Chapter 8
System Identification

Mathematical models are essential to modern science and engineering, and have been very successful in advancing the technology that has had profound impact to our society. A serious question to be addressed in this chapter is how to obtain the model for a given physical process. There are two basic approaches. The first one is based on principles in physics or other sciences. The inverted pendulum in Chap. 1 provides an example to this approach. Its advantage lies in its capability to model nonlinear systems and preservation of the physical parameters. However, this approach can be costly and time consuming. The second approach is based on input and output data to extrapolate the underlying system model. This approach treats the system as a black box and is only concerned with its input–output behaviors. While experiments need to be carried out and restrictions on input signals may apply, the second approach overcomes the weakness of the first approach.

This chapter examines the input/output approach to modeling of the physical system which is commonly referred to as system identification. In this approach, the mathematical model is first parameterized and then estimated based on input/output experimental data. Autoregressive moving average (ARMA) models are often used in feedback control systems due to their ability to capture the system behavior with lower order and fewer parameters than the MA models or transversal filters. On the other hand, wireless channels are more suitable to be described by MA models due to the quick die out of the CIR. Many identification algorithms exist, and most of them uses squared error as the identification criterion. The squared error includes energy or mean power of the model matching error that results in least squares (LS), or total LS (TLS), or MMSE algorithms. These algorithms will be presented and analyzed in two different sections. For ease of the presentation, only real matrices and variable are considered, but the results are readily extendable to the case of complex valued signals and systems. A very important result in estimation is the well-known Cramér–Rao lower bound (CRLB). See Sect. 6 in Appendix B for details.

G. Gu, *Discrete-Time Linear Systems: Theory and Design with Applications*,
DOI 10.1007/978-1-4614-2281-5_8, © Springer Science+Business Media, LLC 2012

8.1 Least-Squares-Based Identification

Consider an m-input/p-output system with plant model

$$\mathbf{H}(z) = \mathbf{M}(z)^{-1}\mathbf{N}(z) = \left(I - \sum_{k=1}^{n_\mu} M_k z^{-k}\right)^{-1} \left(\sum_{k=1}^{n_\nu} N_k z^{-k}\right).$$

The parameter matrices are those of $\{M_k\}_{k=1}^{n_\mu}$ with dimension $p \times p$, and of $\{N_k\}_{k=1}^{n_\nu}$ with dimension $p \times m$. Due to the existence of measurement error, the input and output are related through the following difference equation:

$$\mathbf{y}(t) = \sum_{k=1}^{n_\mu} M_k \mathbf{y}(t-k) + \sum_{k=1}^{n_\nu} N_k \mathbf{u}(t-k) + \mathbf{v}(t), \tag{8.1}$$

where $\mathbf{v}(t)$ is a WSS process with mean zero and covariance $\sigma^2 I$. Define

$$\Theta = \begin{bmatrix} M_1 \cdots M_{n_\mu} \; N_1 \cdots N_{n_\nu} \end{bmatrix}$$

as the true parameter matrix of dimension $p \times (n_\mu + n_\nu)m$, and

$$\underline{\phi}(t) = \begin{bmatrix} \mathbf{y}(t-1)' \cdots \mathbf{y}(t-n_\mu)' \; \mathbf{u}(t-1)' \cdots \mathbf{u}(t-n_\nu)' \end{bmatrix}'$$

as the regressor vector of dimension $(n_\mu + n_\nu)m$. There holds

$$\mathbf{y}(t) = \Theta \underline{\phi}(t) + \mathbf{v}(t). \tag{8.2}$$

The linearity is owing to the linearity of the system. Although the above signal model is derived from the input/output model (8.1), this section assumes temporarily that $\underline{\phi}(t)$ is noise-free. This assumption holds for the FIR model, including wireless channels. The dependence of the signal model (8.2) on input/output model (8.1) will be revisited in the next section.

8.1.1 LS and RLS Algorithms

The LS algorithm is perhaps the most widely adopted in the practice of system identification. It has an interpretation of maximum likelihood estimate (MLE), if the observation noise is Gauss distributed. Consider Gauss-distributed $\mathbf{v}(t)$ that is temporally white with mean zero and covariance $\Sigma_v = \sigma^2 I$ that is known. The MLE is equivalent to minimizing the squared error index

$$J_\Theta(t_0, t_f) = \frac{1}{2\sigma^2} \sum_{t=t_0}^{t_f} [\mathbf{y}(t) - \Theta \underline{\phi}(t)]' [\mathbf{y}(t) - \Theta \underline{\phi}(t)]$$

by noting that MLE maximizes the PDF given by

$$f_V(\mathbf{y};\Theta) = \frac{1}{\sqrt{(2\pi\sigma^2)^{(t_f-t_0+1)p}}}\exp\left\{-J_\Theta(t_0,t_f)\right\},$$

and by noting the independence of $\{\mathbf{v}(t)\}$. For convenience, denote

$$\begin{aligned} Y_{0,f} &= \left[\mathbf{y}(t_0)\ \mathbf{y}(t_0+1)\cdots\mathbf{y}(t_f)\right], \\ \Phi_{0,f} &= \left[\underline{\phi}(t_0)\ \underline{\phi}(t_0+1)\cdots\underline{\phi}(t_f)\right]. \end{aligned} \tag{8.3}$$

Then the squared error index can be rewritten as

$$J_\Theta(t_0,t_f) = \frac{1}{2\sigma^2}\mathrm{Tr}\left\{\left(Y_{0,f}-\Theta\Phi_{0,f}\right)\left(Y_{0,f}-\Theta\Phi_{0,f}\right)'\right\}. \tag{8.4}$$

The LS solution $\Theta = \Theta_{\mathrm{LS}}$ minimizes $J_\Theta(t_0,t_f)$ and is the MLE.

Recall the definition in (B.50). It is left as an exercise to show (Problem 8.1)

$$\begin{aligned} \frac{\partial \mathrm{Tr}\{AXB\}}{\partial X} &= A'B', \quad \frac{\partial \mathrm{Tr}\{AX'B\}}{\partial X} = BA, \\ \frac{\partial \mathrm{Tr}\{AXBX'\}}{\partial X} &= A'XB' + AXB. \end{aligned}$$

See Sect. B.5 in Appendix B. The next result provides the MLE in the general case.

Theorem 8.1. *Consider the signal model in (8.2) where* $\mathbf{v}(t)$ *is* Gauss *distributed with mean zero and covariance* $\Sigma_v = \mathrm{diag}(\sigma_1^2,\cdots,\sigma_p^2)$. *Then the* MLE *estimate based on* $\{(\mathbf{y}(t),\underline{\phi}(t)\}_{t=t_0}^{t_f}$ *is the* LS *solution and given by*

$$\Theta_{\mathrm{LS}} = Y_{0,f}\Phi_{0,f}'(\Phi_{0,f}\Phi_{0,f}')^{-1}, \tag{8.5}$$

provided that $(\Phi_{0,f}\Phi_{0,f}')$ *is invertible. Moreover, let* $\underline{\theta}_i$ *be the ith row of* Θ. *Then* $\sigma_i^{-2}\Phi_{0,f}\Phi_{0,f}'$ *is the* FIM *associated with estimation of* $\underline{\theta}_i'$ *for* $1 \le i \le p$, *and thus,* $\sigma_i^2(\Phi_{0,f}\Phi_{0,f}')^{-1}$ *is the corresponding* CRLB.

Proof. Suppose that Σ_v is known. Then the Gauss assumption and (8.3) imply that the MLE minimizes

$$\begin{aligned} J_\Theta &= \frac{1}{2}\mathrm{Tr}\left\{\Sigma_v^{-1}\left(Y_{0,f}-\Theta\Phi_{0,f}\right)\left(Y_{0,f}-\Theta\Phi_{0,f}\right)'\right\} \\ &= \frac{1}{2}\mathrm{Tr}\left\{\Sigma_v^{-1}\left(Y_{0,f}Y_{0,f}' - \Theta\Phi_{0,f}Y_{0,f}' - Y_{0,f}\Phi_{0,f}'\Theta' + \Theta\Phi_{0,f}\Phi_{0,f}'\Theta'\right)\right\}. \end{aligned}$$

Direct calculation shows

$$\frac{\partial J_\Theta}{\partial \Theta} = \Sigma_v^{-1}\left(\Theta\Phi_{0,f}\Phi_{0,f}' - Y_{0,f}\Phi_{0,f}'\right).$$

Setting the above to zero yields the MLE in (8.5). Since the MLE is independent of Σ_v, $\Theta_{\rm LS}$ is indeed the MLE. With partition row-wise,

$$Y_{0,f} = \begin{bmatrix} \underline{y}_1(t_0,t_f) \\ \vdots \\ \underline{y}_p(t_0,t_f) \end{bmatrix}, \quad \Theta = \begin{bmatrix} \underline{\theta}_1 \\ \vdots \\ \underline{\theta}_p \end{bmatrix},$$

and $\underline{\varepsilon}_i = \underline{y}_i(t_0,t_f) - \underline{\theta}_i\Phi_{0,f}$ for $1 \le i \le p$. There holds

$$\begin{aligned} \ln f_V(\mathbf{y};\Theta) &= -J_\Theta(t_0,t_f) = -\frac{1}{2}\sum_{i=1}^{p}\sigma_i^{-2}\underline{\varepsilon}_i\underline{\varepsilon}_i' \\ &= -\frac{1}{2}\sum_{i=1}^{p}\sigma_i^{-2}\left(\underline{y}_i(t_0,t_f) - \underline{\theta}_i\Phi_{0,f}\right)\left(\underline{y}_i(t_0,t_f) - \underline{\theta}_i\Phi_{0,f}\right)' \end{aligned}$$

by $\Sigma_v = \mathrm{diag}(\sigma_1^2,\cdots,\sigma_p^2)$. It can be easily verified that

$$\frac{\partial \ln f_V(\mathbf{y};\Theta)}{\partial \underline{\theta}_i'} = -\sigma_i^{-2}\Phi_{0,f}\left(\underline{y}_i(t_0,t_f) - \underline{\theta}_i\Phi_{0,f}\right)' = -\sigma_i^{-2}\Phi_{0,f}\underline{\varepsilon}_i'.$$

By recognizing $\mathrm{E}\{(\underline{y}_i(t_0,t_f) - \underline{\theta}_i\Phi_{0,f})'(\underline{y}_i(t_0,t_f) - \underline{\theta}_i\Phi_{0,f})\} = \sigma_i^2 I$, the above yields the FIM for $\underline{\theta}_i'$:

$$\mathrm{FIM}(\underline{\theta}_i') = \mathrm{E}\left\{\frac{\partial \ln f_V(\mathbf{y};\Theta)}{\partial \underline{\theta}_i'}\frac{\partial \ln f_V(\mathbf{y};\Theta)}{\partial \underline{\theta}_i}\right\} = \sigma_i^{-2}\Phi_{0,f}\Phi_{0,f}'.$$

