zeta_{2q_0n_0+1} & \zeta_{2q_0n_0+2} & \dots & \zeta_{2q_0+1)n_0} \end{bmatrix}.$$

Denote $r(i)$ as the ith column of $(\Re^{-1})'$. Then, by $b_{q_0} = 1$, we have

$$\tau^{\{q_0\}} = [\tau^{\{q_0,1\}}, \dots, \tau^{\{q_0,n_0\}}]' = \Xi' r(q_0).$$

Note that $u \in \mathcal{U}(n_0, q_0)$ implies that V^{q_0} is full rank. Then, by (12.7), one can get $\theta^* = V_{q_0}^{-1}\tau^{\{q_0\}}$.

(ii) To obtain η^*.

By Assumption (A12.2), $\sum_{i=0}^{n_0-1} a_i \neq 0$, or $V^0\theta \neq \mathbf{0}_{n_0}$. For $u \in \mathcal{U}(n_0, q_0)$ and $j = 1, \dots, q_0$, $V^j = T([v_{n_0}^j, \dots, v_1^j])$ is full rank by Definition 12.1, and so $V^j\theta \neq \mathbf{0}_{n_0}$. Thus, for each $j = 0, \dots, q_0$, $\tau^{\{j\}} = V^j\theta$ has a nonzero component $\tau^{\{j,i_N^*(j)\}}$. For any given positive integer k and $j = 1, \dots, k$, let $\beta_j(k)$ be a k-dimensional vector with all components being zero except the jth being 1, that is,

$$\beta_j(k) = [\underbrace{0, \dots, 0}_{j-1}, 1, \underbrace{0, \dots, 0}_{k-j}]'.$$

Then, from (12.37), we have

$$b_j\tau^{\{j,i_N^*(j)\}} = \beta_j'(m_0+1)\Re^{-1}\Xi\beta_{i^*(j)}(n_0), \quad j = 0, \dots, q_0,$$

which gives b_j, $j = 0, \dots, q_0$, since $\tau^{\{j,i_N^*(j)\}}$ can be calculated from V^j and θ^* via (12.7). Thus, η^* is obtained. □

A particular choice of the scaling factors ρ_j is $\rho_j = \lambda^j$, $j = 0, 1, \dots, q_0$, for some $\lambda \neq 0$ and $\lambda \neq 1$. In this case, the period of input u can be shortened to $n_0(q_0+2)$ under a slightly different condition.

Identification Algorithms and Convergence Properties

The $\zeta_N = [\zeta_N^{\{1\}}, \dots, \zeta_N^{\{2n_0(q_0+1)-1\}}]'$ in (12.12) has $2n_0(q_0+1)$ components for a strongly scaled q_0 full-rank signal $u \in \mathcal{U}(n_0, q_0)$.

Let

$$V_{q_0} = T([v_{n_0}^{q_0}, \ldots, v_1^{q_0}]), \qquad [r_1, \ldots, r(q_0)] := (\Re')^{-1},$$

$$\Xi_N = \begin{bmatrix} \zeta_N^{\{1\}} & \zeta_N^{\{2\}} & \cdots & \zeta_N^{\{n_0\}} \\ \zeta_N^{\{2n_0+1\}} & \zeta_N^{\{2n_0+2\}} & \cdots & \zeta_N^{\{3n_0\}} \\ & & \vdots & \\ \zeta_N^{\{2q_0n_0+1\}} & \zeta_N^{\{2q_0n_0+2\}} & \cdots & \zeta_N^{\{(2q_0+1)n_0\}} \end{bmatrix}.$$

Then, we have the following identification algorithm:

(i) *Estimate* θ*.* The estimate of θ is taken as

$$\theta_N = V_{q_0}^{-1} \Xi_N' r(q_0). \tag{12.38}$$

(ii) *Estimate* η. Let $b_0(j) = 0$ and

$$b_N(j) = \begin{cases} [\zeta_N^{\{i_N^*(j)\}}, \ldots, \zeta_N^{\{2q_0n_0+i_N^*(j)\}}] r_N(i_N^*(j))/\tau_N^{\{j,i_N^*(j)\}}, & \\ & \text{if } \tau^{\{j,i_N^*(j)\}} \neq 0, \\ b_{N-1}(j), & \text{if } \tau^{\{j,i_N^*(j)\}} = 0, \end{cases}$$

where

$$i_N^*(j) = \min\{\text{argmax}_{1 \le i \le n_0} |\tau^{\{j,i\}}|\}, \quad j = 0, 1, \ldots, q_0 - 1; \tag{12.39}$$

$r(i_N^*(j))$ is the $i_N^*(j)$th column of $(\Re')^{-1}$, and $\tau^{\{j,i_N^*(j)\}}$ is the $i_N^*(j)$-th component of $\tau_N^{\{j\}} = V^j \theta_N$. Then, the estimate of η is taken as

$$\eta_N = [b_N(0), \ldots, b_N(q_0 - 1)]'. \tag{12.40}$$

Theorem 12.20. *Suppose* $u \in \mathcal{U}(n_0, q_0)$*. Then, under Assumptions* (A12.1) *and* (A12.2)*,*

$$\theta_N \to \theta \quad \textit{and} \quad \eta_N \to \eta \quad \textit{w.p.1 as} \quad N \to \infty.$$

Proof. By (12.13), $\zeta_N \to \zeta$ w.p.1. as $N \to \infty$. So,

$$\theta_N = V_{q_0}^{-1} \Xi_N' r_{q_0} \to V_{q_0}^{-1} \Xi' r(q_0) = \theta,$$

which in turn leads to

$$\tau_N^{\{j\}} = [\tau_N^{\{j,1\}}, \ldots, \tau_N^{\{j\}}(n_0)]' := V^j \theta_N \to V^j \theta = \tau^{\{j\}} \quad \text{w.p.1.}$$

and $\tau^{\{j,i\}} \to \tau^j(i)$ w.p.1 for $i = 1, \ldots, n_0$. Thus, for $j = 0, \ldots, q_0 - 1$, we have

$$i_N^*(j) = \min\{\text{argmax}_{1 \le i \le n_0} |\tau^{\{j,i\}}|\} \to \min\{\text{argmax}_{1 \le i \le n_0} |\tau^j(i)|\} := i^*(j),$$

and

$$\tau_N^{\{j,i_N^*(j)\}} \to \tau^{\{j,i_N^*(j)\}} \neq 0.$$

This means that with probability 1, there exists $N_0 > 0$ such that

$$\tau_N^{\{j,i_N^*(j)\}} \neq 0, \ \forall N \geq N_0.$$

Let

$$b_N(j) = \frac{r(i_N^*(j))}{\tau_N^{\{j,i_N^*(j)\}}}[\zeta_N(i_N^*(j)), \zeta_N(2n_0 + i_N^*(j)), \ldots, \zeta_N(2q_0n_0 + i_N^*(j)].$$

Then by

$$\begin{aligned} b_N(j)\tau_N^{\{j,i_N^*(j)\}} &\to [\zeta_{i^*(j)}, \zeta_{2n_0+i^*(j)}, \ldots, \zeta_{2q_0n_0+i^*(j)}]r(i^*(j)) \\ &= b_j\tau^{\{j,i_N^*(j)\}}, \end{aligned}$$

we have $b_N(j) \to b_j$ for $j = 0, \ldots, q_0 - 1$. Hence, $\eta_N \to \eta$ w.p.1 as $N \to \infty$. □

Algorithms under Exponentially Scaled Inputs

Let u be $n_0(q_0+2)$-periodic with one-period values $(v, \lambda v, \ldots, \lambda^{q_0} v, \lambda^{q_0+1} v)$. The $\overline{\zeta}_N = [\overline{\zeta}_N^{\{1\}}, \ldots, \overline{\zeta}_N^{\{q_0+1\}}]'$ can be estimated by the algorithms in Section 12.4 with dimension changed from $2n_0(q_0+1)$ to $n_0(q_0+2)$, and

$$\overline{\zeta}_N \to \overline{\zeta} = \sum_{j=0}^{q_0} b_j \overline{\Phi}^j \theta.$$

