9

Worst-Case Identification under Binary-Valued Observations

This chapter focuses on the identification of systems where the disturbances are formulated in a deterministic framework as unknown but bounded. Different from the previous chapters, here the identification error is measured by the radius of the set that the unknown parameters belong to, which is a worst-case measure of the parameter uncertainties. By considering several different combinations of the disturbances and unmodeled dynamics, a number of fundamental issues are studied in detail: When only binary-valued observations are available, how accurately can one identify the parameters of the system? How fast can one reduce uncertainty on model parameters? What are the optimal inputs for fast identification? What is the impact of unmodeled dynamics and disturbances on identification accuracy and time complexity?

The rest of this chapter is arranged as follows. In Section 9.1, the problem is formulated and the main conditions on the disturbances and unmodeled dynamics are given. In Section 9.2, lower bounds on the identification errors and time complexity of the identification algorithms are established, underscoring an inherent relationship between identification time complexity and the Kolmogorov ε-entropy. Identification input design and upper bounds on identification errors are then derived in Section 9.3, demonstrating that the Kolmogorov ε-entropy indeed defines the complexity rates. For the single parameter case, the results are refined further in Section 9.4. Section 9.5 presents a comparison between the stochastic and deterministic frameworks. In contrast to the common perception that these two are competing frameworks, we show that they complement each other in binary sensor identification.

L.Y. Wang et al., *System Identification with Quantized Observations*, Systems & Control: Foundations & Applications, DOI 10.1007/978-0-8176-4956-2_9,

9.1 Worst-Case Uncertainty Measures

Recall the linear system defined in Chapter 2 and further detailed in Chapter 4,

$$y_k = \sum_{i=0}^{\infty} a_i u_{k-i} + d_k, \qquad k = k_0+1, k_0+2, \ldots,$$

where $\{d_k\}$ is a sequence of disturbances, $\{u_k\}$ is the input with $u_k = 0$, $k < k_0$, and $a = \{a_i, i = 0, 1, \ldots\}$, satisfying $\|a\|_1 = \sum_{i=0}^{\infty} |a_i| < \infty$, is the vector-valued parameter.

As in Chapter 4, for a given model order n_0, the system parameters can be decomposed into the modeled part $\theta = [a_0, \ldots, a_{n_0-1}]'$ and the unmodeled dynamics $\widetilde{\theta} = [a_{n_0}, a_{n_0+1}, \ldots]'$, and the system input–output relationship can be expressed as

$$y_k = \phi_k'\theta + \widetilde{\phi}_k'\widetilde{\theta} + d_k, \quad k = k_0+1, k_0+2, \ldots, \tag{9.1}$$

where

$$\phi_k = [u_k, u_{k-1}, \ldots, u_{k-n_0+1}]',$$
$$\widetilde{\phi}_k = [u_{k-n_0}, u_{k-n_0-1}, \ldots]'.$$

Under a selected input sequence u_k, the output s_k from a binary-valued sensor of threshold C is measured for $k = k_0+1, \ldots, k_0+N$. We would like to estimate θ on the basis of input–output observations on u_k and s_k. The issues of identification accuracy, time complexity, and input design will be discussed.

Because some results in this chapter will be valid under any l^p norm, the following assumption is given in a generic l^p norm. The norm will be further specified if certain results are valid only for some p values.

(A9.1) For a fixed $p \geq 1$, to be specified later,

1. the unmodeled dynamics $\widetilde{\theta}$ is bounded in the l^p norm by $\|\widetilde{\theta}\|_p \leq \eta$;
2. the disturbance d is uniformly bounded in the l^∞ norm by $\|d\|_\infty \leq \delta$;
3. the prior information on θ is given by $\Omega_0 = \mathrm{Ball}_p(\theta_0, \varepsilon_0) \subset \mathbb{R}^{n_0}$ for $\theta_0 \in \mathbb{R}^{n_0}$ and $\varepsilon_0 > 0$.

For a selected input sequence u_k, let $s = \{s_k,\ k = k_0+1, \ldots, k_0+N\}$ be the observed output. Define

$$\Omega_N(k_0, u, s) = \{\theta : \ s_k = I_{\{\phi_k'\theta + \widetilde{\phi}_k'\widetilde{\theta} + d_k \leq C\}} \text{ for some } \|\widetilde{\theta}\|_p \leq \eta, \\ \|d\|_\infty \leq \delta \text{ and } k = k_0+1, \ldots, k_0+N\}$$

and

$$e_N = \inf_{\|u\|_\infty \leq u_{\max}} \sup_{k_0} \sup_{s} \mathrm{Rad}_p\left(\Omega_N(k_0, u, s) \cap \mathrm{Ball}_p(\theta_0, \varepsilon_0)\right),$$

where $\text{Rad}_p(\Omega)$ is the radius of Ω in the l_p norm. e_N is the optimal (in terms of the input design) worst-case (with respect to initial time, unmodeled dynamics, and disturbances) uncertainty after N steps of observations. For a given desired identification accuracy ε, the time complexity of $\text{Ball}_p(\theta_0, \varepsilon_0)$ is defined as

$$N(\varepsilon) = \min\{N : \; e_N \leq \varepsilon\}.$$

We will characterize e_N, determine optimal or suboptimal inputs u, and derive bounds on time complexity $N(\varepsilon)$.

9.2 Lower Bounds on Identification Errors and Time Complexity

We will show in this section that identification time complexity is bounded below by the Kolmogorov entropy of the prior uncertainty set.

Noise-Free and No Unmodeled Dynamics

Theorem 9.1. *Assume Assumption* (A9.1)*. Let $\delta = 0$ and $\eta = 0$. Suppose that for a given $p \geq 1$ the prior uncertainty $\Omega_0 = \text{Ball}_p(\theta_0, \varepsilon_0)$. Then, for any $\varepsilon < \varepsilon_0$, the time complexity $N(\varepsilon)$ is bounded below by $N(\varepsilon) \geq n_0 \log(\varepsilon_0/\varepsilon)$.*

Proof. $\text{Ball}_p(c, \varepsilon)$ in $\mathbb{R}^{n_0}$ has volume $a_{p,n_0}\varepsilon^{n_0}$, where the coefficient a_{p,n_0} is independent of ε. To reduce the identification error from ε_0 to below ε, the volume reduction must be at least

$$a_{p,n_0}\varepsilon^{n_0}/(a_{p,n_0}\varepsilon_0^{n_0}) = (\varepsilon/\varepsilon_0)^{n_0}.$$

Each binary sensor observation defines a hyperplane in the parameter space $\mathbb{R}^{n_0}$. The hyperplane divides an uncertainty set into two subsets, with the volume of the larger subset at least half of the volume of the original set. As a result, in a worst-case scenario, one binary observation can reduce the volume of a set by 1/2 at best. Hence, the number N of observations required to achieve the required error reduction is at least

$$(1/2)^N \leq (\varepsilon/\varepsilon_0)^{n_0}, \quad \text{or } N \geq n_0 \log(\varepsilon_0/\varepsilon).$$

The proof is concluded. □

It is noted that $n \log(\varepsilon_0/\varepsilon)$ is precisely the Kolmogorov ε-entropy of the prior uncertainty set Ω_0 [50, 125]. Hence, Theorem 9.1 provides an interesting new interpretation of the Kolmogorov entropy in system identification, beyond its application in characterizing model complexity [125]. Theorem 9.1 establishes a lower bound of exponential rates of time complexity. Upon obtaining an upper bound of the same rates in the next section, we will show that the Kolmogorov ε-entropy indeed defines the time complexity rates in this problem. Next, we present an identifiability result, which is limited to $p = 1$.

Proposition 9.2. *The uncertainty set* $\mathrm{Ball}_1(0, C/u_{\max})$ *is not identifiable.*

Proof. For any $\theta \in \mathrm{Ball}_1(0, C/u_{\max})$, the output

$$y_k = \phi_k'\theta \leq \|\phi_k\|_\infty \|\theta\|_1 \leq u_{\max}\frac{C}{u_{\max}} = C.$$

It follows that $s_k = 1, \forall k$. Hence, the observations could not provide further information to reduce uncertainty. □

Bounded Disturbances

In the case of noisy observations, the input–output relationship becomes

$$y_k = \phi_k'\theta + d_k, \quad s_k = I_{\{y_k \leq C\}}, \tag{9.2}$$

where $|d_k| \leq \delta$. For any given ϕ_k, an observation on s_k from (9.2) defines, in a worst-case sense, two possible uncertainty half-planes:

$$\begin{aligned}\Omega_1 &= \{\theta \in \mathbb{R}^{n_0} : \ \phi_k'\theta \leq C + \delta\}, \quad s_k = 1,\\ \Omega_2 &= \{\theta \in \mathbb{R}^{n_0} : \ \phi_k'\theta > C - \delta\}, \quad s_k = 0.\end{aligned}$$

Uncertainty reduction via observation is possible only if the uncertainty set before observation is not a subset of each half-plane (so that the intersection of the uncertainty set and the half-plane results in a smaller set).

Theorem 9.3. *If* $\varepsilon \leq \delta/u_{\max}$*, then for any* $\theta_0 \in \mathbb{R}^{n_0}$*, either* $\mathrm{Ball}_1(\theta_0, \varepsilon) \subseteq \Omega_1$ *or* $\mathrm{Ball}_1(\theta_0, \varepsilon) \subseteq \Omega_2$*. Consequently, in a worst-case sense,* $\mathrm{Ball}_1(\theta_0, \varepsilon)$ *is not identifiable.*

Proof. Suppose that $\mathrm{Ball}_1(\theta_0, \varepsilon) \nsubseteq \Omega_1$. Then, there exists $\theta_1 \in \mathrm{Ball}_1(\theta_0, \varepsilon)$ such that $\phi_k'\theta_1 > C + \delta$. $\theta \in \mathrm{Ball}_1(\theta_0, \varepsilon)$ satisfies $\|\theta - \theta_1\|_1 \leq 2\varepsilon$. We have

$$\begin{aligned}\phi_k'\theta \ &= \phi_k'\theta_1 + \phi_k'(\theta - \theta_1)\\ &> C + \delta + \phi_k'(\theta - \theta_1)\\ &\geq C + \delta - u_{\max}2\varepsilon \geq C - \delta,\end{aligned}$$

for any $\theta \in \mathrm{Ball}_1(\theta_0, \varepsilon)$. This implies that $\mathrm{Ball}_1(\theta_0, \varepsilon) \subseteq \Omega_2$. Likewise, we can show that if $\mathrm{Ball}_1(\theta_0, \varepsilon) \not\subset \Omega_2$, then it is contained in Ω_1. □

Theorem 9.3 shows that worst-case disturbances introduce irreducible identification errors of size at least $\delta/u_{\max}$. This is a general result. A substantially higher lower bound can be obtained in the special case of $n_0 = 1$.