The corresponding CRLB is thus $\sigma_i^2(\Phi_{0,f}\Phi_{0,f}')^{-1}$ that concludes the proof. □

It needs to keep in mind that the LS solution may not have the MLE interpretation, if $\underline{\phi}(t)$ involves random noises. Consider the case when the covariance of $\mathbf{v}(t)$ is $\Sigma_v(t)$, and thus, $\{\mathbf{v}(t)\}$ is not a WSS process. Suppose that $\Sigma_v(t) = \sigma_t^2 I$ that is known for all t. Then J_Θ is replaced by

$$J_\Theta = \frac{1}{2}\sum_{t=t_0}^{t_f}\mathrm{Tr}\left\{\sigma_t^{-2}[\mathbf{y}(t) - \Theta\underline{\phi}(t)][\mathbf{y}(t) - \Theta\underline{\phi}(t)]'\right\}. \tag{8.6}$$

The MLE minimizes the above J_Θ and satisfies

$$\sum_{t=t_0}^{t_f}\sigma_t^{-2}\left[\Theta_{\rm LS}\underline{\phi}(t)\underline{\phi}(t)' - \mathbf{y}(t)\underline{\phi}(t)'\right] = \mathbf{0}.$$

Hence, the MLE is the weighted LS with weighting $\{\sigma_t^{-2}\}$. Without loss of generality, assume that $t_0 = 0$. For $k > 0$, denote

$$P_k = \left[\sum_{t=0}^{k-1} \sigma_t^{-2} \underline{\phi}(t)\underline{\phi}(t)'\right]^{-1}, \quad Q_k = \sum_{t=0}^{k-1} \sigma_t^{-2} \mathbf{y}(t)\underline{\phi}(t)'. \tag{8.7}$$

By the proof of Theorem 8.1, $\widehat{\Theta}_{t_f} = \Theta_{\rm LS} = Q_{t_f+1}P_{t_f+1}$ is the MLE.

The simplicity form of the LS solution allows its recursive computation with low complexity. To be specific, denote $\widehat{\Theta}_k$ as the LS solution based on the measurement data over the time horizon $[0,\ k)$ for some $k > 0$. Suppose that new input and output measurements are obtained at k. There hold

$$\widehat{\Theta}_k = Q_k P_k, \quad \widehat{\Theta}_{k+1} = Q_{k+1}P_{k+1}.$$

The recursive LS (RLS) algorithm is aimed at computing $\widehat{\Theta}_{k+1}$ based on $\widehat{\Theta}_k$ and the updated regressor $\underline{\phi}(k)$ without explicitly computing $Q_{k+1}P_{k+1}$. In this regard, RLS is similar to Kalman filtering in Theorem 5.6. The key is computation of P_{k+1} based on P_k and $\underline{\phi}(k)$.

First, it is noted that the covariance type matrix P_{k+1} can be written as

$$\begin{aligned} P_{k+1} &= [P_k^{-1} + \sigma_k^{-2}\underline{\phi}(k)\underline{\phi}(k)]^{-1} \\ &= P_k - P_k\underline{\phi}(k)[\sigma_k^2 + \underline{\phi}(k)'P_k\underline{\phi}(k)]^{-1}\underline{\phi}(k)'P_k \end{aligned}$$

by the matrix inversion formula in Appendix A. See also Problem 8.4 in Exercises. The above can be rewritten as

$$P_{k+1} = P_k - P_k\underline{\phi}(k)\mathbf{g}_k, \quad \mathbf{g}_k = [\sigma_k^2 + \underline{\phi}(k)'P_k\underline{\phi}(k)]^{-1}\underline{\phi}(k)'P_k. \tag{8.8}$$

The derivation next shows the relation between $\underline{\phi}(k)'P_{k+1}$ and $\mathbf{g}_k$:

$$\begin{aligned} \underline{\phi}(k)'P_{k+1} &= \underline{\phi}(k)'P_k - \underline{\phi}(k)'P_k\underline{\phi}(k)[\sigma_k^2 + \underline{\phi}(k)'P_k\underline{\phi}(k)]^{-1}\underline{\phi}(k)'P_k \\ &= \{I - \underline{\phi}(k)'P_k\underline{\phi}(k)[\sigma_k^2 + \underline{\phi}(k)'P_k\underline{\phi}(k)]^{-1}\}\underline{\phi}(k)'P_k \\ &= \sigma_k^2[\sigma_k^2 + \underline{\phi}(k)'P_k\underline{\phi}(k)]^{-1}\underline{\phi}(k)'P_k = \sigma_k^2\mathbf{g}_k. \end{aligned}$$

It follows from $Q_{k+1} = Q_k + \sigma_k^{-2}\mathbf{y}(k)\underline{\phi}(k)'$ that

$$\begin{aligned} \widehat{\Theta}_{k+1} &= Q_{k+1}\left[P_k - P_k\underline{\phi}(k)\mathbf{g}_k\right] \\ &= \widehat{\Theta}_k - \widehat{\Theta}_k\underline{\phi}(k)\mathbf{g}_k + \sigma_k^{-2}\mathbf{y}(k)\underline{\phi}(k)'P_{k+1} \\ &= \widehat{\Theta}_k - \widehat{\Theta}_k\underline{\phi}(k)\mathbf{g}_k + \mathbf{y}(k)\mathbf{g}_k = \widehat{\Theta}_k + [\mathbf{y}(k) - \widehat{\mathbf{y}}(k)]\mathbf{g}_k, \end{aligned}$$

where $\widehat{\mathbf{y}}(k) = \widehat{\Theta}_k \underline{\phi}(k)$ can be regarded as the predicted output. The above and (8.8) form the RLS algorithm. If no knowledge is available at $k = 0$, then $\widehat{\Theta}_0 = \mathbf{0}$ and $P_0 = \rho^2 I$ with large ρ can be employed which admit a similar interpretation to that in Kalman filter.

An important problem in parameter estimation is the convergence of the estimate. Consider the case when $\sigma_t^2 = \sigma^2 > 0$ is a constant. Then the signal is called persistent exciting (PE), if $P_t \to \mathbf{0}$ as $t \to \infty$. Accordingly, the PE condition implies $\widehat{\Theta}_t \to \Theta$ asymptotically based on the facts that $\mathbf{g}_t \to \mathbf{0}$ as $t \to \infty$ and that LS algorithm yields the MLE. As a result $\widehat{\Theta}_t$ stops updating eventually. In fact, the LS estimate $\widehat{\Theta}_t$ may stop updating very quickly when the signal is rich in its information content.

While the convergence is welcomed, it does not suit to estimation of the time-varying parameter matrix. Basically, the RLS algorithm fails to track the underlying parameter matrix, when the PE condition holds. One may periodically reset P_t to prevent it from being zero. However, this method is not suggested due to the loss of past information. A more sophisticated method considers the following performance index:

$$J_\Theta(t_f) = \frac{1}{2}\|\mathbf{y}(t_f) - \Theta \underline{\phi}(t_f)\|^2 + \gamma_{t_f} J_\Theta(t_f - 1), \tag{8.9}$$

where $\gamma_t \in (0,\ 1)$ is referred to as the forgetting factor at time t. In the case when $\gamma_t = \gamma$ is a constant and $0 < \gamma < 1$,

$$J_\Theta(t_f) = \frac{1}{2}\sum_{t=0}^{t_f} \gamma^{t_f - t}\|\mathbf{y}(t_f) - \Theta \underline{\phi}(t_f)\|^2.$$

Thus, the terms in the distant past decay exponentially with respect to the time duration. The resultant minimizer is very similar to the RLS algorithm derived earlier with an appropriate modification.

To derive the RLS with the forgetting factor, it is noted that

$$J_\Theta(t-1) = C_t + \Theta P_t^{-1}\Theta' - \Theta Q_t' - Q_t\Theta'$$

for some time-dependent constant matrices P_t, Q_t, and C_t. Consequently,

$$J_\Theta(t) = \frac{\gamma_t}{2}\mathrm{Tr}\left\{C_t + \Theta P_t^{-1}\Theta' - \Theta Q_t' - Q_t\Theta'\right\} + \frac{1}{2}\|\mathbf{y}(t_f) - \Theta \underline{\phi}(t_f)\|^2.$$

Its partial derivative can be easily computed and is given by

$$\left.\frac{\partial J_\Theta(t)}{\partial \Theta}\right|_{\Theta=\widehat{\Theta}_{t+1}} = \widehat{\Theta}_{t+1}[\gamma_t P_t^{-1} + \underline{\phi}(t)\underline{\phi}(t)'] - [\gamma_t Q_t + \mathbf{y}(t)\underline{\phi}(t)'].$$

Setting the above to zero yields $\widehat{\Theta}_{t+1} = \tilde{Q}_{t+1}\tilde{P}_{t+1}$ with

$$\tilde{P}_{t+1} = \gamma_t P_t + \underline{\phi}(t)\underline{\phi}(t)', \quad \tilde{Q}_{t+1} = \gamma_t Q_t + \mathbf{y}(t)\underline{\phi}(t)'.$$

It is interesting to observe that

$$\tilde{P}_{t+1} = \gamma_t P_{t+1}, \quad \tilde{Q}_{t+1} = \gamma_t Q_{t+1}$$

where P_t and Q_t are defined in (8.7) by taking $\gamma_t = \sigma_t^2$. It is left as an exercise (Problem 8.6) to show that the RLS with forgetting factor is given by

$$\widehat{\Theta}_{t+1} = \Theta_t + [\mathbf{y}(t) - \widehat{\mathbf{y}}(t)]\,\mathbf{g}_t, \;\; \widehat{\mathbf{y}}(t) = \widehat{\Theta}_t \underline{\phi}(t), \tag{8.10}$$

$$P_{t+1} = \gamma_t^{-1}\left[P_t - P_t \underline{\phi}(t)\mathbf{g}_t\right], \;\; \mathbf{g}_t = [\gamma_t + \underline{\phi}(t)' P_t \underline{\phi}(t)]^{-1} \underline{\phi}(t)' P_t. \tag{8.11}$$

Two examples are used to illustrate the LS algorithm next.