Partition $\overline{\Phi}^j$ into (q_0+2) submatrices $\overline{\Phi}^j(i)$, $i = 1, \ldots, q_0+2$, of dimension $n_0 \times n_0$:

$$\overline{\Phi}^j = [(\overline{\Phi}^j(1))', (\overline{\Phi}^j(2))', \ldots, (\overline{\Phi}^j(q_0+2))']'.$$

If $u \in \mathcal{U}_\lambda(n_0, q_0)$, then it can be directly verified that

$$\begin{aligned} \overline{\Phi}^j(l+1) &= \lambda^{jl}\overline{\tau}^{\{j\}}, \quad l = 0, 1, \ldots, q_0, \\ \overline{\Phi}^j(q_0+2) &= \lambda^{j(q_0+2)} T(\lambda^{-j(q_0+2)}, [v_{n_0}, \ldots, v_1]), \end{aligned}$$

where $\overline{\tau}^{\{j\}} = T(\lambda^j, [v_{n_0}^j, \ldots, v_1^j])$. With these notations, we have the following result, whose proof is similar to that of Theorem 12.19, and hence, is omitted.

Theorem 12.21. *Suppose $u \in \mathcal{U}_\lambda(n_0, q_0)$. Then, under Assumptions* (A12.1) *and* (A12.2),

$$\overline{\Psi}_\theta[\eta', 1]' = \overline{\zeta}$$

has a unique solution (θ^*, η^*), where

$$\overline{\Phi}_\theta = [\overline{\Phi}(0)\theta, \overline{\Phi}(1)\theta, \ldots, \overline{\Phi}(q_0)\theta].$$

Let

$$\overline{\zeta}_N = [\overline{\zeta}_N^{\{1\}}, \dots, \overline{\zeta}_N^{\{n_0(q_0+1)\}}]',$$

and

$$\overline{V}^{q_0} = T(\lambda^{q_0}, [v_{n_0}^{q_0}, \dots, v_1^{q_0}]), \qquad [r_1, \dots, r(q_0)] := (\Re')^{-1},$$

$$\overline{\Xi}_N = \begin{bmatrix} \overline{\zeta}_N^{\{1\}} & \overline{\zeta}_N^{\{2\}} & \dots & \overline{\zeta}_N^{\{n_0\}} \\ \overline{\zeta}_N^{\{n_0+1\}} & \overline{\zeta}_N^{\{n_0+2\}} & \dots & \overline{\zeta}_N^{\{2n_0\}} \\ & & \vdots & \\ \overline{\zeta}_N^{\{q_0n_0+1\}} & \overline{\zeta}_N^{\{q_0n_0+2\}} & \dots & \overline{\zeta}_N^{\{n_0(q_0+1)\}} \end{bmatrix}.$$

Then, we have the following identification algorithm:

(i) *Estimate* θ*.* The estimate of θ is taken as

$$\theta_N^e = (\overline{\Phi}(q_0))^{-1}(\overline{\Xi}_N)' r_N(q_0).$$

(ii) *Estimate* η*.* Let $b_0^e(j) = 0$ and

$$b_N^e(j) = \begin{cases} [\overline{\zeta}_N^{\{i_N^e(j)\}}, \overline{\zeta}_N^{2n_0+i_N^e(j)}, \dots, \\ \overline{\zeta}_N^{\{2q_0n_0+i_N^e(j)\}}] r_N(i_N^e(j))/\overline{\tau}_N^{\{j,i_N^e(j)\}}, & \text{if } \overline{\tau}_N^{\{j,i_N^e(j)\}} \neq 0, \\ b_{N-1}^e(j), & \text{if } \overline{\tau}_N^{\{j,i_N^e(j)\}} = 0, \end{cases}$$

where

$$i_N^e(j) = \min\{\operatorname{argmax}_{1\le i\le n_0} |\overline{\tau}_N^{\{j\}}(i)|\}, \quad j = 0, 1, \dots, q_0 - 1,$$