Consider the system $y_k = au_k + d$. Suppose that at time k the prior information on a is that $a \in \Omega = [\underline{a}, \overline{a}]$ with $\underline{a} > C/u_{\max}$ for identifiability (see Proposition 9.2). The uncertainty set has center $a_0 = (\underline{a} + \overline{a})/2$ and radius $\varepsilon = (\overline{a} - \underline{a})/2$. To minimize the posterior uncertainty in the worst-case sense, the optimal u_k can be easily obtained as $u_k = C/a_0$.

Theorem 9.4. *If $\delta < C$, then the uncertainty set $[\underline{a}, \overline{a}]$ cannot be reduced if*

$$\varepsilon \leq \frac{\delta/u_{\max}}{1 - \delta/C}.$$

Proof. Let $\varepsilon = \frac{\delta/u_{\max}}{1-\delta/C}$. Then, $\delta = \frac{\varepsilon C}{C/u_{\max}+\varepsilon}$. For any $a \in [\underline{a}, \overline{a}]$, noting $a_0 = \underline{a} + \varepsilon$, we have $|a - a_0| \leq \varepsilon$, and

$$\begin{aligned} au_k &= a\frac{C}{a_0} = (a_0 + (a - a_0))\frac{C}{a_0} = C + (a - a_0)\frac{C}{a_0} \\ &\leq C + \frac{\varepsilon C}{\underline{a} + \varepsilon} < C + \frac{\varepsilon C}{C/u_{\max} + \varepsilon} = C + \delta. \end{aligned}$$

Hence, the observation $s_k = 1$ does not provide any information. Similarly, if $s_k = 0$, we can show that all $\theta \in [\underline{a}, \overline{a}]$ will result in $au_k > C - \delta$. Again, the observation does not reduce uncertainty. □

At present, it remains an open question if Theorem 9.4 holds for higher-order systems.

Unmodeled Dynamics

When the system contains unmodeled dynamics, the input–output relationship becomes

$$y_k = \phi_k'\theta + \widetilde{\phi}_k'\widetilde{\theta}, \qquad s_k = I_{\{y_k \leq C\}}, \tag{9.3}$$

where $\|\widetilde{\theta}\|_1 \leq \eta$. We will show that unmodeled dynamics will introduce an irreducible identification error on the modeled part.

For any $\widetilde{\phi}_k$, the set $\{\widetilde{\phi}_k'\widetilde{\theta} : \|\widetilde{\theta}\|_1 \leq \eta\} = [-\eta m_k, \eta m_k]$, where $m_k = \|\widetilde{\phi}_k\|_\infty$.

Theorem 9.5. *If $\varepsilon \leq \eta$, then in a worst-case sense, for any θ_0,* $\mathrm{Ball}_1(\theta_0, \varepsilon)$ *is not identifiable.*

Proof. Under (9.3), an observation on s_k provides observation information

$$\Omega_1 = \{\theta \in \mathbb{R}^{n_0} : \phi_k'\theta \leq C + \eta m_k\}, \quad s_k = 1,$$
$$\Omega_2 = \{\theta \in \mathbb{R}^{n_0} : \phi_k'\theta > C - \eta m_k\}, \quad s_k = 0.$$

In the worst-case sense, $\mathrm{Ball}_1(\theta_0, \varepsilon)$ can be reduced by this observation only if $\mathrm{Ball}_1(\theta_0, \varepsilon)$ is a subset of neither Ω_1 nor Ω_2.

Suppose that $\mathrm{Ball}_1(\theta_0, \varepsilon) \not\subseteq \Omega_2$. We will show that $\mathrm{Ball}_1(\theta_0, \varepsilon) \subseteq \Omega_1$. Indeed, in this case, there exists $\theta_1 \in \mathrm{Ball}_1(\theta_0, \varepsilon)$ such that $\phi_k'\theta_1 \leq C - \eta m_k$. Since any $\theta \in \mathrm{Ball}_1(\theta_0, \varepsilon)$ satisfies $\|\theta - \theta_1\|_1 \leq 2\varepsilon$, we have

$$\begin{aligned} \phi_k'\theta &= \phi_k'\theta_1 + \phi_k'(\theta - \theta_1) \\ &\leq C - \eta m_k + \phi_k'(\theta - \theta_1) \\ &\leq C - \eta m_k + m_k 2\varepsilon \\ &\leq C + \eta m_k. \end{aligned}$$

This implies $\mathrm{Ball}_1(\theta_0, \varepsilon) \subseteq \Omega_1$. □

9.3 Upper Bounds on Time Complexity

In this subsection, general upper bounds on identification errors or time complexity will be established.

For a fixed $p \geq 1$, suppose that the prior information on θ is given by $\mathrm{Ball}_p(\theta_0, \varepsilon_0)$. For identifiability, assume that the signs of a_i have been detected and

$$\underline{a} = \min\{|\underline{a}_i|, i = 1, \ldots, n\} > \frac{C}{u_{\max}}.$$

The sign of a_i can be obtained easily by choosing an initial testing sequence of u. Also, those parameters with $|\underline{a}_i| < C/u_{\max}$ can be easily detected. Since uncertainty on these parameters cannot be further reduced (see Proposition 9.2), they will be left as remaining uncertainty. $\underline{a}$ defined here will be applied to the rest of the parameters. The detail is omitted for brevity. Denote

$$\overline{a} = \max_{\theta \in \mathrm{Ball}_p(\theta_0, \varepsilon_0)} \|\theta\|_\infty.$$

We will establish upper bounds on the time complexity $N(\varepsilon)$ to reduce the size of the uncertainty from ε_0 to ε, in the l^p norm.

Noise-Free and No Unmodeled Dynamics

Let $\eta = 0$ and $\delta = 0$ and consider $y_k = \phi_k'\theta$.

Theorem 9.6. *Suppose that $u_{\max} > C/\underline{a}$. Then the time complexity to reduce the uncertainty from ε_0 to ε is bounded by*

$$N(\varepsilon) \leq ({n_0}^2 - n_0 + 1) \left\lceil \frac{1}{p} \log n_0 + \log \frac{\varepsilon_0}{\varepsilon} \right\rceil. \tag{9.4}$$

Since n_0 is a constant independent of N, this result, together with Theorem 9.1, confirms that the Kolmogorov entropy defines the time complexity rates in binary sensor identification. The accurate calculation for $N(\varepsilon)$ remains an open and difficult question, except for $n_0 = 1$ (gain uncertainty), which is discussed in the next section.

The proof of Theorem 9.6 utilizes the following lemma. Consider the first-order system $y_k = au_k$, $s_k = I_{\{y_k \le C\}}$, where $a \in [\underline{a}, \overline{a}]$ and $\underline{a} > C/u_{\max} > 0$. Let $\varepsilon_0 = (\overline{a} - \underline{a})/2$.

Lemma 9.7. *There exists an input sequence u such that N observations on s_k can reduce the radius of uncertainty to $\varepsilon = 2^{-N}\varepsilon_0$.*

Proof. Let $[\underline{a}_k, \overline{a}_k]$ be the prior uncertainty before a measurement on s_k. Then $\varepsilon_k = (\overline{a}_k - \underline{a}_k)/2$. By choosing $u_k = C/(\underline{a}_k + \varepsilon_k)$, the observation on s_k will determine uniquely either $a \in [\underline{a}_k, \underline{a}_k + \varepsilon_k]$ if $s_k = 1$; or $a \in [\overline{a}_k - \varepsilon_k, \overline{a}_k]$ if $s_k = 0$. In either case, the uncertainty is reduced by half. Iterating on the number of observations leads to the conclusion. □

The proofs of this section rely on the following idea. Choose $u_k = 0$ except those with index $j({n_0}^2 - n_0 + 1) + i$, $i = 1, n_0 + 1, \ldots, (n_0 - 1)n_0 - n_0 + 3$, $j = 0, 1, \ldots$ This input design results in a specific input–output relationship:

$$\begin{cases} y_{j({n_0}^2-n_0+1)+n_0} = a_{n_0-1} u_{j({n_0}^2-n_0+1)+1}, \\ y_{j({n_0}^2-n_0+1)+n_0+1} = a_0 u_{j({n_0}^2-n_0+1)+n_0+1}, \\ \quad\vdots \\ y_{j({n_0}^2-n_0+1)+(n_0-1)n_0+1} = a_{n_0-2} u_{j({n_0}^2-n_0+1)+(n_0-1)n_0-n_0+3}. \end{cases} \tag{9.5}$$

In other words, within each block of ${n_0}^2 - n_0 + 1$ observations, each model parameter can be identified individually once. Less conservative inputs can be designed. However, they are more problem-dependent and ad hoc, and will not be presented here.

Proof of Theorem 9.6. By Lemma 9.7, the uncertainty radius on each parameter can be reduced by a factor of 2^{-N_1} after N_1 observations. This implies that by using the input (9.5), after $N = ({n_0}^2 - n_0 + 1)N_1$ observations, the uncertainty radius can be reduced to

$$\begin{aligned} \mathrm{rad}_p(\Omega_N) &\le {n_0}^{1/p}\mathrm{rad}_\infty(\Omega_N) \le {n_0}^{1/p} 2^{-\frac{N}{{n_0}^2-n_0+1}} \mathrm{rad}_\infty(\Omega_0) \\ &\le {n_0}^{1/p} 2^{-\frac{N}{{n_0}^2-n_0+1}} \mathrm{rad}_p(\Omega_0) = {n_0}^{1/p} 2^{-\frac{N}{{n_0}^2-n_0+1}} \varepsilon_0. \end{aligned}$$

Hence, for

$${n_0}^{1/p} 2^{-\frac{N}{{n_0}^2-n_0+1}} \varepsilon_0 \le \varepsilon,$$

it suffices to have

$$N = ({n_0}^2 - n_0 + 1)\left\lceil \frac{1}{p}\log n + \log\frac{\varepsilon_0}{\varepsilon} \right\rceil.$$

The desired result follows. □

Bounded Disturbances

Consider $y_k = \phi_k'\theta + d_k$, where $|d_k| \le \delta$.