Example 8.2. Let $\{H_i\}_{i=0}^2$ be the CIR with one input and two outputs. Thus, each H_i has dimension 2×1, specified by

$$H_0 = \begin{bmatrix} 0.6360 \\ 0.0636 \end{bmatrix}, \;\; H_1 = \begin{bmatrix} -0.3552 \\ 0.2439 \end{bmatrix}, \;\; H_2 = \begin{bmatrix} 0.5149 \\ -0.3737 \end{bmatrix}. \tag{8.12}$$

Hence, $\|\mathbf{H}\|_2 = 1$. The training sequence of binary symbols $\pm P_\mathrm{b}$ is transmitted with length ranging from 10 to 20 bits. The received signals are corrupted by i.i.d. Gauss noises with variance σ^2. Since $\|\mathbf{H}\|_2 = 1$, the SNR is the same as the ratio of P_b to σ^2. Numerical simulations are carried out for the cases when SNR $= 10$ dB and when SNR $= 20$ dB. A total of 500 ensemble runs are used to evaluate the channel estimation performance. Let $\{\widehat{H}_i^{(k)}\}$ be the estimated CIR at the kth ensemble run. The RMSE is computed according to

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{k=1}^{T} \mathrm{Tr}\left\{\sum_{i=0}^{2}\left(H_k - \widehat{H}_i^{(k)}\right)\left(H_k - \widehat{H}_i^{(k)}\right)'\right\}}.$$

The results are plotted in the following figure.

The upper two curves correspond to the case of SNR $= 10$ dB. The dashed line marked with diamond shows the RMSE, while the solid line marked with circle is the corresponding CRLB curve, defined by $\sqrt{\mathrm{Tr}\{\mathrm{FIM}^{-1}\}}$ that is the lower bound for RMSE. These two curves are close to each other. In fact, the two curves overlap for large T (Problem 8.5 in Exercises). The lower two curves in Fig. 8.1 correspond to the case of SNR $= 20$ dB. The dashed line marked with square shows the RMSE, while the solid line marked with $*$ shows the corresponding CRLB.

The simulation result validates the MLE nature of the LS solution in the case of FIR models that hold for the wireless channels (Problem 8.8 in Exercises). However, if the temporary assumption on noise-free $\{\underline{\phi}(t)\}$ is removed that is the case for IIR models (refer to (8.1) in which $M_k \neq \mathbf{0}$ for at least one k), the MLE interpretation of the LS solution will be lost. Specifically, the physical systems in feedback control are generically described by IIR models. Figure 8.2 illustrates the plant model with input/output signals together with observation noises.

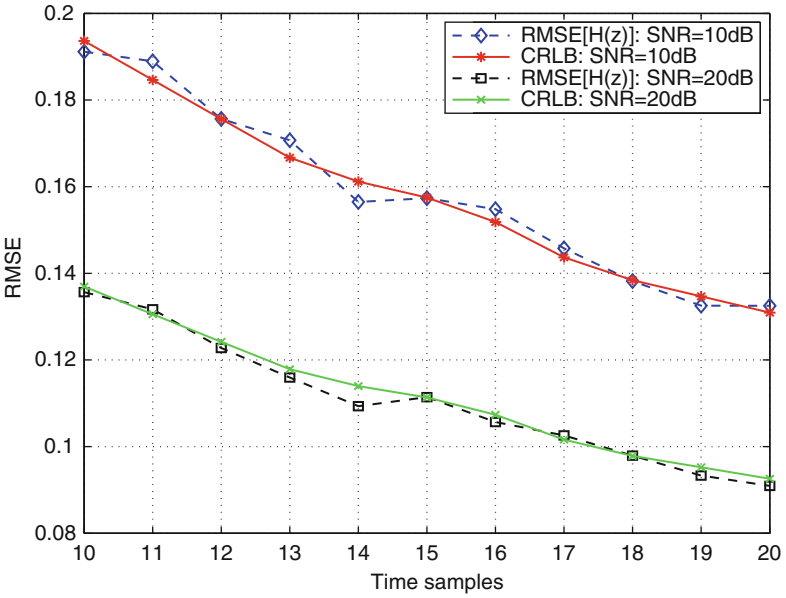
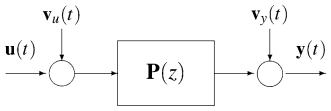


Fig. 8.1 RMSE plots for LS-based channel estimation

Fig. 8.2 Plant model with noisy measurement data



It is clear that both observations $\mathbf{u}(t)$ and $\mathbf{y}(t)$ are no longer the actual input and output of the plant. In fact, the signal model in (8.1) is replaced by

$$\mathbf{y}(t) = \mathbf{v}_y(t) + \sum_{k=1}^{n_\mu} M_k \mathbf{y}(t-k) + \sum_{k=1}^{n_\nu} N_k[\mathbf{u}(t-k) + \mathbf{v}_u(t-k)]. \tag{8.13}$$

Hence, when the LS algorithm is applied to estimate the system parameters, the estimation performance is different from the case for FIR models. The next example illustrates the identification results.

Example 8.3. Consider identification of a SISO plant represented by its transfer function

$$P(z) = \frac{0.3624z + 0.1812}{z^2 - 1.5z + 0.7}.$$

The input $u(t)$ is chosen as white Gauss with variance 1 that results in output variance about 4 dB. Observation noises are added to input, or output, or both in the same way as shown in Fig. 8.2. The corrupting noises are white and Gauss distributed with variance 0.1, or -20 dB. The corresponding RMSE curves are

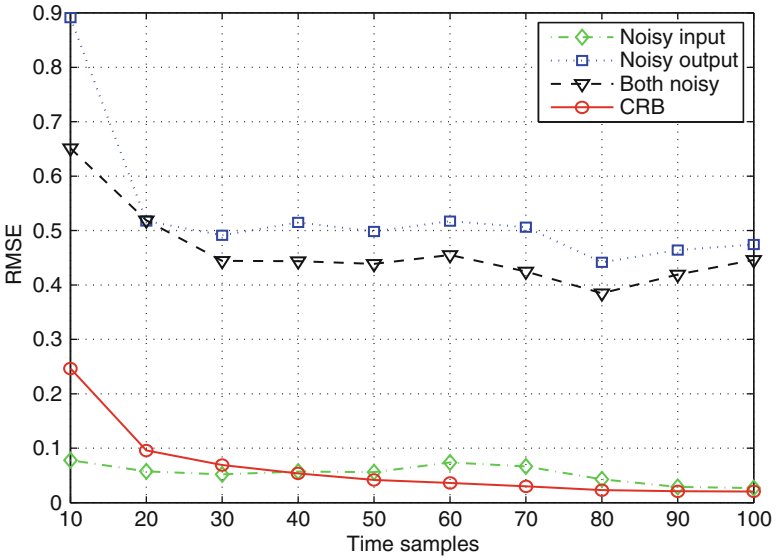


Fig. 8.3 RMSE curves for LS-based IIR model estimation

plotted in Fig. 8.3 with a larger observation interval than the previous example, in order to illustrate the trend of the identification error. It is clearly seen that the RMSE curves do not monotonically decline anymore. The simulation results indicate that the LS solution is biased for identification of IIR models in the presence of output observation noises. In fact, the corruption noise at the output impacts more negatively than at the input in terms of the estimation performance.

It needs to be pointed out that the CRLB curve used in the figure is the same as $\sigma_v\sqrt{\mathrm{Tr}\{(\Phi_{0,f}\Phi_{0,f}')^{-1}\}}$. This expression is actually not the true CRLB anymore due to the observation noise involved in $\Phi_{0,f}$. The derivation of the CRLB for the case of noisy $\Phi_{0,f}$ will be investigated in a later subsection.

8.1.2 MMSE Algorithm

For the signal model $\mathbf{y}(t)=\Theta\underline{\phi}(t)+\mathbf{v}(t)$ studied in the previous section, $\{\underline{\phi}(t)\}$ is likely to involve observation noises, in which case the LS solution is not the MLE, and the MLE solution is difficult to compute in general. An alternative to MLE is the MMSE estimate that minimizes $\mathrm{E}\{\|\mathbf{y}(t)-\Theta\underline{\phi}(t)\|^2\}$. Assume that both $\underline{\phi}(t)$ and $\mathbf{v}(t)$ are WSS processes, and denote

$$R_{\mathbf{y}}=\mathrm{E}\{\mathbf{y}(t)\mathbf{y}(t)'\},\quad R_{\phi}=\mathrm{E}\{\underline{\phi}(t)\underline{\phi}(t)'\},\quad R_{\mathbf{y},\phi}=\mathrm{E}\{\mathbf{y}(t)\underline{\phi}(t)'\}.$$

The following provides the MMSE estimate.

Theorem 8.4. *Suppose that both* $\underline{\phi}(t)$ *and* $\mathbf{v}(t)$ *are* WSS *processes, and* R_ϕ *is nonsingular. Then* $\Theta_{\text{MMSE}} = R_{\mathbf{y}\phi}R_\phi^{-1}$ *is the* MMSE *estimate and minimizes* $\mathrm{E}\{\|\mathbf{y}(t) - \Theta\underline{\phi}(t)\|^2\}$. *Moreover, the* MSE *associated with* Θ_{MMSE} *is given by*

$$\varepsilon_{\text{MMSE}} = \min_{\Theta} \mathrm{E}\{\|\mathbf{y}(t) - \Theta\underline{\phi}(t)\|^2\} = \mathrm{Tr}\left\{R_{\mathbf{y}} - R_{\mathbf{y}\phi}R_\phi^{-1}R_{\mathbf{y}\phi}'\right\}. \tag{8.14}$$

Proof. Let $\varepsilon_{\text{MSE}} = \mathrm{E}\{\|\mathbf{y}(t) - \Theta\underline{\phi}(t)\|^2\}$ be the performance index for the MMSE estimation. Then

$$\varepsilon_{\text{MSE}} = \mathrm{Tr}\left\{R_{\mathbf{y}} + \Theta R_\phi \Theta' - R_{\mathbf{y}\phi}\Theta' - \Theta R_{\mathbf{y}\phi}'\right\}.$$

Since the MMSE estimate minimizes ε_{MSE}, it can be computed from

$$\frac{\partial \varepsilon_{\text{MSE}}}{\partial \Theta} = 2\left(\Theta R_\phi - R_{\mathbf{y}\phi}\right) = \mathbf{0} \tag{8.15}$$

that shows $\Theta_{\text{MMSE}} = R_{\mathbf{y}\phi}R_\phi^{-1}$. Substituting $\Theta = \Theta_{\text{MMSE}}$ into the expression of ε_{MSE} yields (8.14). □