$r(i_N^e(j))$ is the $i_N^e(j)$-th column of $(\Re')^{-1}$, and $\overline{\tau}_N^{\{j,i_N^e(j)\}}$ is the $i_N^e(j)$-th component of

$$\overline{\tau}_N(j) = \overline{\tau}^{\{j\}}\theta_N.$$

Then, the estimate of η is taken as

$$\eta_N^e = [b_N^e(0), \dots, b_N^e(q_0 - 1)]'.$$

Theorem 12.22. *Suppose* $u \in \mathcal{U}_\lambda(n_0, q_0)$*. Then, under Assumptions* (A12.1) *and* (A12.2)*,*

$$\theta_N^e \to \theta \quad \textit{and} \quad \eta_N^e \to \eta \quad \textit{w.p.1 as} \quad N \to \infty.$$

12.6 Examples

In this section, we illustrate the convergence of the estimates given by the algorithms described above. The noise is Gaussian with known mean and variance. In Example 12.23, the identification algorithm with quantized sensors is shown. Example 12.24 concerns the identification of systems with non-monotonic nonlinearities. Example 12.25 illustrates an algorithm based on the prior information, which is more simplified than the one described by (12.38)–(12.40). The parameter estimates are shown to be convergent in all cases.

Example 12.23. Consider a gain system $y_k = a + d_k$. Here the actual value of the unknown a is 5. The disturbance is a sequence of i.i.d. Gaussian variables with zero mean and standard deviation $\sigma = 5$. The sensor has three switching thresholds, $C_1 = 2$, $C_2 = 6$, and $C_3 = 10$. Then, the recursive algorithm in Section 12.4 is used to generate quasi-convex combination estimates. For comparison, estimates derived by using each threshold individually (i.e., binary-valued sensors) are also calculated. Figure 12.2 compares quasi-convex combination estimates to those using each threshold. It is shown that the estimate with three thresholds converges faster than the ones with each threshold individually. The weights of the estimates of each threshold are shown in Figure 12.3, which illustrates that the weights are not sure to be positive.