Theorem 9.8. *Suppose $\delta < C$. Let*

$$\beta = \frac{\delta}{C}, \quad \rho = \frac{1}{2}(1-\beta), \quad \textit{and} \quad \sigma = \frac{\delta\overline{a}}{2C(1-\rho)} = \frac{\overline{a}\beta}{1+\beta}.$$

If $\varepsilon_0 > \varepsilon > \sigma$ and $u_{\max} > C/\underline{a}$, then the time complexity $N(\varepsilon)$ to reduce the uncertainty from ε_0 to ε is bounded in the l^p norm by

$$N(\varepsilon) \le ({n_0}^2 - n_0 + 1)\left\lceil \frac{1}{p}\log n_0 + \frac{\log\frac{\varepsilon-\sigma}{\varepsilon_0-\sigma}}{\log\rho}\right\rceil. \tag{9.6}$$

Proof. Using the input in (9.5), the identification of the n parameters a_0, ..., a_{n_0-1} is reduced to identifying each parameter individually. Now for identification of a single parameter $y_k = au_k + d_k$, we can derive the following iterative uncertainty reduction relationship. If the prior uncertainty at k is $[a_k - \varepsilon_k, a_k + \varepsilon_k]$, then the optimal worst-case input u_k can be shown as $u_k = C/a_k$. (More detailed derivations are given in the next section.) The posterior uncertainty will be either $[a_k - \varepsilon_k, (1+\beta)a_k]$, if $s_k = 1$; or $[(1-\beta)a_k, a_k + \varepsilon_k]$, if $s_k = 0$. Both have the radius

$$\varepsilon_{k+1} = \frac{1}{2}\left(\varepsilon_k + \beta a_k\right) = \frac{1-\beta}{2}\varepsilon_k + \frac{\beta}{2}(a_k + \varepsilon_k) \le \rho\varepsilon_k + \frac{\beta\overline{a}}{2}.$$

Starting from ε_0, after N_1 observations, we have

$$\begin{aligned}\varepsilon(N_1) &\le \rho^{N_1}\varepsilon_0 + \frac{\beta\overline{a}}{2}\sum_{i=0}^{N_1-1}\rho^i = \rho^{N_1}\varepsilon_0 + \frac{\beta\overline{a}}{2}\frac{1-\rho^{N_1}}{1-\rho}\\ &= \rho^{N_1}\varepsilon_0 + \sigma(1-\rho^{N_1}) = \rho^{N_1}(\varepsilon_0 - \sigma) + \sigma.\end{aligned}$$

To achieve $\varepsilon(N_1) \le \varepsilon$, it suffices that

$$\rho^{N_1}(\varepsilon_0 - \sigma) + \sigma \le \varepsilon \quad \text{or} \quad N_1 \ge \frac{\log\frac{\varepsilon-\sigma}{\varepsilon_0-\sigma}}{\log\rho}.$$

Following the same arguments as in the proof of Theorem 9.6, we conclude that

$$N = ({n_0}^2 - n_0 + 1)\left\lceil \frac{1}{p}\log n_0 + \frac{\log\frac{\varepsilon-\sigma}{\varepsilon_0-\sigma}}{\log\rho}\right\rceil$$

suffices to reduce the uncertainty from ε_0 to ε in the l^p norm. □

Unmodeled Dynamics

Consider $y_k = \phi_k' \theta + \widetilde{\phi}_k' \widetilde{\theta}$. The results of this case hold for $p = 1$ only. The unmodeled dynamics introduce an uncertainty on the observation on y_k: $\{\widetilde{\phi}_k' \widetilde{\theta} : \|\widetilde{\theta}\|_1 \leq \eta\} = [-\eta m_k, \eta m_k]$, $m_k = \|\phi_k\|_\infty$.

Theorem 9.9. *Suppose* $0 < \eta < C/u_{\max}$. *Let*

$$\rho_1 = \frac{1}{2}\left(1 - \frac{\eta u_{\max}}{C}\right), \quad \sigma_1 = \frac{\eta u_{\max} \overline{a}}{2C(1-\rho_1)}.$$

Then

$$N(\varepsilon) \leq ({n_0}^2 - n_0 + 1)\left\lceil \log n_0 + \frac{\log \frac{\varepsilon - \sigma_1}{\varepsilon_0 - \sigma_1}}{\log \rho_1} \right\rceil. \tag{9.7}$$

Proof. By using the input (9.5), the identification of θ is reduced to each of its components. For a scalar system $y_k = a u_k + \widetilde{\phi}_k' \widetilde{\theta}$, since $|\widetilde{\phi}_k' \widetilde{\theta}| \leq \eta u_{\max}$, we can apply Theorem 9.8 with δ replaced by $\eta u_{\max}$. Inequality (9.7) then follows from Theorem 9.8. □

9.4 Identification of Gains

In the special case $n = 1$, explicit results and tighter bounds can be obtained. When $n = 1$, the observation equation becomes

$$y_k = a u_k + \widetilde{\phi}_k' \widetilde{\theta} + d_k.$$

Assume that the initial information on a is that $\underline{a}_0 \leq a \leq \overline{a}_0$, $\underline{a}_0 \neq 0$, $\overline{a}_0 \neq 0$, with radius $\varepsilon_0 = (\overline{a}_0 - \underline{a}_0)/2$.

Case 1: $y_k = a u_k$

It is noted that this is a trivial identification problem when regular sensors are used: After one input $u_0 \neq 0$, a can be identified uniquely.

Theorem 9.10. *The following assertions hold.*

(1) *Suppose the sign of a is known, say, $\underline{a}_0 > 0$, and $u_{\max} \geq C/\underline{a}_0$. Then the optimal identification error is $e_N = 2^{-N} e_0$ and the time complexity is $N(\varepsilon) = \lceil \log(\varepsilon_0/\varepsilon) \rceil$.*

If, at $k-1$, the information on a is that $a \in [\underline{a}_{k-1}, \overline{a}_{k-1}]$, then the one-step optimal u_k is

$$u_k = \frac{2C}{\underline{a}_{k-1} + \overline{a}_{k-1}}, \tag{9.8}$$

where $\underline{a}_k$ and $\overline{a}_k$ are updated by

$$\underline{a}_k = \begin{cases} (\underline{a}_{k-1} + \overline{a}_{k-1})/2, & \textit{if } s_k = 0, \\ \underline{a}_{k-1}, & \textit{if } s_k = 1; \end{cases}$$

$$\overline{a}_k = \begin{cases} \overline{a}_{k-1}, & \textit{if } s_k = 0, \\ (\underline{a}_{k-1} + \overline{a}_{k-1})/2, & \textit{if } s_k = 1. \end{cases}$$

(2) *If $\underline{a}_0$ and $\overline{a}_0$ have opposite signs and*

$$\delta_l = \max\left\{\underline{a}_0, -\frac{C}{u_{\max}}\right\}, \quad \delta_h = \min\left\{\overline{a}_0, \frac{C}{u_{\max}}\right\},$$

then the uncertainty interval (δ_l, δ_h) is not identifiable. Furthermore, in the case of $\underline{a}_0 \le \delta_l$ and $\overline{a}_0 \ge \delta_h$, if $\delta_h - \delta_l \le \varepsilon$ and $\varepsilon_0 \ge 2\varepsilon$, then the time complexity $N(\varepsilon)$ is bounded by

$$\left\lceil \log \frac{\varepsilon_0}{\varepsilon} \right\rceil \le N(\varepsilon) \le \left\lceil \log \frac{\varepsilon_0 - (\delta_h - \delta_l)}{\varepsilon} \right\rceil + 2.$$

Proof. The proof is divided into a couple of steps.

(1) The identification error and time complexity follow directly from Theorems 9.1 and 9.6 with $n = 1$. As for the optimal input, note that starting from the uncertainty $[\underline{a}_k, \overline{a}_k]$, an input u_k defines a testing point C/u_k on a. The optimal worst-case input is then obtained by placing the testing point at the middle. That is,

$$\frac{C}{u_k} = \frac{1}{2}(\underline{a}_k + \overline{a}_k),$$

which leads to the optimal input and results in posterior uncertainty sets.

(2) When the input is bounded by $u_k \in [-u_{\max}, u_{\max}]$, the testing points cannot be selected in the interval $[-C/u_{\max}, C/u_{\max}]$. Consequently, this uncertainty set cannot be further reduced by identification. Furthermore, by using $u_1 = -u_{\max}$ and $u_2 = u_{\max}$ as the first two input values, a can be determined as belonging uniquely to one of the three intervals:

$$[\underline{a}_0, -C/u_{\max}), \ [-C/u_{\max}, C/u_{\max}], \ [C/u_{\max}, \overline{a}_0].$$

By taking the worst-case scenario of

$$\overline{a}_0 - C/u_{\max} = \varepsilon_0 - (\delta_h - \delta_l),$$

the time complexity for reducing the remaining uncertainty to ε is $\left\lceil \log \frac{\varepsilon_0 - (\delta_h - \delta_l)}{\varepsilon} \right\rceil$. This leads to the upper bound on $N(\varepsilon)$. The lower bound follows from Theorem 9.1 with $n = 1$.

□

In this special case, the actual value $C > 0$ does not affect the identification accuracy. This is due to noise-free observation. The value C will become essential in deriving optimal identification errors when observation noises are present. $C = 0$ is a singular case in which the uncertainty on a cannot be reduced (in the sense of the worst-case scenario). Indeed, in this case, one can only test the sign of a. It is also observed that the optimal u_k depends on the previous observation s_{k-1}. As a result, u_k can be constructed causally and sequentially, but not off-line.

Case 2: $y_k = au_k + d_k$

Here we assume $|d_k| \le \delta < C$. The prior information on a is given by $a \in \Omega_0 = [\underline{a}_0, \overline{a}_0]$, and $\underline{a}_0 > 0$.