The autocorrelation matrices R_ϕ and $R_{\mathbf{y}\phi}$ are often unavailable in practice. Estimates based on N samples of measurements over $[0,\ t_f]$ with $t_f = N-1$ can be used:

$$R_\phi \approx \frac{1}{N}\sum_{t=0}^{N-1} \underline{\phi}(t)\underline{\phi}(t)' = \frac{P_N^{-1}}{N}, \quad R_{\mathbf{y}\phi} \approx \frac{1}{N}\sum_{t=0}^{N-1} \mathbf{y}(t)\underline{\phi}(t)' = \frac{Q_N}{N} \tag{8.16}$$

where P_N and Q_N are the same as defined in (8.7). If

$$R_\phi = \lim_{N\to\infty} \frac{1}{N}\sum_{t=0}^{N-1} \underline{\phi}(t)\underline{\phi}(t)', \quad R_{\mathbf{y}\phi} = \lim_{N\to\infty} \frac{1}{N}\sum_{t=0}^{N-1} \mathbf{y}(t)\underline{\phi}(t)',$$

then $\{\underline{\phi}(t)\}$ and $\{\mathbf{y}(t)\}$ are called ergodic processes. The PE condition is clearly necessary for R_ϕ to be nonsingular. Under both WSS and ergodic assumptions, the RLS solution approaches the MMSE estimate, i.e.,

$$\widehat{\Theta}_N = P_N Q_N = \left(\frac{1}{N}\sum_{t=0}^{N-1} \underline{\phi}(t)\underline{\phi}(t)'\right)^{-1}\left(\frac{1}{N}\sum_{t=0}^{N-1} \mathbf{y}(t)\underline{\phi}(t)'\right) \ \to\ \Theta_{\text{MMSE}}$$

as $N \to \infty$. It follows that (cf. Problem 8.10 in Exercises)

$$\widehat{\Theta}_N \ \to\ \Theta + \lim_{N\to\infty} \mathrm{E}\left\{V_{0,f}\Phi_{0,f}'\right\}\left(\mathrm{E}\left\{\Phi_{0,f}\Phi_{0,f}'\right\}\right)^{-1} \tag{8.17}$$

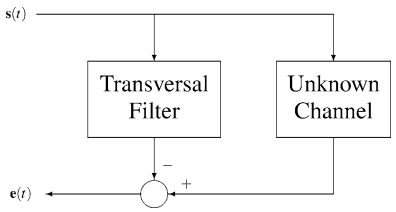


Fig. 8.4 Schematic illustration for channel estimation

where $N = t_f - t_0 + 1$. Hence, if $\mathrm{E}\{V_{0,f}\Phi'_{0,f}\} \to \mathbf{0}$ as $N \to \infty$, then the LS algorithm is asymptotically unbiased. Otherwise, the LS solution is biased in which case large number of samples does not help to eliminate the bias in the LS solution.

The next example illustrates the use of MMSE estimation.

Example 8.5. In wireless communications, channel information is essential for reliable data detection. While pilot tones such as training sequence can be used to estimate the CIR, it is desirable to estimate the CIR based on statistical information of the data sequence. Consider MIMO channel estimation in Fig. 8.4 with $\mathrm{E}\{\mathbf{s}(t)\mathbf{s}(t-k)'\} = R_{\mathbf{s}}(k)$ assumed to be known for each integer k. The received signal at the output of the channel is given by

$$\mathbf{y}(t) = \sum_{k=1}^{L} H_i \mathbf{s}(t-i) + \mathbf{v}(t) = \Theta \underline{\phi}(t) + \mathbf{v}(t)$$

for some white Gauss noise $\mathbf{v}(t)$ where

$$\Theta = \begin{bmatrix} H_1 & \cdots & H_L \end{bmatrix}, \quad \phi(t) = \begin{bmatrix} \mathbf{s}(t-1) \\ \vdots \\ \mathbf{s}(t-L) \end{bmatrix}.$$

It follows from $\mathrm{E}\{\mathbf{s}(t)\mathbf{s}(t-k)'\} = R_{\mathbf{s}}(k)$ that

$$R_\phi = \underline{R}_{\mathbf{s}} = \mathrm{E}\{\underline{\phi}(t)\underline{\phi}(t)'\} = [\, R_{\mathbf{s}}(k-i)\,]_{i,k=1,1}^{L,L}$$

is a block Toeplitz matrix known at the receiver. However, $R_{\mathbf{y}\phi}$ has to be estimated using

$$R_{\mathbf{y}\phi} \approx \frac{1}{N} \sum_{t=0}^{N-1} \mathbf{y}(t)\underline{\phi}(t)'$$

for some integer $N >> 1$. In this case, the MMSE estimate for CIR can be obtained from $\Theta_{\mathrm{MMSE}} = R_{\mathbf{y}\phi} R_\phi^{-1}$. If $\{R_{\mathbf{s}}(k)\}_{k=0}^{L-1}$ are not available at the receiver, they need to be estimated as well. A commonly seen method employs the past detected symbols, assumed to be correct, which yields effective way to estimate R_ϕ.

It is worth to pointing out the difference between the MMSE estimation in this section and that in Chap. 5. Recall that in Chap. 5, the state vector under estimation is random, whereas the parameters under estimation in this section are deterministic.

8.2 Subspace-Based Identification

For the signal model $\mathbf{y}(t) = \Theta\underline{\phi}(t) + \mathbf{v}(t)$ studied in the previous section, the MLE interpretation for the LS algorithm does not hold in general, if the noise sequence $\{\mathbf{v}(t)\}$ is not Gauss distributed. Although the LS solution can still be used to extrapolate the system model, the estimate is not unbiased anymore for IIR models, because $\underline{\phi}(t)$ involves $\{\mathbf{y}(k)\}$ for $k < t$ (Problem 8.10 in Exercises). It turns out that it is the TLS algorithm that yields the unbiased estimate asymptotically which will be used for system identification in this section.

8.2.1 TLS Estimation Algorithm

The TLS algorithm arises from the class of error-in-variable (EIV) models. Let Θ be the parameter matrix of interest satisfying $\Theta A_0 = B_0$ where A and B are wide matrices. The precise values of A_0 and B_0 are not available. Instead, only their measurements, denoted by A and B, respectively, are available given by the EIV model

$$\begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} A_0 \\ B_0 \end{bmatrix} + M\Delta. \tag{8.18}$$

To be specific, the dimensions of A, B, and M are $L \times N$, $p \times N$, and $p \times \ell$ with $N > (L + p)$. Thus, Δ has dimension $\ell \times N$. The elements of Δ are assumed to be independent and identically distributed (i.i.d.) random variables with mean zero and variance σ^2. The goal is to estimate Θ of dimension $p \times L$ based on A, B, and M.

The previous section studied the estimation problem for the case of $A = A_0$. The only measurement error comes from B. Recall the deterministic assumption for $\Phi_{0,f}$ by taking $A = A_0 = \Phi_{0,f}$ and $B = Y_{0,f} \neq B_0$. The LS algorithm finds $\widehat{B}$ closest to B such that

$$\text{rank}\left\{\begin{bmatrix} A \\ \widehat{B} \end{bmatrix}\right\} = \text{rank}\{A\},$$

and then solve for $\widehat{\Theta}$ from $\Theta A = \widehat{B}$. It is noted that $B = \widehat{B} + \widehat{B}_\perp$ of which each row of $\widehat{B}$ lies in the row space of A, and $\widehat{B}_\perp A' = \mathbf{0}$. That is, the row spaces of $\widehat{B}$ and $\widehat{B}_\perp$ are orthogonal to each other. Consequently, $\Theta AA' = BA' = \widehat{B}A'$ yielding the LS solution $\Theta_{\text{LS}} = BA'(AA')^{-1}$. If the measurement errors are Gauss distributed, then the LS algorithm yields the MLE estimate.

When $A \neq A_0$ in addition to $B \neq B_0$, both $\widehat{A}$ and $\widehat{B}$ closest to A and B, respectively, are searched for such that

$$\text{rank}\left\{\begin{bmatrix}\widehat{A}\\ \widehat{B}\end{bmatrix}\right\} = \text{rank}\{\widehat{A}\}$$

prior to solving for Θ from $\Theta\widehat{A} = \widehat{B}$ that is the essence of the TLS. Let $M = I$ to begin with. The TLS algorithm is aimed at minimizing

$$J_{A,B} := \left\|\begin{bmatrix}A\\ B\end{bmatrix} - \begin{bmatrix}\widehat{A}\\ \widehat{B}\end{bmatrix}\right\|_{\mathscr{F}} \quad \text{subject to} \quad \text{rank}\left\{\begin{bmatrix}\widehat{A}\\ \widehat{B}\end{bmatrix}\right\} = L \tag{8.19}$$

with $\|X\|_{\mathscr{F}} = \sqrt{\text{Tr}\{X'X\}}$ the Frobenius norm. A similar problem is encountered in Hankel-norm approximation. See (4.58) in Sect. 4.3 and the discussion therein. Hence, SVD can be used to compute such a pair of $\widehat{A}$ and $\widehat{B}$. As a result, there exists a unique solution pair (X_1, X_2) with X_2 of dimension $p \times p$ to $X_1\widehat{A} = X_2\widehat{B}$. If X_2 is nonsingular, then $\widehat{\Theta} = X_2^{-1}X_1$ is the TLS solution.

A formal procedure for TLS is stated next. Define

$$W := \begin{bmatrix}A\\ B\end{bmatrix}\begin{bmatrix}A' & B'\end{bmatrix}, \tag{8.20}$$

and let $W = G'\Lambda G$ be the eigenvalue decomposition with

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_{L+p})$$

arranged in descending order. Partition the eigenvector matrix G and eigenvalue matrix Λ according to

$$G = \begin{bmatrix}G_{11} & G_{12}\\ G_{21} & G_{22}\end{bmatrix}, \quad \Lambda = \begin{bmatrix}\Lambda_1 & \mathbf{0}\\ \mathbf{0} & \Lambda_2\end{bmatrix}, \tag{8.21}$$

where G_{11} and Λ_1 have the same dimension of $L \times L$. Then

$$\Theta_{\text{TLS}} = G_{21}G_{11}^{-1} = -(G_{22}')^{-1}G_{12}' \tag{8.22}$$

is the TLS estimate, provided that G_{22} is nonsingular. It can be shown that $\det(G_{22}) \neq 0$ has probability 1, but the proof is skipped because of the involvement of more specialized mathematical background. The next result shows the MLE property when the elements of Δ are Gauss distributed.