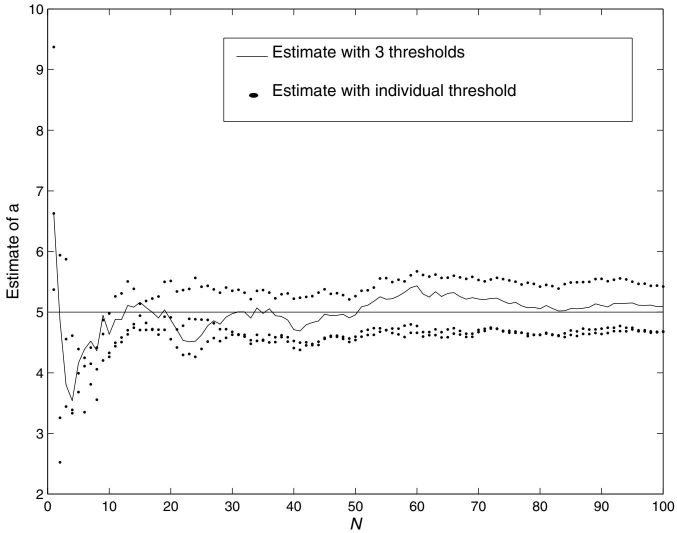


FIGURE 12.2. Identification with quantized output observations

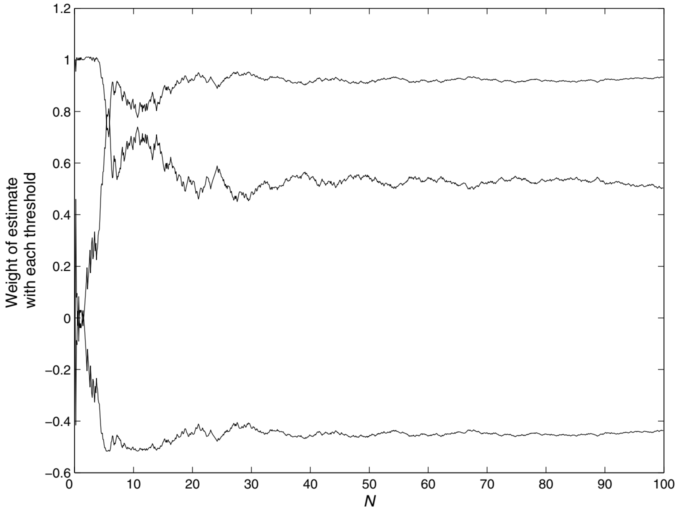


FIGURE 12.3. Weights of estimates with each threshold

Example 12.24. Consider

$$\begin{cases} y_k = a_0 x_k + a_1 x_{k-1} + d_k, \\ x_k = b_0 + b_1 u_k + b_2 u_k^2 + u_k^3, \end{cases}$$

where the noise $\{d_k\}$ is a sequence of i.i.d. Gaussian variables with $Ed_1 = 0$, $\sigma_d^2 = 1$. The output is measured by a binary-valued sensor with threshold $C = 13$. The linear subsystem has order $n_0 = 2$, and the nonlinear function has order $q_0 = 2$. The prior information on a_i, $i = 0, 1$, is that $a_i \in [0.5, 5]$. Suppose the true values of unknown parameters are $\theta = [a_0, a_1]' = [1.31, 0.85]'$ and $\eta = [b_0, b_1, b_2]' = [4, 1.4, -3]'$. The nonlinearity is not monotone, which is illustrated in Figure 12.4. It is shown that not all values of $v, \rho_1 v, \rho_2 v, \rho_3 v$ are situated in the same monotone interval of the nonlinearity.

The input is chosen to be $2n_0(q_0 + 1) = 12$-periodic with one period $(v, v, \rho_1 v, \rho_1 v,\ \rho_2 v, \rho_2 v)$, where $v = [1.2, 0.85]$, $\rho_1 = 0.5$, $\rho_2 = 1.65$, and $\rho_3 = 0.75$. Define the block variables X_l, Y_l, Φ_l^j, D_l, and S_l, in the case of a six-periodic input. Using (12.12), we can construct the algorithms (12.38)–(12.40) to identify θ and η.

The estimation errors of θ and η are illustrated in Figure 12.5, where the errors are measured by the Euclidean norm. Both parameter estimates of the linear and nonlinear subsystems converge to their true values, despite the nonlinearity being non-monotonic.

Example 12.25. For some prior information, algorithms (12.38)–(12.40) can be simplified. For example, the estimation algorithms of η can be simplified when the prior information on θ is known to be positive and the

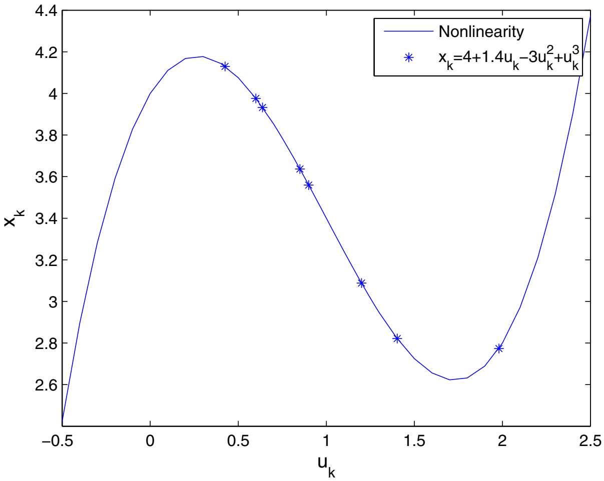


FIGURE 12.4. Nonmonotonic nonlinearity

periodic input u is positive. Both the mean and the variance of disturbance are not zero in this example.