Theorem 9.11. *Suppose that*

$$u_{\max} \ge \frac{C}{\underline{a}_0} \quad \textit{and} \quad \frac{\overline{a}_0}{\underline{a}_0} \ge \frac{1+\beta}{1-\beta}.$$

Then

(1) *the optimal input u_k is given by the causal mapping from the available information at $k-1$:*

$$u_k = \frac{2C}{\underline{a}_{k-1} + \overline{a}_{k-1}}.$$

The optimal identification error satisfies the iteration equation

$$e_k = \frac{1}{2} e_{k-1} + \frac{1}{2}\beta(\overline{a}_{k-1} + \underline{a}_{k-1}), \tag{9.9}$$

where $\overline{a}_k$ and $\underline{a}_k$ are updated by the rules

$$\begin{aligned} \overline{a}_k &= \overline{a}_{k-1}, \quad \underline{a}_k = \frac{C-\delta}{u_k}, \quad \text{if } s_k = 0, \\ \underline{a}_k &= \underline{a}_{k-1}, \quad \overline{a}_k = \frac{C+\delta}{u_k}, \quad \text{if } s_k = 1. \end{aligned}$$

(2)

$$\frac{\overline{a}(k)}{\underline{a}(k)} \ge \frac{1+\beta}{1-\beta} \quad \textit{for all} \quad k \ge 1;$$

$\{\underline{a}_k\}$ *is monotonically increasing,* $\{\overline{a}_k\}$ *and* $\left\{\frac{\overline{a}_k}{\underline{a}_k}\right\}$ *are monotonically decreasing;*

$$\lim_{k\to\infty} \frac{\overline{a}_k}{\underline{a}_k} = \frac{1+\beta}{1-\beta}.$$

(3) *At each time k, uncertainty reduction is possible if and only if*

$$\frac{\overline{a}_{k-1}}{\underline{a}_{k-1}} > \frac{1+\beta}{1-\beta}.$$

Proof. (1) Since $u_k > 0$, the relationship (9.2) can be written as $a = \frac{y_k - d_k}{u_k}$. The observation outcome $y_k \geq C$ will imply that

$$a \geq \frac{C - d_k}{u_k} \geq \frac{C-\delta}{u_k},$$

which will reduce uncertainty from $a \in [\underline{a}_{k-1}, \overline{a}_{k-1}]$ to $[\frac{C-\delta}{u_k}, \overline{a}_{k-1}]$ with error $e_1(k) = \overline{a}_{k-1} - \frac{C-\delta}{u_k}$. Similarly, $y < C$ implies $a < \frac{C+\delta}{u_k}$ and $a \in [\underline{a}_{k-1}, \frac{C+\delta}{u_k}]$ with $e_2(k) = \frac{C+\delta}{u_k} - \underline{a}_{k-1}$. In a worst-case scenario,

$$e_k = \max\{e_1(k), e_2(k)\}.$$

Consequently, the optimal u_k can be derived from $\inf_{u_k} e_k$. Hence, the optimal u_k is the one that causes $e_1(k) = e_2(k)$, namely,

$$\frac{C+\delta}{u_k} - \underline{a}_{k-1} = \overline{a}_{k-1} - \frac{C-\delta}{u_k},$$

or

$$u_k = \frac{2C}{\underline{a}_{k-1} + \overline{a}_{k-1}}.$$

The optimal identification error is then

$$\begin{aligned} e_k &= \frac{(C+\delta)(\overline{a}_{k-1} + \underline{a}_{k-1})}{2C} - \underline{a}_{k-1} \\ &= \left(\frac{1}{2} + \frac{\beta}{2}\right)(\overline{a}_{k-1} + \underline{a}_{k-1}) - \underline{a}_{k-1} \\ &= \frac{1}{2}e_{k-1} + \frac{\beta}{2}(\overline{a}_{k-1} + \underline{a}_{k-1}). \end{aligned}$$

(2) We prove $\frac{\overline{a}_k}{\underline{a}_k} \geq \frac{1+\beta}{1-\beta}$ by induction. Suppose that $\frac{\overline{a}_{k-1}}{\underline{a}_{k-1}} \geq \frac{1+\beta}{1-\beta}$. Then we have $u_k \overline{a}_{k-1} \geq C + \delta$ and $u_k \underline{a}_{k-1} \leq C - \delta$, which, respectively, leads to $\frac{\overline{a}_k}{\underline{a}_k} = \frac{u_k \overline{a}_{k-1}}{C-\delta} \geq \frac{1+\beta}{1-\beta}$ in the case of $s_k = 1$, and $\frac{\overline{a}_k}{\underline{a}_k} = \frac{C+\delta}{u_k \underline{a}_{k-1}} \geq \frac{1+\beta}{1-\beta}$ in the case of $s_k = 0$. Thus, by the initial condition that $\varepsilon_0 \geq \frac{1+\beta}{1-\beta}$, we have $\frac{\overline{a}_k}{\underline{a}_k} \geq \frac{1+\beta}{1-\beta}$ for all $k \geq 1$.

By $\frac{\overline{a}_{k-1}}{\underline{a}_{k-1}} \geq \frac{1+\beta}{1-\beta}$, we have $u_k \underline{a}_{k-1} \leq C - \delta$ and $u_k \overline{a}_{k-1} \geq C + \delta$, which gives $\overline{a}_k = \overline{a}_{k-1}$ and $\frac{\underline{a}_k}{\underline{a}_{k-1}} = \frac{C-\delta}{u_k \underline{a}_{k-1}} \geq 1$ in the case of $s_k = 1$, and $\underline{a}_k = \underline{a}_{k-1}$ and $\frac{\overline{a}_k}{\overline{a}_{k-1}} = \frac{C+\delta}{u_k \overline{a}_{k-1}} \leq 1$ in the case of $s_k = 0$. Thus, $\{\underline{a}_k\}$ is monotonically increasing and $\{\overline{a}_k\}$ is monotonically decreasing.

Furthermore, by $\frac{a_k}{\underline{a}_{k-1}} \geq 1$ and $\frac{\overline{a}_{k-1}}{\overline{a}_k} \geq 1$, we obtain $\frac{\underline{a}_k \overline{a}_{k-1}}{\overline{a}_k \underline{a}_{k-1}} \geq 1$, i.e., $\frac{\overline{a}_{k-1}}{\underline{a}_{k-1}} \geq \frac{\overline{a}_k}{\underline{a}_k}$. Hence, $\left\{\frac{\overline{a}_k}{\underline{a}_k}\right\}$ is monotonically decreasing.

The dynamic expression (9.9) can be modified as

$$e_k = \frac{1}{2}(1-\beta)\, e_{k-1} + \beta \overline{a}_{k-1}, \tag{9.10}$$

or

$$e_k = \frac{1}{2}(1+\beta)\, e_{k-1} + \beta \underline{a}_{k-1}. \tag{9.11}$$

By taking $k \to \infty$ on both sides of (9.10) and (9.11), we obtain $\overline{a}(\infty) = \frac{C+\delta}{2\delta} e(\infty)$ and $\underline{a}(\infty) = \frac{C-\delta}{2\delta} e(\infty)$. This leads to $\lim_{k\to\infty} \frac{\overline{a}_k}{\underline{a}_k} = \frac{1+\beta}{1-\beta}$.

(3) From (9.9) it follows that the uncertainty is reducible if and only if

$$\beta(\overline{a}_{k-1} + \underline{a}_{k-1}) < e_{k-1} = \overline{a}(k-1) - \underline{a}_{k-1}.$$

This is equivalent to

$$\frac{\overline{a}_{k-1}}{\underline{a}_{k-1}} > \frac{1+\beta}{1-\beta}.$$

□

Theorem 9.12. *Let*

$$\alpha_1 = \frac{1}{2}(1-\beta), \quad \alpha_2 = \frac{1}{2}(1+\beta).$$

Then under the conditions and notation of Theorem 9.11*,*

(1) *for $k \geq 1$, the optimal identification error e_k is bounded by*

$$\begin{aligned} \alpha_1^k e_0 + \beta \frac{a(1-\alpha_1^k)}{\alpha_2} &\leq \alpha_1^k e_0 + \beta \frac{\overline{a}_{k-1}(1-\alpha_1^k)}{\alpha_2} \\ &\leq e_k \leq \alpha_2^k e_0 + \beta \frac{\underline{a}_{k-1}(1-\alpha_2^k)}{\alpha_1} \\ &\leq \alpha_2^k e_0 + \beta \frac{a(1-\alpha_2^k)}{\alpha_1}; \end{aligned} \tag{9.12}$$

(2) *let $\varepsilon_0 = e_0/2$ and $\varepsilon_0 > \varepsilon > \frac{\beta a}{\alpha_1} = \frac{2\beta a}{1-\beta}$. Then the time complexity $N(\varepsilon)$ for reducing the uncertainty from ε_0 to ε is bounded by*

$$\left\lceil \frac{\log \frac{\varepsilon - \frac{\beta a}{\alpha_2}}{\varepsilon_0 - \frac{\beta a}{\alpha_2}}}{\log \alpha_1} \right\rceil \leq N \leq \left\lceil \frac{\log \frac{\varepsilon - \frac{\beta a}{\alpha_1}}{\varepsilon_0 - \frac{\beta a}{\alpha_1}}}{\log \alpha_2} \right\rceil;$$

(3) *there exists an irreducible relative error*

$$\frac{2\beta}{1+\beta} \leq \frac{e(\infty)}{a} \leq \frac{2\beta}{1-\beta}; \tag{9.13}$$

(4) *the parameter estimation error is bounded by*

$$0 \le \frac{\overline{a}(\infty) - a}{\overline{a}(\infty)} \le \frac{2\beta}{1+\beta}, \quad 0 \le \frac{a - \underline{a}(\infty)}{\underline{a}(\infty)} \le \frac{2\beta}{1-\beta}. \tag{9.14}$$

Proof. We prove the assertions step by step as follows.

(1) From (9.10) and the monotonically decreasing property of $\overline{a}_k$, we have

$$e_k \ge \alpha_1^k e_0 + \frac{\delta \overline{a}_{k-1}}{C} \sum_{i=0}^{k-1} \alpha_1^i,$$

and from (9.11) and the monotonically increasing property of $\underline{a}_k$,

$$e_k \le \alpha_2^k e_0 + \frac{\delta \underline{a}_{k-1}}{C} \sum_{i=0}^{k-1} \alpha_2^i.$$

The results follow from $\sum_{i=0}^{k-1} \alpha_1^i = \frac{1-\alpha_1^k}{1-\alpha_1}$, $\sum_{i=0}^{k-1} \alpha_2^i = \frac{1-\alpha_2^k}{1-\alpha_2}$, $1-\alpha_1 = \alpha_2$, and $\underline{a}_k \le a \le \overline{a}_k$.

(2) From item (2) of Theorem 9.11, it follows that the error $e_k = \overline{a}_k - \underline{a}_k$ is monotonically decreasing. Thus, the upper bound on the time complexity is obtained by solving the inequality for the smallest N satisfying

$$e_N \le \alpha_2^N \varepsilon_0 + \frac{\beta a(1-\alpha_2^N)}{\alpha_1} \le \varepsilon.$$

Similarly, the lower bound can be obtained by calculating the largest N satisfying

$$\varepsilon \le \alpha_1^N \varepsilon_0 + \frac{\beta a(1-\alpha_1^N)}{\alpha_2} \le e_k.$$

(3) This follows from (9.12) and item (2) of Theorem 9.11, which implies the existence of $\lim_{t\to\infty} e_k$.

(4) From the last two lines of the proof of item (2) of Theorem 9.11, it follows that $\overline{a}(\infty) = \frac{C+\delta}{2\delta} e(\infty)$ and $\underline{a}(\infty) = \frac{C-\delta}{2\delta} e(\infty)$. This, together with (9.13), gives (9.14).