Theorem 8.6. *Suppose that $M = I$ and elements of Δ are normal* i.i.d. *with mean zero and variance σ^2. Let W be defined in (8.20), $W = G'\Lambda G$ be the eigenvalue decomposition with diagonal elements of Λ arranged in descending order. Partition*

G and Λ as in (8.21) where G_{11} and Λ_1 are of dimension $L \times L$. Then G_{11} and G_{22} are nonsingular w.p. 1 (*with probability* 1), *and*

$$\widehat{\Theta} = G_{21}G_{11}^{-1} = -(G_{22}')^{-1}G_{12}', \quad \widehat{\sigma}^2 = \frac{\mathrm{Tr}\{\Lambda_2\}}{(L+p)N} \tag{8.23}$$

are MLEs *for Θ and σ^2, respectively.*

Proof. The PDF for the measurement data A and B is given by

$$f_\Delta(\{\delta_{ij}\}) = \frac{1}{\left(\sqrt{2\pi\widehat{\sigma}^2}\right)^{(L+p)N}} \exp\left\{ -\frac{1}{2\widehat{\sigma}^2} \left\| \begin{bmatrix} A \\ B \end{bmatrix} - \begin{bmatrix} \widehat{A} \\ \widehat{B} \end{bmatrix} \right\|_{\mathscr{F}}^2 \right\}.$$

The MLE searches $\widehat{A}, \widehat{B}$ and $\widehat{\sigma}^2$ which maximize $f_\Delta(\{\delta_{ij}\})$. Since $J_{A,B}$ defined in (8.19) is independent of σ^2, its minimum is $\mathrm{Tr}\{\Lambda_2\}$ by taking $\widehat{\Theta} = G_{21}G_{11}^{-1}$. See Problem 8.13 in Exercises. Hence,

$$\max f_\Delta(\{\delta_{ij}\}) = \max_{\widehat{\sigma}^2} \frac{1}{\left(\sqrt{2\pi\widehat{\sigma}^2}\right)^{(L+p)N}} \exp\left\{ -\frac{1}{2\widehat{\sigma}^2} \mathrm{Tr}\{\Lambda_2\} \right\}.$$

Taking derivative with respect to $\widehat{\sigma}$ and setting it to zero yield

$$\frac{(L+p)N}{\widehat{\sigma}} - \frac{\mathrm{Tr}\{\Lambda_2\}}{\widehat{\sigma}^3} = 0.$$

Hence, $\widehat{\sigma}^2$ in (8.23) is the MLE for σ^2 that concludes the proof. □

Theorem 8.6 shows the MLE of the TLS, but whether or not it is unbiased, estimate remains unknown. The following shows that the TLS solution is asymptotically unbiased.

Theorem 8.7. *Suppose that the same hypotheses of* Theorem 8.6 *hold, and assume that*

$$\text{(i)}\ \ \Pi_0 := \lim_{N\to\infty} \frac{A_0A_0'}{N} > \mathbf{0}, \quad \text{(ii)}\ \lim_{N\to\infty} \frac{\Delta \begin{bmatrix} A_0' & B_0' \end{bmatrix}}{N} = \mathbf{0}. \tag{8.24}$$

Then the TLS *estimate $\Theta_{\mathrm{TLS}} \to \Theta$ as $N \to \infty$. In addition, there holds*

$$\lim_{N\to\infty} \frac{W}{N} = \sigma^2 I + \begin{bmatrix} I \\ \Theta \end{bmatrix} \Pi_0 \begin{bmatrix} I & \Theta' \end{bmatrix}. \tag{8.25}$$

Proof. By the EIV model (8.18), $M = I$, W in (8.20), and $\Theta A_0 = B_0$,

$$\frac{W}{N} = \frac{1}{N} \begin{bmatrix} A_0 \\ B_0 \end{bmatrix} \begin{bmatrix} A_0' & B_0' \end{bmatrix} + \frac{\Delta\Delta'}{N} + \frac{\Delta}{N} \begin{bmatrix} A_0' & B_0' \end{bmatrix} + \frac{1}{N} \begin{bmatrix} A_0 & B_0 \end{bmatrix} \Delta'.$$

Taking limit $N \to \infty$ and using the two assumptions in (8.24) arrive at

$$\lim_{N\to\infty} \frac{W}{N} = \lim_{N\to\infty} \frac{1}{N} \begin{bmatrix} A_0 \\ B_0 \end{bmatrix} \begin{bmatrix} A_0' & B_0' \end{bmatrix} + \frac{\Delta\Delta'}{N}$$
$$= \sigma^2 I + \begin{bmatrix} I \\ \Theta \end{bmatrix} \Pi_0 \begin{bmatrix} I & \Theta' \end{bmatrix}$$

by the assumption on the i.i.d. of elements of Δ, which verifies (8.25). Hence, $\lambda_{L+i} \to \sigma^2$ as $N \to \infty$ for $1 \le i \le p$. Denote $\mathscr{R}(\cdot)$ for the range space. Then

$$\mathscr{R}\left(\begin{bmatrix} G_{11} \\ G_{12} \end{bmatrix} \right) \longrightarrow \mathscr{R}\left(\begin{bmatrix} I \\ \Theta \end{bmatrix} \right)$$

as $N \to \infty$. The asymptotic convergence of $\widehat{\Theta}_{\text{TLS}}$ to Θ thus follows. □

Two comments are made. The first regards the assumption in (8.24): Statement (ii) is in fact implied by (i). But because the proof is more involved, it is skipped. The second is the convergence of the MLE for the noise variance σ^2. The proof of Theorem 8.7 indicates that

$$\lim_{N\to\infty} \frac{(L+p)\widehat{\sigma}^2}{p} = \lim_{N\to\infty} \frac{\text{Tr}\{\Lambda_2\}}{pN} = \frac{\text{Tr}\{\sigma^2 I_p\}}{p} = \sigma^2 \tag{8.26}$$

where $\widehat{\sigma}^2$ is the MLE for σ^2 in Theorem 8.6. Therefore, the MLE for the σ^2 is not an asymptotically unbiased estimate. The regularity condition breaks down for estimation of σ^2.

Theorems 8.6 and 8.7 address the estimation problem for the EIV model in the case of $M = I$. If $M \neq I$ is a full rank and possibly wide matrix, it can be converted to the estimation problem of $M = I$.

Corollary 8.1. *Under the same conditions and hypotheses of* Theorem 8.6 *except that* $M(\neq I)$ *has the full row rank, the expressions of* MLEs *in (8.23) hold, provided that the eigenvalue decomposition of* W *is replaced by that of* $W_0 = \Sigma_0^{-1/2} W \Sigma_0^{-1/2}$ *where* $MM' = \Sigma_0$. *In addition, the* MLE $\widehat{\Theta}$ *is an asymptotically unbiased estimate for* Θ, *and* $\frac{(L+p)\widehat{\sigma}^2}{p}$ *is an asymptotically unbiased estimate for* σ^2.

Proof. It is important to note that $\sigma^2 \Sigma_0 = \sigma^2 MM'$ can be regarded as the common covariance for each column of ΔM, and thus,

$$\Sigma_0^{-1/2} \begin{bmatrix} A \\ B \end{bmatrix} = \Sigma_0^{-1/2} \begin{bmatrix} A_0 \\ B_0 \end{bmatrix} + U\Delta \tag{8.27}$$

with $\Sigma_0^{1/2}$ symmetric and $U = \Sigma_0^{-1/2} M$ satisfying $UU' = I$. Theorem 8.6 can now be applied, leading to

$$\left(\lambda_i I - \Sigma_0^{-1/2} W \Sigma_0^{-1/2}\right) \mathbf{v}_i = 0 \tag{8.28}$$

for $i = 1, 2, \cdots, L+p$. Hence, by setting

$$G = \left[\, \mathbf{v}_1 \; \cdots \; \mathbf{v}_{L+p} \,\right], \quad \Lambda = \text{diag}(\lambda_1, \cdots, \lambda_{L+p}),$$

the proof of the corollary can be concluded. □

It is noted that the eigenvalue/eigenvector equation in (8.28) is the same as the following generalized eigenvalue/eigenvector equation:

$$(\lambda_i \Sigma_0 - W)\, \mathbf{g}_i = \mathbf{0} \tag{8.29}$$

by taking $\mathbf{g}_i = \Sigma_0^{-1/2} \mathbf{v}_i$ for $i = 1, 2, \cdots, L+p$. The above is more convenient to compute than (8.28), and avoids the potential numerical problem associated with the inversion of $\Sigma_0^{1/2}$.

In the case when $M (\neq I)$ is a wide and full rank matrix, the MLE for Θ corresponds to the generalized TLS solution. It requires to compute the p smallest (generalized) eigenvalues and their respective eigenvectors in (8.29) in order to obtain the MLE. Let $\text{diag}(X)$ be a diagonal matrix using diagonal elements of X. It is interesting to observe that the p eigenvectors associated with the p smallest (generalized) eigenvalues in (8.29) solve the following minimization problem (Problem 8.12 in Exercises):

$$\min_H \left\{ \text{Tr}\{H' W H\} : \;\; \text{diag}\left(H' \Sigma_0 H\right) = I_p \right\}. \tag{8.30}$$

Indeed, by denoting $\mathbf{h}_i$ as the ith column of H, and by setting the cost index

$$J = \sum_{i=1}^{p} \left[\mathbf{h}_i' W \mathbf{h}_i + \gamma_i (1 - \mathbf{h}_i' \Sigma_0 \mathbf{h}_i) \right] \tag{8.31}$$

with $\{\gamma_i\}_{i=1}^p$ Lagrange multipliers, the constrained minimization in (8.30) is equivalent to the unconstrained minimization of J in (8.31). Carrying out computation of the necessary condition leads to $(\gamma_i \Sigma_0 - W)\mathbf{h}_i = 0$ that has the same form as (8.29). Hence, the optimality is achieved by taking $\gamma_i = \lambda_{L+i}$ and $\mathbf{h}_i = \mathbf{g}_{L+i}$ for $1 \leq i \leq p$ that are the p smallest (generalized) eigenvalues and their respective eigenvectors in (8.29). In the next two subsections, the results of the TLS solution will be applied to channel estimation in wireless communications, and also to system identification in feedback control systems.

8.2.2 Subspace Method

In wireless communications, the CIR has finite duration. Let $\{H_i\}$ be the CIR of a MIMO channel with M input and P output. The received signal is mathematically described by

$$\mathbf{y}(t) = \sum_{i=0}^{L} H_i \mathbf{s}(t-i) + \mathbf{v}(t), \tag{8.32}$$

where $\{\mathbf{s}(k)\}$ is the sequence of the transmitted symbols and $\{\mathbf{v}(k)\}$ is the sequence of i.i.d. with normal distribution. Denote

$$\underline{\mathbf{y}}(t) = \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{y}(t-1) \\ \vdots \\ \mathbf{y}(t-q) \end{bmatrix}, \quad \underline{\mathbf{v}}(t) = \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{v}(t-1) \\ \vdots \\ \mathbf{v}(t-q) \end{bmatrix}, \quad \underline{\mathbf{s}}(t) = \begin{bmatrix} \mathbf{s}(t) \\ \mathbf{s}(t-1) \\ \vdots \\ \mathbf{s}(t-L_q) \end{bmatrix}$$

with $L_N = L+q$. Let $T_{\mathscr{H}}$ be a block Toeplitz matrix defined by

$$T_{\mathscr{H}} = \begin{bmatrix} H_0 & \cdots & H_L & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & H_0 & \cdots & H_L \end{bmatrix} \tag{8.33}$$

that has dimension $(q+1)P \times (q+L+1)M$. There holds

$$\underline{\mathbf{y}}(t) = T_{\mathscr{H}} \underline{\mathbf{s}}(t) + \underline{\mathbf{v}}(t). \tag{8.34}$$

Training signals are often employed to estimate the CIR. However, the use of training signals consume precious channel bandwidth. There is thus a strong incentive to estimate the channel blindly, given the statistics of the symbol sequence $\{\mathbf{s}(t)\}$. A common assumption is that $\mathbf{s}(t)$ is a WSS process and has mean zero and covariance Σ_s that is known at the receiver site. This subsection considers the subspace method for blind channel estimation.