Consider

$$\begin{cases} y_k = a_0 x_k + a_1 x_{k-1} + d_k, \\ x_k = b_0 + b_1 u_k + u_k^2, \end{cases}$$

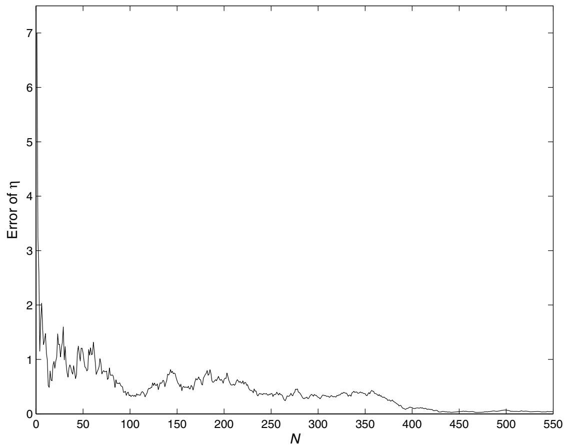
where the noise $\{d_k\}$ is a sequence of i.i.d. Gaussian variables with $Ed_1 = 2$, $\sigma_d^2 = 4$. The output is measured by a binary-valued sensor with threshold $C = 13$. The linear subsystem has order $n_0 = 2$, and the nonlinear function has order $q_0 = 2$. The prior information on a_i, $i = 0, 1$, is that $a_i \in [0.5, 5]$. Suppose the true values of the unknown parameters are $\theta = [a_0,\ a_1]' = [1.17,\ 0.95]'$ and $\eta = [b_0,\ b_1]' = [3,\ 1.3]'$.

The input is 12-periodic with one period $(v, v, \rho_1 v, \rho_1 v, \rho_2 v, \rho_2 v)$, where $v = [1.2, 0.85]$, $\rho_1 = 0.65$, and $\rho_2 = 1.25$. Define the block variables X_l, Y_l, Φ_l^j, D_l and S_l, in the case of a 12-periodic input. Using (12.12), we can construct the algorithms (12.38)–(12.40) to identify θ.

Considering the prior information on θ, a more simplified algorithm can be constructed to identify η than the one given by (12.38)–(12.40). Note that $a_i \in [0.5, 5]$, $i = 1, 2$, and u is positive. Then, $\tau^{\{j,1\}}$, the first component of $V^j\theta$, is

$$\tau^{\{j,1\}} = a_0 v_2^2 + a_1 v_1^2 \geq 0.5(v_2^2 + v_1^2) \neq 0,$$

where the last inequality is derived from the fact that v is strongly 2 full rank. So, it is not necessary to calculate $i_N^*(j)$ in (12.39), which aims to



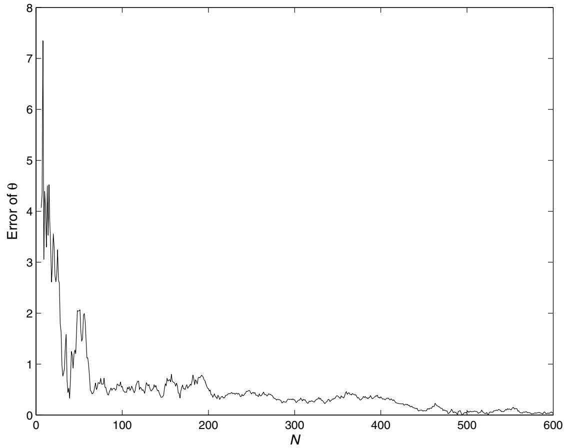


FIGURE 12.5. Identification errors of θ and η with nonmonotonic nonlinearity

find the nonzero component of $\tau^{\{j\}}$. And η can be estimated as follows:

$$\eta_0 = 0$$

$$\eta_N = \begin{cases} \Lambda_N \Re^c [\zeta_N^{\{1\}}, \zeta_N^{\{2n_0+1\}}, \ldots, \zeta_N^{\{2q_0 n_0+1\}}]', & \text{if } \prod_{j=0}^{q_0-1} \tau_N^{\{j,1\}} \neq 0, \\ \eta_{N-1}, & \text{if } \prod_{j=0}^{q_0-1} \tau_N^{\{j,1\}} = 0, \end{cases}$$

where $\Lambda_N = \text{diag}^{-1}(\tau_N^{\{0,1\}}, \ldots, \tau_N^{\{q_0-1,1\}})$, $\Re^c$ is a $q_0 \times (q_0+1)$ matrix containing the first to q_0-1th rows of $\Re^{-1}$, and $\tau_N^{\{j,1\}}$ is the first component of $\tau_N^{\{j\}} = V^j \theta_N$.

The estimation errors of θ and η are shown in Figure 12.6, where the errors are measured by the Euclidean norm. Both parameter estimates of the linear and nonlinear subsystems converge to their true values.

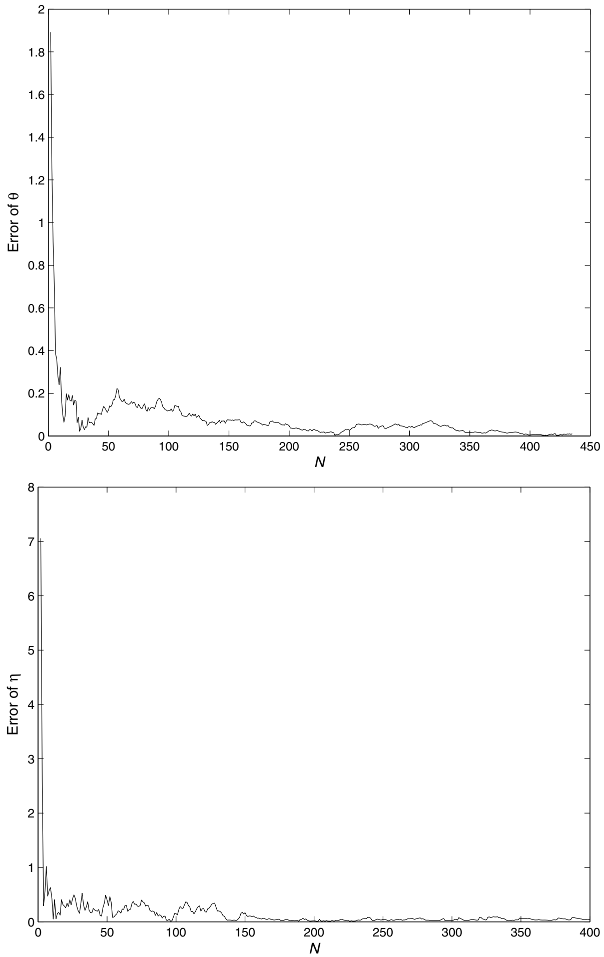


FIGURE 12.6. Identification errors of θ and η

12.7 Notes

In this chapter, the identification of Hammerstein systems with quantized output observations is studied. The development follows [128]. Hammerstein systems have been used to model practical systems with memoryless actuators and have been studied extensively in system identification, see for example [3, 45, 69, 71].

Structurally, a Hammerstein system with a quantized sensor may be viewed as Hammerstein–Wiener system which contains both input and output nonlinearity. However, our approaches are quite different from typical studies of such nonlinear system identification problems in which the output nonlinearities usually contain some sections of smooth functions. Unlike traditional approximate gradient methods or covariance analysis, we employ the methods of empirical measures and parameter mappings. Under assumptions of known noise distribution functions and strongly scaled full-rank inputs, identification algorithms, convergence properties, and the estimation efficiency are derived.