□

Remark 9.13. It is noted that the bounds in item (2) of Theorem 9.12 can be easily translated to sequential information bounds by replacing a with the on-line inequalities $\underline{a}_{k-1} \le a \le \overline{a}_{k-1}$.

Case 3: $y_k = au_k + \widetilde{\phi}_k'\widetilde{\theta}$

Let $u_k = \{u_\tau,\ \tau \le k\}$. Then $\|u_k\|_\infty$ is the maximum $|u_\tau|$ up to time k. Since we assume no information on $\widetilde{\theta}$, except that $\|\widetilde{\theta}\|_1 \le \eta$, it is clear that $\sup_{\|\widetilde{\theta}\|_1 \le \eta} |\widetilde{\phi}_k'\widetilde{\theta}| = \eta m_k$, where $m_k = \|\widetilde{\phi}_k\|_\infty$. Let $w_k = \widetilde{\phi}_k'\widetilde{\theta}$. Then

$$\{\widetilde{\phi}_k'\widetilde{\theta}:\ \|\widetilde{\theta}\|_1 \le \eta\} = \{w_k:\ |w_k| \le \eta m_k\}.$$

Theorem 9.14. *Suppose that $\underline{a}_0 > 0$, $u_{\max} \ge C/\underline{a}_0$, $\eta < \underline{a}_0$. Then*

(1) *the optimal input u_k, which minimizes the worst-case uncertainty at k, is given by the causal mapping from the available information at $k-1$:*

$$u_k = \frac{2C}{\underline{a}_{k-1} + \overline{a}_{k-1}}. \tag{9.15}$$

The optimal identification error at k satisfies the iteration equation

$$e_k = \frac{1}{2}e_{k-1} + \frac{\eta m_k}{2C}(\overline{a}_{k-1} + \underline{a}_{k-1}), \tag{9.16}$$

where $\overline{a}_k$ and $\underline{a}_k$ are updated by the rules

$$\begin{aligned} \overline{a}_k &= \overline{a}_{k-1}, \quad \underline{a}_k = \frac{C - \eta m_k}{u_k}, \quad \text{if } s_k = 1, \\ \underline{a}_k &= \underline{a}_{k-1}, \quad \overline{a}_k = \frac{C + \eta m_k}{u_k}, \quad \text{if } s_k = 0; \end{aligned}$$

(2) *the uncertainty is reducible if and only if $\overline{a}_{k-1} > \underline{a}_{k-1} + 2\eta$;*

(3) *for $k \ge 1$, the optimal identification error e_k is bounded by*

$$\begin{aligned} &\left(\prod_{j=1}^{k}\beta_1(j)\right)e_0 + \frac{\eta a}{C}\sum_{i=1}^{k} m_i \prod_{j=i+1}^{k}\beta_1(j) \\ &\le e_k \le \left(\prod_{j=1}^{k}\beta_2(j)\right)e_0 + \frac{\eta a}{C}\sum_{i=1}^{k} m_i \prod_{j=i+1}^{k}\beta_2(j), \end{aligned} \tag{9.17}$$

where $\beta_1(k) = \frac{1}{2}\left(1 - \frac{\eta m_k}{C}\right)$ and $\beta_2(k) = \frac{1}{2}\left(1 + \frac{\eta m_k}{C}\right)$;

(4) *let $\varepsilon_0 = e_0/2$ and $\varepsilon_0 > \varepsilon > \frac{2\eta\overline{a}(0)}{\underline{a}_0 - \eta}$. Also, denote $\beta_1 = \frac{1}{2}\left(1 - \frac{\eta}{\underline{a}_0}\right)$, $\beta_2 = \frac{1}{2}\left(1 + \frac{\eta}{\underline{a}_0}\right)$. Then the time complexity $N(\varepsilon)$ for reducing the uncertainty from ε_0 to ε is bounded by*

$$\left\lceil \frac{\log \frac{\varepsilon - \frac{\eta a}{\overline{a}_0\beta_2}}{\varepsilon_0 - \frac{\eta a}{\overline{a}_0\beta_2}}}{\log \beta_1} \right\rceil \le N(\varepsilon) \le \left\lceil \frac{\log \frac{\varepsilon - \frac{\eta a}{\underline{a}_0\beta_1}}{\varepsilon_0 - \frac{\eta a}{\underline{a}_0\beta_1}}}{\log \beta_2} \right\rceil. \tag{9.18}$$

Proof. The proof is arranged as follows.

(1) The results follow from the definition of m_k and Theorem 9.12, with δ replaced by ηm_k.

(2) From (9.16) and (9.15), it follows that the uncertainty is reducible if and only if $\frac{\eta m_k}{u_k} < \frac{1}{2} e_{k-1} = \frac{1}{2}(\overline{a}_{k-1} - \underline{a}_{k-1})$. This is equivalent to $\eta < \frac{1}{2}(\overline{a}_{k-1} - \underline{a}_{k-1})$ or $\overline{a}_{k-1} > \underline{a}_{k-1} + 2\eta$, since $\frac{m_k}{u_k} \geq 1$.

(3) By (9.16), we have

$$e_k = \frac{1}{2}\left(1 + \frac{\eta_k}{C}\right) e_{k-1} + \frac{\eta m_k}{C} \underline{a}_{k-1} \tag{9.19}$$

and

$$e_k = \frac{1}{2}\left(1 - \frac{\eta m_k}{C}\right) e_{k-1} + \frac{\eta m_k}{C} \overline{a}_{k-1}. \tag{9.20}$$

Furthermore, from $\underline{a}_k \leq a \leq \overline{a}_k$ for all $k \geq 0$,

$$e_k \leq \beta_2(k) e_{k-1} + \frac{\eta m_k}{C} a$$

and

$$e_k \geq \beta_1(k) e_{k-1} + \frac{\eta m_k}{C} a.$$

Then, the inequalities in (9.17) can be obtained by iterating the above two inequalities in k.

(4) Since for all $k \geq 1$, $\overline{a}_0 \geq \overline{a}_k \geq \underline{a}_k \geq \underline{a}_0$,

$$u_k = \frac{2C}{\underline{a}_{k-1} + \overline{a}_{k-1}} \leq \frac{C}{\underline{a}_0},$$

which implies that $\frac{C}{\overline{a}_0} \leq u_k \leq \frac{C}{\underline{a}_0}$. This leads to

$$\beta_1(k) \geq \beta_1 = \frac{1}{2}(1 - \frac{\eta}{\underline{a}_0})$$

and

$$\beta_2(k) \leq \beta_2 = \frac{1}{2}\left(1 + \frac{\eta}{\underline{a}_0}\right).$$

Hence,

$$\beta_1 e_{k-1} + \frac{\eta a}{\overline{a}_0} \leq e_k \leq \beta_2 e_{k-1} + \frac{\eta a}{\underline{a}_0} \quad \text{for all } k \geq 1. \tag{9.21}$$

As a result, the inequalities of Theorem 9.12 can be adopted here to get (9.18).

□

Note that $\beta_2(k) \geq \beta_1(k)$ and $\beta_1(k) + \beta_2(k) = 1$; and $\beta_1 \to \beta_2$ as $\eta \to 0$, uniformly in k.

9.5 Identification Using Combined Deterministic and Stochastic Methods

This section highlights the distinctive underlying principles used in designing inputs and deriving posterior uncertainty sets in the stochastic and deterministic information frameworks.

In the deterministic worst-case framework, the information on noise is limited to its magnitude bound. Identification properties must be evaluated against worst-case noise sample paths. As shown earlier, the input is designed on the basis of choosing an optimal worst-case testing point (a hyperplane) for the prior uncertainty set. When the prior uncertainty set is large, this leads to an exponential rate of uncertainty reduction. However, when the uncertainty set is close to its irreducible limits due to disturbances or unmodeled dynamics, its rate of uncertainty reduction decreases dramatically due to its worst-case requirements. Furthermore, when the disturbance magnitude is large, the irreducible uncertainty will become too large for identification error bounds to be practically useful.

In contrast, in a stochastic framework, noise is modeled by a stochastic process and identification errors are required to be small with a large probability. Binary sensor identification in this case relies on the idea of averaging. Typically, in identification under stochastic setting, the input is designed to provide sufficient excitation for asymptotic convergence, rather than fast initial uncertainty reduction. Without effective utilization of prior information in designing the input during the initial time interval, the initial convergence can be slow. This is especially a severe problem in binary sensor identification since a poorly designed input may result in a very imbalanced output of the sensor in its 0 or 1 values, leading to a slow convergence rate. In the case of large prior uncertainty, the selected input may result in nonswitching at the output, rendering the stochastic binary-sensor identification inapplicable. On the other hand, averaging disturbances restores estimation consistency and overcomes a fundamental limitation of the worst-case identification.

Consequently, it seems a sensible choice of using the deterministic framework initially to achieve fast uncertainty reduction when the uncertainty set is large, then using the stochastic framework to modify estimation consistency. In fact, we shall demonstrate by an example that these two frameworks complement each other precisely, in the sense that when one framework fails, the other starts to be applicable.

9.5.1 Identifiability Conditions and Properties under Deterministic and Stochastic Frameworks

We first establish identifiability conditions of the two frameworks for a gain system

$$y_k = au_k + d_k, \quad k = 1, 2, \ldots, \tag{9.22}$$

where $\{d_k\}$ is a sequence of disturbances, and a is an unknown parameter. The prior information on a is given by $a \in [\underline{a}, \overline{a}]$, with $0 < \underline{a} \leq \overline{a} < \infty$. $u_k > 0$ is the input. The output y_k is measured by a binary-valued sensor with threshold C.

Deterministic Framework.

The idea of deterministic framework is to reduce the parameter uncertainty based on the bound of disturbances. Denote $r_k = \underline{a}_k/\overline{a}_k$ as the relative error.