Suppose that $P > M$. Then $T_{\mathscr{H}}$ is a strictly tall matrix, if $(q+1)(P-M) > LM$. Since L and M are fixed for each MIMO channel, the block Toeplitz matrix $T_{\mathscr{H}}$ can be made strictly tall by taking large q.

Lemma 8.1. *Let $T_{\mathscr{H}}$ of dimension $(q+1)P \times (q+L+1)M$ be the block Toeplitz matrix defined in (8.33) with $P > M$ and $(q+1)(P-M) > LM$. Suppose that both H_0 and H_L have the full column rank. Then $T_{\mathscr{H}}$ has the full column rank, if*

$$\operatorname{rank}\{\, \mathbf{H}(z) \,\} = M \ \ \forall\, z \in \mathbb{C} \tag{8.35}$$

where $\mathbf{H}(z) = H_0 + H_1 z^{-1} + \cdots + H_L z^{-L}$ is the channel transfer matrix.

Proof. The contrapositive argument will be used for the proof. Suppose that $T_{\mathscr{H}}$ has the full column rank, but the rank condition (8.35) fails. Since $\mathbf{H}(z)$ losses its rank for some $z = z_0$, there exists $\mathbf{s} \neq \mathbf{0}$ such that $\mathbf{H}(z_0)\mathbf{s} = 0$. The hypothesis that both H_0 and H_L have the full column rank implies that $z_0 \neq 0$ and $z_0 \neq \infty$. As a result,

$$\sum_{i=0}^{L} H_i z_0^{-(i+k)} \mathbf{s} = \mathbf{0}$$

for each positive integer k. By taking $\mathbf{s}(t-k) = z_0^{-k}\mathbf{s}$ for each element of $\underline{\mathbf{s}}(t)$ in (8.34) yields $T_{\mathscr{H}}\underline{\mathbf{s}}(t) = \mathbf{0}$, contradicting to the full column rank assumption for $T_{\mathscr{H}}$ at the beginning. The proof is now complete. □

A common assumption on the measurement noise is that it is not only temporally but also spatially white. Hence, $\mathrm{E}\{\mathbf{v}(t)\mathbf{v}(t)'\} = \sigma_v^2 I$. Under the condition that $\{\mathbf{s}(t)\}$ and $\{\mathbf{v}(t)\}$ are independent random processes,

$$\Sigma_y = \mathrm{E}\left\{\mathbf{y}(t)\mathbf{y}(t)'\right\} = T_{\mathscr{H}}\Sigma_s T'_{\mathscr{H}} + \sigma^2 I.$$

Both $\{\mathbf{s}(t)\}$ and $\{\mathbf{v}(t)\}$ are assumed to be not only WSS, but also ergodic. Consequently, there holds

$$\hat{\Sigma}_y = \frac{1}{N}\sum_{t=0}^{N-1} \mathbf{y}(t)\mathbf{y}(t)' \;\longrightarrow\; \Sigma_y = T_{\mathscr{H}}\Sigma_s T'_{\mathscr{H}} + \sigma^2 I, \tag{8.36}$$

as $N \to \infty$. Applying eigenvalue decomposition to Σ_y yields

$$\Sigma_y = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} \begin{bmatrix} \Lambda_1 & \mathbf{0} \\ \mathbf{0} & \sigma_v^2 I_\nu \end{bmatrix} \begin{bmatrix} G'_{11} & G'_{21} \\ G'_{12} & G'_{22} \end{bmatrix}$$

where the partitions are compatible and $\nu = (q+1)(P-M) - LM > 0$. Recall $\mathscr{R}\{\cdot\}$ for the range space and $\mathscr{N}\{\cdot\}$ for the null space. There hold

$$\mathscr{R}\{T_{\mathscr{H}}\} = \mathscr{R}\left\{\begin{bmatrix} G_{11} \\ G_{21} \end{bmatrix}\right\}, \quad \mathscr{N}\{T'_{\mathscr{H}}\} = \mathscr{R}\left\{\begin{bmatrix} G_{12} \\ G_{22} \end{bmatrix}\right\}. \tag{8.37}$$

The former is termed signal subspace and the latter termed noise subspace. The orthogonality of the two subspaces leads to the subspace method for channel estimation. Specifically, $T_{\mathscr{H}} = T_{\mathscr{H}}(\{H_i\})$, and there holds

$$\begin{bmatrix} G'_{12} & G'_{22} \end{bmatrix} T_{\mathscr{H}}(\{H_i\}) = \mathbf{0}. \tag{8.38}$$

However, the precise Σ_y is not available due to finitely many samples of the received signal and the existence of the measurement error. Hence, the relation in (8.38) does not hold, if the eigen-matrix G is computed based on the estimated Σ_y. Nevertheless,

(8.38) suggests an effective way for channel estimation by searching for $\{\widehat{H}_i\}$ to minimize

$$J_{\mathscr{F}} = \left\| \left[G'_{12} \; G'_{22} \right] T_{\mathscr{H}}(\{\widehat{H}_i\}) \right\|_{\mathscr{F}} \text{ subject to } \sum_{i=0}^{L} \widehat{H}'_i \widehat{H}_i = I. \tag{8.39}$$

The normalization constraint is necessary to prevent the CIR estimates $\{\widehat{H}_i\}$ from being the meaningless zero. Although other norms and normalization constraints can be adopted, the constrained minimization of $J_{\mathscr{F}}$ in (8.39) leads to a simpler solution for the CIR estimates $\{\widehat{H}_i\}$.

Specifically, let $C = \left[G'_{12} \; G'_{22} \right]$ that has dimension $\nu \times (q+1)P$. It can be partitioned into $(q+1)$ blocks of the same size as follows:

$$\left[G'_{12} \; G'_{22} \right] = \left[C_0 \; C_1 \; \cdots \; C_q \right].$$

Thus, each C_i has dimension $\nu \times P$. Let $F = [G'_{12} \; G'_{22}] T_{\mathscr{H}}(\{\widehat{H}_i\})$. Recall that $T_{\mathscr{H}}(\{\widehat{H}_i\})$ has dimension $(q+1)P \times (q+L+1)M$. There exists a partition

$$\left[G'_{12} \; G'_{22} \right] T_{\mathscr{H}}(\{\widehat{H}_i\}) = \left[F_0 \; F_1 \; \cdots \; F_{q+L+1} \right] \tag{8.40}$$

with $\{F_i\}$ of the same dimension of $\nu \times M$. Denote Θ as the parameter matrix of dimension $(L+1)P \times M$ with H_i being the $(i+1)$th block, $\underline{F}$ as a $(q+L+1)\nu \times M$ matrix with F_i as the $(i+1)$th block, and $T_{\mathscr{C}}$ as a block Toeplitz matrix consisting of $\{C_i\}$ as shown next:

$$\widehat{\Theta} = \begin{bmatrix} \widehat{H}_0 \\ \widehat{H}_1 \\ \vdots \\ \widehat{H}_L \end{bmatrix}, \quad \underline{F} = \begin{bmatrix} F_0 \\ F_1 \\ \vdots \\ F_{N+L+1} \end{bmatrix}, \quad T_{\mathscr{C}} = \begin{bmatrix} C_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \vdots & & & C_0 \\ C_N & & & \vdots \\ \mathbf{0} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & C_N \end{bmatrix},$$

assuming that $q \geq L$. It can be verified that $\underline{F} = T_{\mathscr{C}} \widehat{\Theta}$ and

$$J_{\mathscr{F}} = \left\| \left[G'_{12} \; G'_{22} \right] T_{\mathscr{H}} \left(\{\widehat{H}_i\} \right) \right\|_{\mathscr{F}} = \sum_{i=0}^{q+L+1} \mathrm{Tr} \left\{ F'_i F_i \right\} = \underline{F}' \underline{F}.$$

Therefore, the constrained minimization of $J_{\mathscr{F}}$ is the same as minimization of

$$J_{\mathscr{F}} = \mathrm{Tr} \left\{ \widehat{\Theta}' T'_{\mathscr{C}} T_{\mathscr{C}} \widehat{\Theta} \right\} \text{ subject to } \widehat{\Theta}' \widehat{\Theta} = I. \tag{8.41}$$

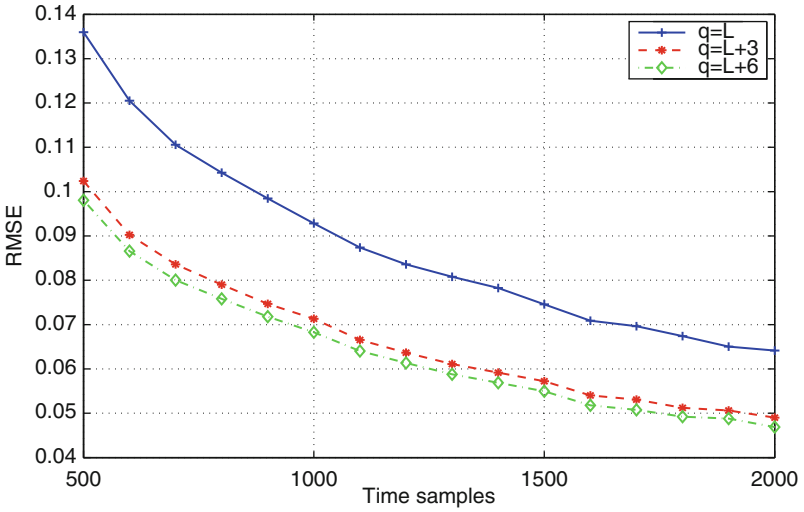


Fig. 8.5 RMSE for blind channel estimation

It is important to observe that $T'_{\mathscr{C}}T_{\mathscr{C}}$ is a block Toeplitz matrix. The minimizer consists of the M right singular vectors $\{\mathbf{v}_i\}$ corresponding to the M smallest nonzero singular values $\{\sigma_i\}$ that can be obtained via SVD of $T_{\mathscr{C}} = USV'$ and $V = [\,\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_{(L+1)M}\,]$. Since the true second-order statistics Σ_y is not available, its sampled version $\hat{\Sigma}_y$ defined in (8.36) has to be used. As a result, $\widehat{\Theta}_{\text{opt}} \neq \Theta$ in general with $\widehat{\Theta}_{\text{opt}}$ the solution to the constrained minimization in (8.41). The performance of the subspace algorithm depends on SNR and on the estimation of the noise subspace.