Starting from the initial uncertainty $\Omega_0 = [\underline{a}, \overline{a}]$ and input $u_0 = u^*$, we check the output of binary sensor. If $s_1 = 0$, which means $au^* + d_1 > 0$, we obtain $au^* + \delta \geq au^* + d_1 > C$. Hence, $a > (C - \delta)/u^*$ and $e_1 = \overline{a} - (C - \delta)/u^* < e_0$. This means the parameter bound is reducible if $\underline{a} < (C - \delta)/u^*$. Otherwise, we have $s_1 = 1$. Then, $au^* - \delta \leq au^* + d_1 \leq C$; hence, $a \leq (C + \delta)/u^*$ and $e_1 = (C + \delta)/u^* - \underline{a} < e_0$ if $\overline{a} > (C + \delta)/u^*$. So, the parameter bound is reducible if

$$\underline{a} < \frac{C - \delta}{u^*} \quad \text{and} \quad \overline{a} > \frac{C + \delta}{u^*}$$

in the worst-case sense, or equivalently,

$$r_0 < \frac{C - \delta}{C + \delta} := \Delta. \tag{9.23}$$

Furthermore, by the above analysis, we arrive at the new uncertainty set

$$e_1 = \max\left\{\overline{a} - \frac{C - \delta}{u^*}, \frac{C + \delta}{u^*} - \underline{a}\right\}$$

in the worst-case sense. The uncertainty set is minimized at the optimal input

$$u_1^* = \frac{2C}{\underline{a} + \overline{a}} \quad \text{and} \quad e_1^* = \frac{(1 + \beta)\overline{a} - (1 - \beta)\underline{a}}{2} \tag{9.24}$$

with $\beta = \delta/C$.

The one-step optimal input design and parameter error was introduced in [111]. This, however, is not an overall optimal design if N steps are considered. The N-step optimal input design was developed in [14].

Theorem 9.15 [14]. *For binary observations with threshold C, the optimal parameter bound is*

$$e_N^* = 2\beta \frac{\overline{a}(1 + \beta)^{(2^N - 1)} - \underline{a}(1 - \beta)^{(2^N - 1)}}{(1 + \beta)^{(2^N)} - (1 - \beta)^{(2^N)}} \tag{9.25}$$

and the optimal inputs are

$$u_k^* = \frac{C}{\widetilde{a}_{k|N}}, \quad k = 1, 2, \ldots, N,$$

where

$$\widetilde{a}_{k|N} = \frac{\overline{a}_{k-1}(1+\beta)^{(2^{N-k}-1)} + \underline{a}_{k-1}(1-\beta)^{(2^{N-k}-1)}}{(1+\beta)^{(2^{N-k})} + (1-\beta)^{(2^{N-k})}}.$$

Example 9.16. For system (9.22) with $C = 40$, $\delta = 0.5$, $\underline{a} = 1$, and $\overline{a} = 10$, the optimal error provided by (9.25) is shown in Figure 9.1. It is shown that at first the uncertainty is reduced very fast, but uncertainty reduction gradually slows down toward an irreducible error bound.

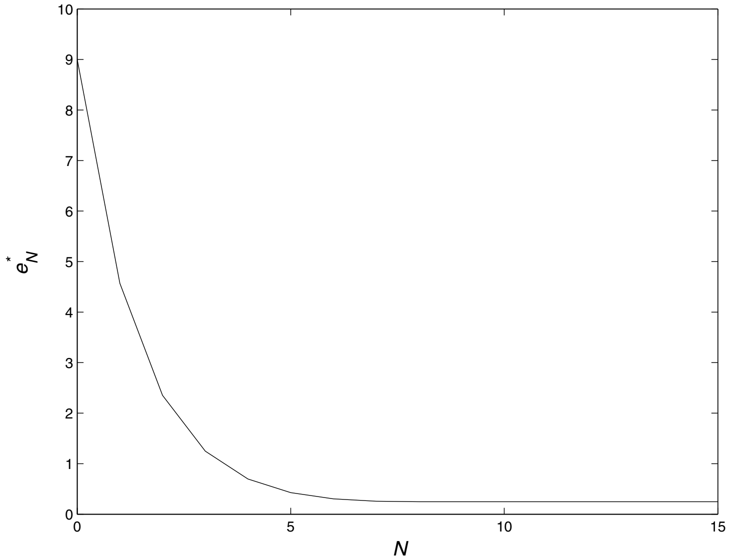

FIGURE 9.1. Optimal parameter error

Stochastic Framework

The essence of stochastic framework is to utilize the probabilistic properties of disturbances. Define the empirical measure $\xi_N^0 = \sum_{k=1}^{N} s_k/N$. If there exists u^* such that $C - au^*$ is on the support of $F(\cdot)$, which means

$$-\delta < C - au^* < \delta, \tag{9.26}$$

then ξ_N^0 is the empirical measure of $F(\cdot)$ at $C - au^*$, and

$$\xi_N^0 \to F(C - au^*), \quad \text{w.p.1.} \tag{9.27}$$

(A9.2) The noise $\{d_k\}$ is a sequence of i.i.d. random variables bounded by $|d_k| \le \delta$ whose distribution function $F(x)$, $x \in (-\delta, \delta)$, and its inverse $F^{-1}(\cdot)$ are twice continuously differentiable in $(-\delta, \delta)$ and known.

Throughout the rest of the chapter, we assume Assumption (A9.2) also holds. Note that F is a monotone function in view of Assumption (A9.2). If a is bounded, then there exists $z > 0$ such that $p = F(C - au^*)$ is bounded by

$$z < p < 1 - z. \tag{9.28}$$

Since $F(\cdot)$ is not invertible at 0 and 1, we modify ξ_N^0 to avoid these points as in (3.1):

$$\xi_N = \begin{cases} \xi_N^0, & \text{if} \quad z \le \xi_N^0 \le 1 - z, \\ z, & \text{if} \quad \xi_N^0 < z, \\ 1 - z, & \text{if} \quad \xi_N^0 > 1 - z. \end{cases} \tag{9.29}$$

As shown in Chapter 3, $\xi_N \to p$ w.p.1. Define

$$\widehat{a}_N = (C - F^{-1}(\xi_N))/u^*. \tag{9.30}$$

Then

$$\widehat{a}_N \to a \text{ w.p.1.}$$

For $a \in [\underline{a}, \overline{a}]$, the identifiability condition (9.26) becomes

$$-\delta < C - \overline{a}u^* \le C - \underline{a}u^* < \delta.$$

So for a given threshold C, u^* can be chosen to construct the estimation algorithm if and only if

$$r_0 > \Delta, \tag{9.31}$$

which complements exactly (9.23) for the deterministic framework. Under (9.31) and $C > \delta$, the admissible input set is

$$u^* \in \Gamma = \left(\frac{C - \delta}{\underline{a}}, \frac{C + \delta}{\overline{a}} \right). \tag{9.32}$$

By Chapter 6, for a given u^*, the optimal CR lower bound with binary-valued observations is

$$\eta_N^*(a, u^*) = E\left(\widehat{a}_N^* - a\right)^2 = \frac{F(C - au^*)(1 - F(C - au^*))}{N(u^*)^2 f^2(C - au^*)} \tag{9.33}$$

and $N(\eta_N - \eta_N^*(a, u^*)) = N[E(\widehat{a}_N - a)^2 - \eta_N^*] \to 0$ as $N \to \infty$, which means the algorithm (9.30) of the stochastic framework is asymptotically efficient.

Remark 9.17. The foregoing analysis indicates that the identifiability condition for the deterministic framework is that $r_0 < \Delta$ in the worst case and $r_0 > \Delta$ for the stochastic framework. Due to the strict inequalities, there is a dividing line $r_0 = \Delta$ between the two frameworks under binary observations. A key problem in combining the two frameworks is to find a way to connect the two sets of identifiability regions.

9.5.2 Combined Deterministic and Stochastic Identification Methods

In this subsection, we introduce a method to connect the two frameworks and develop the criteria for switching from one framework to another.

Connection of Two Frameworks by Input Design

Since there is no intersection between the two identifiability sets (9.23) and (9.31), one cannot design a strategy to switch from one framework to another. Consequently, it is necessary to find an approach to connect the sets. Here, we modify the stochastic methods by using two input values, rather than one. Since each input value creates one identifiability set, by choosing the inputs appropriately, we can create a scenario that these two sets collectively intersect to the identifiability set of the deterministic method.

For the initial uncertainty $[\underline{a}, \overline{a}]$, let $b \in (\underline{a}, \overline{a})$. Then, b divides the interval into two parts, $[\underline{a}, b]$ and $(b, \overline{a}]$. For $a \in [\underline{a}, b]$, the identifiability condition (9.31) becomes $\underline{a}/b > \Delta$. Similarly, for $a \in (b, \overline{a}]$, the requirement is $b/\overline{a} > \Delta$. Since

$$\max_b \min \left\{ \frac{\underline{a}}{b}, \frac{b}{\overline{a}} \right\} = \sqrt{\frac{\underline{a}}{\overline{a}}}$$

with $b^* = \sqrt{\underline{a}\overline{a}}$, the parameter can be estimated if $r_0 > \Delta^2$. Since $\Delta = (C - \delta)/(C + \delta) < 1$, we have $\Delta^2 < \Delta$; thus, there is an intersection between the identifiability sets of two frameworks.

This analysis indicates that it is possible to connect the two frameworks if two input values are used for the stochastic framework. We discuss next the switching strategies. This will be done by using convergence speeds. We first use an example to illustrate the basic ideas.

Example 9.18. Consider the one-step optimal worst-case error in Theorem 9.15

$$\frac{e_1^*}{e_0^*} = \frac{(1+\beta)\overline{a} - (1-\beta)\underline{a}}{2e_0} = \frac{1-\beta}{2} + \frac{\beta}{1-r_0},$$

which decreases with r_0. For the same system as in Example 9.16, the ratio is plotted as a function of r_0 in Figure 9.2 with $\beta = 0.2$. We can see that the ratio goes to 1 when r_0 approaches $(1-\beta)/(1+\beta)$, which means the uncertainty is almost irreducible.

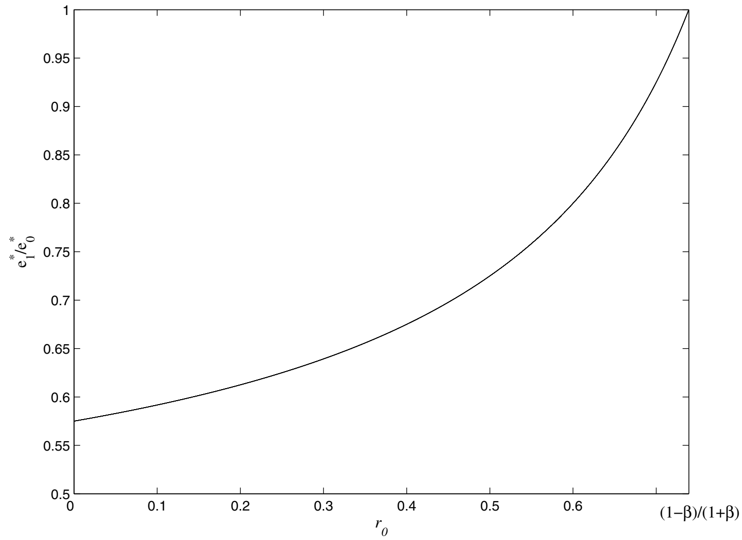

FIGURE 9.2. Optimal parameter reduction ratio

Consider the identifiability condition of the stochastic framework $r_0 > \Delta$. The convergence speed may be slow in the worst case as r_0 is small and close to Δ. The optimal covariance of the stochastic method with threshold C and constant input u^* is

$$\eta_N^*(a,u) = \frac{F(C-au^*)(1-F(C-au^*))}{N(u^*)^2 f^2(C-au^*)}.$$

Let

$$\eta^*(\Omega_0,u^*) = \sup_{a\in\Omega_0} \frac{F(C-au^*)(1-F(C-au^*))}{(u^*)^2 f^2(C-au^*)},$$

and

$$\eta^*(\Omega_0) = \inf_{u^*} \eta^*(\Omega_0,u^*). \tag{9.34}$$

Then, the optimal convergence speed by designing an optimal input value can be derived as

$$\eta_N^*(\Omega_0) = \eta^*(\Omega_0)/N. \tag{9.35}$$

For Ω_0, if we use the $\eta_N^*(\Omega_0)$, we can first design identification algorithms for $\Psi_1 = [\underline{a}, b^*]$ and $\Psi_2 = [b^*, \overline{a}]$, and then calculate $\eta_N^*(\Psi_1)$ and $\eta_N^*(\Psi_2)$. Hence, the switch time N_s can be decided by the following rule:

$$N_s = \min_N \left\{ \varepsilon_{2N}^2 > \min\{\eta_N^*([\underline{a}, \sqrt{\underline{a}\overline{a}}]), \eta_N^*([\sqrt{\underline{a}\overline{a}}, \overline{a}])\} \right\}. \tag{9.36}$$

With this switching rule, the joint identification algorithm can be constructed as follows:

1. In the case of $r_k < \Delta^2$, apply the deterministic method.
2. Denote the first time that $r_k \geq \Delta^2$ as K, and calculate N_s by (9.36) with the information of parameter uncertainty lower and upper bounds at that time K.
3. Keep using deterministic methods for another $N_s - 1$ steps. Then get the parameter lower and upper bounds, namely, $\underline{a}_s$ and $\overline{a}_s$.
4. Switch to the stochastic method.

9.5.3 Optimal Input Design and Convergence Speed under Typical Distributions

We now solve (9.34) concretely for some typical noise distribution functions. We will derive specific expressions for the uniform distribution and truncated normal distribution. For other distributions, similar methods can be used, although derivation details may vary. For simplification, let $\eta_N^* = \eta_N^*(\Omega_0)$ and $\eta^*(u^*) = \eta^*(\Omega_0, u^*)$.

Uniform Distribution
Suppose that the density function of d_k is $f(x) = 1/(2\delta)$ for the support set (i.e., strictly positive) in $(-\delta, \delta)$. Then, $F(x) = \frac{\delta+x}{2\delta}$. We have

$$\begin{aligned} \eta^*(u^*) &= \sup_{a\in\Omega_0} \{\delta^2 - (C - au^*)^2\}, \\ &= \begin{cases} \delta^2, & \text{if } u^* \in \Gamma_1 = (\frac{C}{\overline{a}}, \frac{C}{\underline{a}}), \\ \delta^2 - (C - \overline{a}u^*)^2, & \text{if } u^* < \frac{C}{\overline{a}}, \\ \delta^2 - (C - \underline{a}u^*)^2, & \text{if } u^* > \frac{C}{\underline{a}}. \end{cases} \end{aligned} \tag{9.37}$$

Theorem 9.19. *Suppose d_k has a uniform distribution on $(-\delta, \delta)$. Then for Ω_0, η^* defined in (9.34) can be expressed as*

$$\eta^* = \begin{cases} \frac{\delta^2 \overline{a}^2}{(C+\delta)^2}, & \text{if } r \leq \frac{C}{C+\delta}, \\ \frac{\delta^2 \underline{a}^2}{C^2}, & \text{if } r > \frac{C}{C+\delta} \text{ and } C > \delta, \\ \frac{(\overline{a}-\underline{a})[(C+\delta)\underline{a}-(C-\delta)\overline{a}]}{C+\delta}, & \text{if } r > \frac{C}{C+\delta} \text{ and } C \leq \delta, \end{cases} \tag{9.38}$$

and the optimal input can be derived concretely by the above cases, respectively.

Proof. By (9.32), the feasible input set is $u^* \in \Gamma$. The set is nonempty if and only if $r > \Delta$ in case of $C > \delta$.

Case (i): In case of $\Delta < r \le \frac{C-\delta}{C}$, since $\frac{C-\delta}{C} \le \frac{C}{C+\delta}$, we have $\frac{C-\delta}{\underline{a}} \ge \frac{C}{\overline{a}}$ and $\frac{C+\delta}{\overline{a}} \le \frac{C}{\underline{a}}$, namely, $\Gamma \subset \Gamma_1 = (\frac{C}{\overline{a}}, \frac{C}{\underline{a}})$. So for $\forall u^* \in \Gamma$, there exists $a \in \Omega_0$ such that $a = C/u^*$, which induces $\eta^*(u^*) = \delta^2$. Hence,

$$\eta^* = \inf_{u^* \in \Gamma} \frac{\delta^2}{(u^*)^2} = \frac{\delta^2 \overline{a}^2}{(C+\delta)^2}$$

with $u^* = (C+\delta)/\overline{a}$.

Case (ii): In case of $\frac{C-\delta}{C} < r \le \frac{C}{C+\delta}$, we have $\frac{C-\delta}{\underline{a}} < \frac{C}{\overline{a}}$ and $\frac{C+\delta}{\overline{a}} \le \frac{C}{\underline{a}}$. For $u^* \in \Gamma_2 = (\frac{C}{\overline{a}}, \frac{C+\delta}{\overline{a}})$, we have $\eta^*(u^*) = \delta^2$. So

$$\inf_{u^* \in \Gamma_2} \frac{\delta^2}{(u^*)^2} = \frac{\delta^2 \overline{a}^2}{(C+\delta)^2}.$$

For $u^* \in \Gamma_3 = (\frac{C-\delta}{\underline{a}}, \frac{C}{\overline{a}}]$, notice that $C - au^* > C - a\frac{C}{\overline{a}} \ge 0$, which means $a \le C/u^*$. So $\eta^*(u^*) = \delta^2 - (C - \overline{a}u^*)^2$ for $u^* \in \Gamma_3$. Since

$$\inf_{u^* \in \Gamma_3} \frac{\delta^2 - (C - \overline{a}u^*)^2}{(u^*)^2} = \inf_{u^* \in \Gamma_3} \left\{ \frac{\delta^2 - C_1^2}{(u^*)^2} + \frac{2\overline{a}C}{u^*} - \overline{a}^2 \right\}$$

and $C > \delta$, $\frac{\delta^2 - C_1^2}{(u^*)^2} + \frac{2\overline{a}C}{u^*} - \overline{a}^2$, as a function of $1/u^*$, is symmetric about $1/u^* = \frac{\overline{a}C}{C^2 - \delta^2}$.

Since

$$r = \frac{\underline{a}}{\overline{a}} \le \frac{C}{C+\delta} \le \frac{C^2+\delta^2}{C(C+\delta)},$$

we have

$$\left(\frac{\underline{a}}{C-\delta} - \frac{\overline{a}C}{C^2-\delta^2} \right) - \left(\frac{\overline{a}C}{C^2-\delta^2} - \frac{\overline{a}}{C} \right) = \frac{\underline{a}C(C+\delta) - \overline{a}(C^2+\delta^2)}{C(C^2-\delta^2)} \le 0.$$

As a result,

$$\frac{\delta^2 - C_1^2}{(u^*)^2} + \frac{2\overline{a}C}{u^*} - \overline{a}^2$$

is minimized at $u^* = C/\overline{a}$ on Γ_3, namely,

$$\inf_{u^* \in \Gamma_3} \left\{ \frac{\delta^2 - C_1^2}{(u^*)^2} + \frac{2\overline{a}C}{u^*} - \overline{a}^2 \right\} = \frac{\delta^2 \overline{a}^2}{C^2}.$$

Hence, we have

$$\eta^* = \min \left\{ \frac{\delta^2 \overline{a}^2}{(C+\delta)^2}, \frac{\delta^2 \overline{a}^2}{C^2} \right\} = \frac{\delta^2 \overline{a}^2}{(C+\delta)^2}.$$

Case (iii): In the case of $r > \frac{C}{C+\delta}$, we have $\frac{C-\delta}{\underline{a}} < \frac{C}{\overline{a}}$ and $\frac{C+\delta}{\overline{a}} > \frac{C}{\underline{a}}$. For $u^* \in \Gamma_1$, we have $\eta^*(u^*) = \delta^2$ and $\eta^*(u^*) = \delta^2 - (C - \overline{a}u^*)^2$ for $u^* \in \Gamma_3$. So

$$\inf_{u^* \in \Gamma_1} \frac{\delta^2}{(u^*)^2} = \frac{\delta^2 \underline{a}^2}{C^2} \quad \text{and} \quad \inf_{u^* \in \Gamma_3} \frac{\delta^2 - (C - \overline{a}u^*)^2}{(u^*)^2} = \frac{\delta^2 \overline{a}^2}{C^2}.$$

For $u^* \in \Gamma_4 = [C/\underline{a}, \frac{C+\delta}{\overline{a}})$, note that $C - au^* \le C - a\frac{C}{\underline{a}} \le 0$, which means $a \ge C/u^*$. So $\eta^*(u^*) = \delta^2 - (C - \underline{a}u^*)^2$ for $u^* \in \Gamma_4$. The minimization problem is

$$\inf_{u^* \in \Gamma_4} \frac{\delta^2 - (C - \underline{a}u^*)^2}{(u^*)^2} = \inf_{u^* \in \Gamma_4} \left\{ \frac{\delta^2 - C_1^2}{(u^*)^2} + \frac{2\underline{a}C}{u^*} - \underline{a}^2 \right\}.$$

Since $C > \delta$, $\frac{\delta^2 - C_1^2}{(u^*)^2} + \frac{2\underline{a}C}{u^*} - \underline{a}^2$, as a function of $1/u^*$, is symmetric about $1/u^* = \frac{\underline{a}C}{C^2 - \delta^2}$. Since

$$r > \frac{C}{C+\delta} \ge \frac{C-\delta}{C} \ge \frac{C(C-\delta)}{C^2+\delta^2},$$

we have

$$\left(\frac{\underline{a}}{C} - \frac{\underline{a}C}{C^2 - \delta^2}\right) - \left(\frac{\underline{a}C}{C^2 - \delta^2} - \frac{\overline{a}}{C+\delta}\right)$$
$$= \frac{\overline{a}C(C-\delta) - \underline{a}(C^2 + \delta^2)}{C(C^2 - \delta^2)} \ge 0.$$

As a result,

$$\frac{\delta^2 - C_1^2}{(u^*)^2} + \frac{2\underline{a}C}{u^*} - \underline{a}^2$$

is minimized at $u^* = C/\underline{a}$ on Γ_4, namely,

$$\inf_{u^* \in \Gamma_4} \left\{ \frac{\delta^2 - C_1^2}{(u^*)^2} + \frac{2\underline{a}C}{u^*} - \underline{a}^2 \right\} = \frac{\delta^2 \underline{a}^2}{C^2}.$$

Hence, $\eta^* = \delta^2 \underline{a}^2 / C^2$.