Example 8.8. Consider the same CIR in Example 8.2 with $P = 2, M = 1$, and $L = 2$. The transmitted signal consists of binary symbols that is white. The SNR, defined as the ratio of the signal power to the noise power, is taken to be 0 dB. The received signal is measured at 2000 times samples and used to compute the sampled second-order statistics Σ_y. The constrained minimization of (8.39) is employed to compute the estimated CIR. Let $\{\widehat{H}_i^{(k)}\}$ be the estimated CIR for the kth emsemble run. The RMSE is computed according to

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{k=1}^{T} \text{Tr}\left\{\sum_{i=0}^{L}\left(H_i - \widehat{H}_i^{(k)}\right)'\left(H_i - \widehat{H}_i^{(k)}\right)\right\}}$$

with a total of $T = 2500$ ensemble runs. The simulation results are plotted in the Fig. 8.5 that shows the improvement when N increases. However the improvement due to large N diminishes as SNR increases that is why a small SNR is used in this example.

The subspace algorithm for blind channel estimation is closely related to the TLS algorithm studied in the previous subsection. Both compute the sampled second-order statistics, and both use eigenvalue decomposition. The difference lies in the Toeplitz structure of the second-order statistics for blind channel estimation that prevents the subspace algorithm from being MLE. Nonetheless, the following result is true.

Theorem 8.9. *Let $\widehat{\Theta}_{\text{opt}}$ be the channel estimate for blind channel estimation based on the subspace algorithm. If the input signal is both temporally with nonsingular Σ_s, then the estimate $\widehat{\Theta}_{\text{opt}}$ converges to the true Θ as the number of time samples approaches infinity* w.p.1.

Proof. The assumption on the input signal implies that (8.36) holds. In addition, the convergence has probability 1 which is the same as that for the TLS algorithm. Hence, the strong convergence holds. □

Thus far, the optimality of the subspace method is not addressed. The main hurdle lies in the structure of the signal model that is not in the same form as the EIV model. It will be shown in the next subsection that the subspace is asymptotically optimal in the sense of MLE.

8.2.3 Graph Space Method

For identification of the plant model in feedback control, the subspace method can also be used to estimate the plant parameters. Let $\mathbf{P}(z)$ be the transfer matrix with m input/p output. It is assumed that $\mathbf{P}(z) = \mathbf{B}(z)\mathbf{A}(z)^{-1}$ with

$$\mathbf{A}(z) = I + \sum_{k=1}^{L} A_k z^{-k}, \quad \mathbf{B}(z) = \sum_{k=0}^{L} B_k z^{-k}. \tag{8.42}$$

Even though the physical system is strictly causal, $B_0 \neq \mathbf{0}$ is assumed which can help to reduce the modeling error. It is further assumed that

$$\text{rank}\left\{\begin{bmatrix}\mathbf{A}(z)\\ \mathbf{B}(z)\end{bmatrix}\right\} = m \quad \forall\, z \in \mathbb{C}.$$

The above ensures that $\{\mathbf{A}(z), \mathbf{B}(z)\}$ are right coprime. Because physical systems in practice are more complex than the linear finite dimensional models, this assumption holds for most real systems.

Let $\{\mathbf{u}(t)\}$, and $\{\mathbf{y}(t)\}$ be the input and output of the system, respectively. The graph space associated with the plant model $\mathbf{P}(z)$ is defined as

$$\underline{\mathbf{G}}_P := \left\{\mathbf{z}(t) = \begin{bmatrix}\mathbf{u}(t)\\ \mathbf{y}(t)\end{bmatrix} \middle| : \ \exists\, \mathbf{w}(t) : \ \mathbf{z}(t) = \begin{bmatrix}\mathbf{A}(q)\\ \mathbf{B}(q)\end{bmatrix}\mathbf{w}(t)\right\}. \tag{8.43}$$

The unknown signal $\{\mathbf{w}(t)\}$ will be referred to auxiliary input. For this reason, an FIR transfer matrix $\mathbf{G}(z)$ can be defined via

$$\mathbf{G}(z) = \sum_{k=0}^{L} G_k z^{-k}, \quad G_k = \begin{bmatrix} A_k \\ B_k \end{bmatrix}, \tag{8.44}$$

where $A_0 = I_m$ is taken. By taking $\mathbf{z}(t)$ as the observation and $\mathbf{w}(t)$ as the unknown input, system identification for the plant model $\mathbf{P}(z)$ is converted to parameter estimation for $\{G_k\}$. As a result, the subspace method from the previous subsection can be employed to estimate the system parameters. To emphasize the graph space of the system and to distinguish it from blind channel estimation, the subspace method used for control system identification is termed as the graph space method.

Denote $\Theta = [G_0' \; G_1' \; \cdots \; G_L']'$ as the parameter matrix of the system. The constraint $\Theta'\Theta = I$ from the subspace method is replace by the first square block of Θ being $A_0 = I$. However, this constraint does not change the estimation algorithm. Let $\mathbf{v}(t)$ be the noise vector comprising both measurement errors at the plant input and output. There holds

$$\mathbf{z}(t) = \sum_{k=0}^{L} G_k \mathbf{w}(t-k) + \mathbf{v}(t) \tag{8.45}$$

that is almost identical to (8.32). Define the block Toeplitz matrix $T_{\mathscr{G}}$ of dimension $(q+1)(p+m) \times (L+q+1)m$ as

$$T_{\mathscr{G}} = \begin{bmatrix} G_0 & \cdots & G_L & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & G_0 & \cdots & G_L \end{bmatrix} \tag{8.46}$$

that is identical to (8.33), except that H_i is replaced by G_i for $0 \le i \le L$. Similarly, denote

$$\underline{\mathbf{z}}(t) = \begin{bmatrix} \mathbf{z}(t) \\ \mathbf{z}(t-1) \\ \vdots \\ \mathbf{z}(t-q) \end{bmatrix}, \quad \underline{\mathbf{v}}(t) = \begin{bmatrix} \mathbf{v}(t) \\ \mathbf{v}(t-1) \\ \vdots \\ \mathbf{v}(t-q) \end{bmatrix}, \quad \underline{\mathbf{w}}(t) = \begin{bmatrix} \mathbf{w}(t) \\ \mathbf{w}(t-1) \\ \vdots \\ \mathbf{w}(t-L_q) \end{bmatrix}$$

with $L_q = L+q$. It follows that at each time sample t,

$$\underline{\mathbf{z}}(t) = T_{\mathscr{G}} \underline{\mathbf{w}}(t) + \underline{\mathbf{v}}(t). \tag{8.47}$$

The observation noise vector $\mathbf{v}(t)$ is assumed to be both spatially and temporally white with noise variance σ^2. The sampled second-order statistics can be computed via

$$\hat{\Sigma}_z = \frac{1}{N}\sum_{t=0}^{N-1} \mathbf{z}(t)\mathbf{z}(t)' \;\longrightarrow\; \Sigma_z = T_{\mathscr{G}}\Sigma_w T_{\mathscr{G}}' + \sigma^2 I$$

as $N \to \infty$. The above convergence has probability 1, and is similar to that for blind channel estimation.

For the graph space method, the persistent excitation (PE) for the input signal $\{\mathbf{u}(t)\}$ needs to be assumed which ensures strictly positivity of Σ_w. Hence, under the PE condition and $\mu = (q+1)p - Lm > 0$, the covariance matrix Σ_z has precisely μ zero eigenvalues. Let $\{\mathbf{x}_i\}_{i=1}^{\mu}$ be the corresponding eigenvectors that span the noise subspace of the sampled second-order statistics. Denote

$$X = \left[\, \mathbf{x}_1 \; \mathbf{x}_2 \; \cdots \; \mathbf{x}_\mu \,\right].$$

The graph space method is aimed at searching for $\{\widehat{G}_i\}$ to minimize

$$J_{\mathscr{F}} = \left\| X' T_{\mathscr{G}}\left(\{\widehat{G}_i\}\right) \right\|_{\mathscr{F}}^2 \;\text{ subject to }\; \left[\, I_m \; \mathbf{0} \,\right] \widehat{G}_0 = I_m. \tag{8.48}$$

The matrices X and $F = X' T_{\mathscr{G}}(\{\widehat{G}_i\})$ have dimensions of $(q+1)(p+m) \times \mu$ and $\mu \times (L+q+1)m$, respectively. Partition these two matrices in accordance with

$$X' = \left[\, X_0 \; X_1 \; \cdots \; X_q \,\right],$$
$$F = \left[\, F_0 \; F_1 \; \cdots \; F_{L+q+1} \,\right],$$

of which each X_i has the dimension $\mu \times (p+m)$ and each F_i has the dimension $\mu \times m$. There holds $F = T_{\mathscr{X}}\widehat{\Theta}$ where

$$\widehat{\Theta} = \begin{bmatrix} \widehat{G}_0 \\ \widehat{G}_1 \\ \vdots \\ \widehat{G}_L \end{bmatrix}, \quad F = \begin{bmatrix} F_0 \\ F_1 \\ \vdots \\ F_{q+L+1} \end{bmatrix}, \quad T_{\mathscr{X}} = \begin{bmatrix} X_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \vdots & & & X_0 \\ X_q & & & \vdots \\ \mathbf{0} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & X_q \end{bmatrix},$$

assuming that $q \geq L$. As a result,

$$J_{\mathscr{F}} = \left\| X' T_{\mathscr{G}} \left(\{ \widehat{G}_i \} \right) \right\|_{\mathscr{F}}^2 = \sum_{i=0}^{q+L+1} \mathrm{Tr} \{ F_i' F_i \} = F'F.$$

Therefore, the constrained minimization of $J_{\mathscr{F}}$ in (8.48) is the same as minimization of

$$J_{\mathscr{F}} = \mathrm{Tr} \left\{ \widehat{\Theta}' T_{\mathscr{X}}' T_{\mathscr{X}} \widehat{\Theta} \right\} \text{ subject to } \left[I_m \ \mathbf{0} \right] \widehat{\Theta} = I_m. \tag{8.49}$$

The minimizer is given by $\widehat{\Theta}_{\text{opt}} = \left[\mathbf{v}_{Lm+1} \ \mathbf{v}_{Lm+2} \cdots \mathbf{v}_{(L+1)m} \right] \Omega^{-1}$, consisting of the m right singular vectors $\{\mathbf{v}_i\}_{i=Lm+1}^{(L+1)m}$ corresponding to the m smallest nonzero singular values $\{\sigma_i\}_{i=Lm+1}^{(L+1)m}$. The normalization matrix Ω is used to satisfy the constraint $\left[I_m \ \mathbf{0} \right] \widehat{\Theta} = I_m$. Similar to the subspace algorithm, the right singular vectors can be computed via SVD.