The proof for $C < \delta$ is similar and omitted. □

Truncated Normal Distribution

Suppose d_k has a truncated normal distribution with probability density function

$$f_\sigma(x) = \frac{\frac{1}{\sigma}\lambda\left(\frac{x}{\sigma}\right)}{\Lambda\left(\frac{\delta}{\sigma}\right) - \Lambda\left(\frac{-\delta}{\sigma}\right)},$$

where $x \in (-\delta, \delta)$, $\lambda(\cdot)$ is the probability density function of the standard normal distribution, and $\Lambda(\cdot)$ its cumulative distribution function. Here, we discuss the case of $\sigma = 1$; general cases can be derived similarly. Then, we have the density function

$$f(x) = \frac{\lambda(x)}{\Lambda(\delta) - \Lambda(-\delta)}$$

and the distribution function given by

$$F(x) = \frac{\Lambda(x) - \Lambda(-\delta)}{\Lambda(\delta) - \Lambda(-\delta)}.$$

Denote

$$\lambda_1(x) = \lambda(x)(1 - 2\Lambda(x)),$$
$$\lambda_2(x) = (\Lambda(x) - \Lambda(-\delta))(\Lambda(\delta) - \Lambda(x)),$$

and

$$G(x) = \frac{\lambda_2(x)}{\lambda^2(x)}.$$

Hence,

$$\eta^*(u^*) = \sup_{C-\overline{a}u^* \leq x \leq C-\underline{a}u^*} G(x)$$

and

$$\eta^* = \inf_{u^*} \eta^*(u^*).$$

First, we analyze the property of $G(x)$. The derivative of $G(x)$ can be written as

$$G'(x) = \frac{\lambda_1(x) + 2x\lambda_2(x)}{\lambda^2(x)}.$$

Let

$$g_1(x) = \lambda_1(x) + 2x\lambda_2(x).$$

Then, we have $g_1(0) = 0$ and $g_1(\delta) = \lambda(\delta)(1 - 2\Lambda(\delta)) < 0$.

Note that

$$g_2(x) = g_1'(x) = x\lambda_1(x) - 2\lambda^2(x) + 2\lambda(x).$$

Then

$$g_2(\delta) = \delta\lambda(\delta)(1 - 2\Lambda(\delta)) - 2\lambda^2(\delta) < 0$$

and

$$g_2(0) = 2\left(\Lambda(\delta) - \frac{1}{2}\right)^2 - 2\lambda^2(0) < 0$$

in the case of $\Lambda(\delta) < \frac{1}{2} + \lambda(0)$, and $g_2(0) \geq 0$ in the case of

$$\Lambda(\delta) \geq \frac{1}{2} + \lambda(0).$$

Lemma 9.20. *$g_2(x) < 0$ on $(0, \delta)$ for $\Lambda(\delta) \le \frac{1}{2} + \lambda(0)$. And for $\Lambda(\delta) > \frac{1}{2} + \lambda(0)$, there exists exactly one $x_2 \in (0, \delta)$ such that $g_2(x_2) = 0$, $g_2(x) > 0$ on $(0, x_2)$, and $g_2(x) < 0$ on (x_2, δ).*

Theorem 9.21. *$G'(x) < 0$ on $(0, \delta)$ for $\Lambda(\delta) \le \frac{1}{2} + \lambda(0)$. In addition, for $\Lambda(\delta) > \frac{1}{2} + \lambda(0)$, there exists $x_3 \in (0, \delta)$ such that $G_1'(x_3) = 0$, $G'(x) > 0$ on $(0, x_3)$, and $G'(x) < 0$ on (x_3, δ).*

Proof. Note that $G'(x) = g(x)/\lambda^2(x)$, so we need only prove the same conclusion for $g_1(x)$. By Lemma 9.20, $g_2(x) < 0$ on $(0, \delta)$ for $\Lambda(\delta) \le \frac{1}{2} + \lambda(0)$. In addition, $g_1(0) = 0$, and we have $g_1(x) < 0$ on $(0, \delta)$. For $\Lambda(\delta) > \frac{1}{2} + \lambda(0)$, $g_2(x) > 0$ on $(0, x_2)$ and $g_2(x) < 0$ on (x_2, δ) by Lemma 9.20, so $g_1(x_2) > g_1(0) = 0$. Since $g_1(\delta) < 0$ and $g_2(x) < 0$ on (x_2, δ), the second part is true. □

Here, we only derive the case of $C > \delta$, and $\Lambda(\delta) \le \frac{1}{2} + \lambda(0)$. We can discuss other cases similarly. Recall (9.32); the feasible input set is $u^* \in \Gamma = \left(\frac{C-\delta}{\underline{a}}, \frac{C+\delta}{\overline{a}}\right)$ and the set is nonempty if and only if $r > \Delta$. Then, we have the following theorem.

Theorem 9.22 *Suppose d has a truncated normal distribution on $(-\delta, \delta)$. Then for Ω_0, η^* defined in (9.34) can be expressed as*

$$\eta^* = \begin{cases} \frac{G(0)\overline{a}^2}{(C+\delta)^2}, & \text{if } \Delta < r \le \frac{C-\delta}{C}, \\ \min\{\frac{G(0)\overline{a}^2}{(C+\delta)^2}, H(C - \frac{C-\delta}{\underline{a}}\overline{a})\overline{a}^2\}, & \text{if } \frac{C-\delta}{C} < r \le \frac{C}{C+\delta}, \\ \min\{\frac{G(0)\underline{a}^2}{C^2}, H(0)\overline{a}^2, H(C - \frac{C-\delta}{\underline{a}}\overline{a})\overline{a}^2, \underline{a}^2 H(C - \frac{C+\delta}{\overline{a}}\underline{a})\}, & \\ & \text{if } r > \frac{C}{C+\delta}, \end{cases} \tag{9.39}$$

where

$$H(t) = \frac{\lambda_2(t)}{(C-t)^2\lambda^2(t)}.$$

For system (9.22) with $C = 40$, $\underline{a} = 1$, $\overline{a} = 50$, and the actual parameter $a = 15$. The disturbance has a uniform distribution on $(-\delta, \delta)$ with $\delta = 6$, by the algorithm developed in Section 9.5.2:

We have $K = 2$, $\underline{a}_K = 9.8$, and $\overline{a}_K = 15$. Then, we calculate $N_s = 1$ by (9.36). We turn to stochastic method and get $\widetilde{a}$.

It is shown that the parameter uncertainty is reduced to a certain bound using the deterministic method in the first stage and convergent to the real parameter using the stochastic method afterwards.

9.6 Notes

The material in this chapter is derived mostly from [111]. This chapter presents input design, uncertainty reduction rates, and time complexity

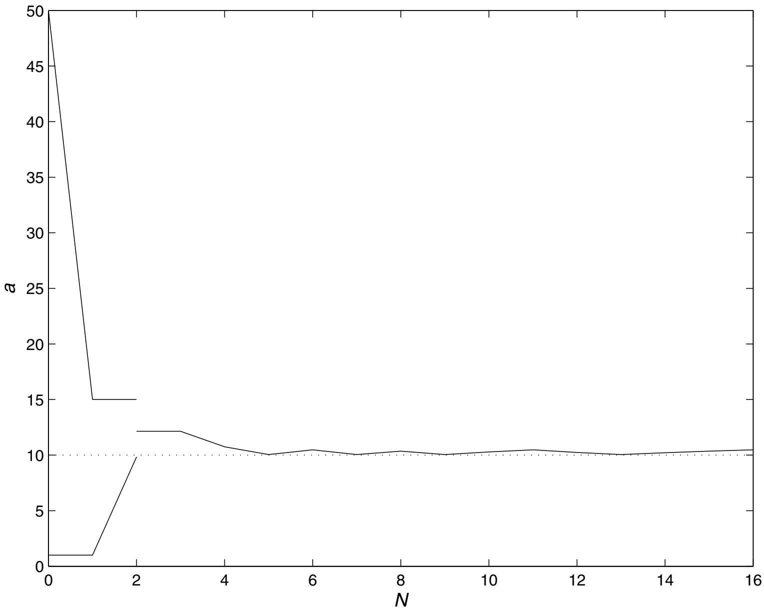

FIGURE 9.3. Simulation on the combined deterministic and stochastic identification methods

for system identification under binary-valued observations. This chapter deals with nonstatistical information from the observed data. We show that to enhance the nonstatistical information, the input must be properly designed.

Our input design is based on the idea of one-step optimal design: From the current uncertainty set on the unknown parameter, we select the best input value of the input such that the next uncertainty set can be maximally reduced, assuming no further information toward the future. Casini, Garulli, and Vicino have shown in [14] that if one has additional information on the number N of remaining steps toward the end of the identification data window, a better input design can be achieved. A dynamic programming method was introduced to optimize such an input design. It can be shown that in that case, the one-step optimal input design employed in Section 9.4 is no longer optimal for this N-step optimal input design. On the other hand, to achieve convergence with growing data sizes (namely, $N \to \infty$, rather than a fixed integer from the outset), the one-step design is a simple choice to achieve exponential convergence toward the irreducible uncertainty set.

The deterministic approaches are subject to an irreducible identification error; hence convergence is lost. They work well when the magnitude of the noise error bounds is relatively small since the irreducible set is a function of the size of the noise. Also, the input design can achieve exponential convergence rates toward the irreducible set, which is much faster than the polynomial rates of convergence in a stochastic framework. On the other hand, stochastic information in the data can produce a convergent estimator. A combined identification algorithm that employs the input design first to reduce the parameter uncertainty set exponentially, followed by a statistical averaging approach to achieve convergence with periodic inputs, seems to be the best choice in overcoming the shortcomings of each individual framework.