Theorem 8.10. *Consider square transfer matrix* $\mathbf{P}(z) = \mathbf{B}(z)\mathbf{A}(z)^{-1}$ *with* $\{\mathbf{A}(z), \mathbf{B}(z)\}$ *specified in (8.42). Suppose that* $\mathbf{G}(z)$ *defined in (8.44) is right coprime, the input* $\{\mathbf{u}(t)\}$ *is* PE*, and the noise vectors* $\{\mathbf{v}(t)\}$ *in (8.45) are both spatially and temporally white and Gauss with variance* σ^2*. If the auxiliary input* $\{\mathbf{w}(t)\}$ *is also spatially and temporally white with covariance identity, then the graph space algorithm in this subsection is asymptotically optimal in the sense of* MLE*, and the estimate* $\widehat{\Theta}_{\text{opt}}$ *converges to the true system parameter matrix* Θ *with probability* 1.

Proof. By the hypothesis, $p = m$, although the result is true for $p \neq m$. The right coprime assumption implies the existence of

$$\mathbf{G}_\ell(z) = \sum_{i=0}^{L} \left[A_{\ell_k} \ B_{\ell_k} \right] z^{-k}$$

such that $\mathbf{G}_\ell(z) J \mathbf{G}(z) \equiv \mathbf{0}$ where

$$J = \begin{bmatrix} \mathbf{0} & I_p \\ -I_m & \mathbf{0} \end{bmatrix}, \quad A_{\ell_0} = I.$$

That is, $\mathbf{A}_\ell(z)\mathbf{B}(z) = \mathbf{B}_\ell(z)\mathbf{A}(z)$ with

$$\mathbf{A}_\ell(z) = I + \sum_{i=1}^{L} A_{\ell_k} z^{-k}, \quad \mathbf{B}_\ell(z) = \sum_{i=0}^{L} B_{\ell_k} z^{-k},$$

and thus, $\{\mathbf{A}_\ell(z), \mathbf{B}_\ell(z)\}$ are left coprime and $\mathbf{P}(z) = \mathbf{A}_\ell(z)^{-1}\mathbf{B}_\ell(z)$. Since $A_{\ell_0} = A_0$ by $p = m$, the left coprime factors $\{\mathbf{A}_\ell(z), \mathbf{B}_\ell(z)\}$ are uniquely determined by right coprime factors $\{\mathbf{A}(z), \mathbf{B}(z)\}$ up to a unitary matrix, and vice versa. As a result,

identification of the right coprime factors is the same as that of the left coprime factors. Consider the case $q = L$. Denote

$$\mathbf{z}_J(t) = J\mathbf{z}(t) = \begin{bmatrix} \mathbf{y}(t) \\ -\mathbf{u}(t) \end{bmatrix}.$$

In the noise-free case, there holds the relation

$$\begin{bmatrix} I & \Theta_\ell \end{bmatrix} \underline{\mathbf{z}}_J(t) = \mathbf{0}, \quad \Theta_\ell = \begin{bmatrix} B_{\ell_0} & A_{\ell_1} & B_{\ell_1} & \cdots & A_{\ell_L} & B_{\ell_L} \end{bmatrix}$$

where $\underline{\mathbf{z}}_J(t)$ is the blocked column vector of $\mathbf{z}_J(t)$ with size $(L+1)(p+m)$. Clearly, $\underline{\mathbf{z}}_J(t)$ is permutation of $\underline{\mathbf{z}}(t)$. In the noisy case, an EIV model is resulted in but the elements of the noise matrix are not i.i.d. anymore.

Without loss of generality, the measurements $\underline{\mathbf{z}}_J(t)\}$ at times samples $[t_0,\ t_f]$ are assumed. Hence, the corresponding EIV model is given by

$$\begin{bmatrix} \underline{\mathbf{z}}_J(t_0) & \cdots & \underline{\mathbf{z}}_J(t_f) \end{bmatrix} = \begin{bmatrix} \underline{\mathbf{z}}_J^{(0)}(t_0) & \cdots & \underline{\mathbf{z}}_J^{(0)}(t_f) \end{bmatrix} + \begin{bmatrix} \underline{\mathbf{v}}_J(t_0) & \cdots & \underline{\mathbf{v}}_J(t_f) \end{bmatrix}$$

with $\underline{\mathbf{z}}_J^{(0)}(t)$ the noise-free blocked graph signal at time t. Indeed, the elements of the noise matrix are not i.i.d., because $\underline{\mathbf{v}}_J(t)$ is a blocked column vector of $\mathbf{v}_J(t) = J\mathbf{v}(t)$ consisting of $\{\mathbf{v}(t-i)\}_{i=0}^{N}$. It follows that the TLS solution is not an MLE for Θ_ℓ. On the other hand, let

$$\varepsilon_N^2 = \frac{1}{N} \sum_{t=t_0}^{t_f} \left\| \underline{\mathbf{z}}_J(t) - \underline{\mathbf{z}}_J^{(0)}(t) \right\|^2, \quad N = t_f - t_0 + 1.$$

Since $\mathbf{v}_J(t) = \underline{\mathbf{z}}_J(t) - \underline{\mathbf{z}}_J^{(0)}(t)$, it can be verified that

$$\varepsilon_N^2 = \frac{q}{N} \sum_{t=t_0}^{t_f-q} \|\mathbf{v}_J(t)\|^2 + \frac{1}{N} \sum_{i=1}^{q-1} \left[(q-i)\|\mathbf{v}_J(t_0-i)\|^2 + i\|\mathbf{v}_J(t_f-i+1)\|^2 \right].$$

Recall $q = L$ that is fixed and finite. The second summation on the right-hand side of the above equation approaches zero as $N \to \infty$. Hence, the TLS solution minimizes ε_T^2 asymptotically in the same spirit of MLE. In addition, the white assumption on $\mathbf{w}(t)$ leads to

$$\frac{1}{N} \sum_{t=t_0}^{t_f} \underline{\mathbf{z}}(t)\underline{\mathbf{z}}(t)' \ \longrightarrow \ Q_J T_{\mathscr{G}} T_{\mathscr{G}}' Q_J' + \sigma^2 I$$

for some permutation matrix Q_J dependent on J. The right-hand side is deterministic. Consequently, the TLS solution for $\widehat{\Theta}_\ell$, and thus the graph space estimate $\widehat{\Theta}_{\text{opt}}$, are indeed the asymptotic MLE. The convergence with probability 1 follows from

the PE condition and identity covariance of $\{\mathbf{w}(t)\}$. The proof for the case of $q > L$ can be covered by adding zero blocks to Θ_ℓ. The proof is now complete. □

It is known that the error covariance for MLE approaches the CRLB asymptotically under certain regularity condition. To compute the CRLB, it is necessary to obtain first the corresponding FIM. Denote $f_V(\{\mathbf{z}(t)\})$ as the joint PDF. By the Gauss assumption and the signal model in (8.45),

$$\ln f_V(\{\mathbf{z}(t)\}) \sim J = -\frac{1}{2\sigma^2}\sum_{t=t_0}^{t_f} \mathrm{Tr}\left\{\left[\mathbf{z}(t)-\tilde{\Theta}\mathbf{w}(t)\right]\left[\mathbf{z}(t)-\tilde{\Theta}\mathbf{w}(t)\right]'\right\}, \tag{8.50}$$

where $\tilde{\Theta} = \left[\, G_0\ G_1\ \cdots\ G_L \,\right]$. Denote

$$\vartheta = \begin{bmatrix} \mathrm{vec}(G_0) \\ \vdots \\ \mathrm{vec}(G_L) \end{bmatrix}, \quad \underline{\mathbf{z}}_N = \begin{bmatrix} \mathbf{z}(t_f) \\ \vdots \\ \mathbf{z}(t_0) \end{bmatrix},$$

$$\underline{\mathbf{w}}_{T+L} = \begin{bmatrix} \mathbf{w}(t_f) & \vdots \\ \mathbf{w}(t_0-L) \end{bmatrix}, \quad \underline{\mathbf{v}}_N = \begin{bmatrix} \mathbf{v}(t_f) \\ \vdots \\ \mathbf{v}(t_0) \end{bmatrix}.$$

Direct calculation yields

$$\begin{aligned} \frac{\partial J}{\partial \vartheta} &= \frac{1}{\sigma^2}\begin{bmatrix} \mathbf{w}_1(t_f)I_{p+m} & \cdots & \mathbf{w}_1(t_0)I_{p+m} \\ \vdots & \cdots & \vdots \\ \mathbf{w}_{(L+1)m}(t_f)I_{p+m} & \cdots & \mathbf{w}_{(L+1)m}(t_0)I_{p+m} \end{bmatrix}\begin{bmatrix} \mathbf{v}(t_f) \\ \vdots \\ \mathbf{v}(t_0) \end{bmatrix} \\ &=: \frac{1}{\sigma^2} M_{\mathbf{w}}\underline{\mathbf{v}}_N \end{aligned} \tag{8.51}$$

where $\mathbf{w}_k(t)$ is the kth component of $\mathbf{w}(t)$. A caution needs to be taken to that the first m rows of ϑ are the same as I_m which are known. Let $\tilde{\vartheta}$ be obtained from ϑ by deleting the m^2 known elements. Then

$$\frac{\partial J(\tilde{\Theta})}{\partial \tilde{\vartheta}} = \frac{1}{\sigma^2}\tilde{M}_{\mathbf{w}}\underline{\mathbf{v}}_N, \quad \tilde{M}_{\mathbf{w}} = \begin{bmatrix} Z & 0 \\ 0 & I_{mpL} \end{bmatrix} M_{\mathbf{w}}, \tag{8.52}$$

and $Z = I_m \otimes \left[\, \mathbf{0}\ I_p \,\right]$.

It is important to notice that the auxiliary input $\{\mathbf{w}(t)\}$ is also unknown. Its impact to the CRLB needs to be taken into account. Let $T_{\mathscr{G}}(\Theta)$ be the same as in (8.46) but with blocking size $q = N$. There holds

$$\underline{\mathbf{z}}_N = T_{\mathscr{G}}(\Theta)\underline{\mathbf{w}}_{N+L} + \underline{\mathbf{v}}_N.$$

Then the likelihood function in (8.50) can be written alternatively as

$$J(\Theta,\mathbf{w}) = -\frac{1}{2\sigma^2}\left[\underline{\mathbf{z}}_N - T_{\mathscr{G}}(\Theta)\underline{\mathbf{w}}_{N+L}\right]'\left[\underline{\mathbf{z}}_N - T_{\mathscr{G}}(\Theta)\underline{\mathbf{w}}_{N+L}\right]. \tag{8.53}$$

Thus, the partial derivative of $J(\Theta,\mathbf{w})$ with respect to $\underline{\mathbf{w}}_{N+L}$ is given as

$$\frac{\partial J(\Theta,\mathbf{w})}{\partial \underline{\mathbf{w}}_{N+L}} = \frac{1}{\sigma^2}T_{\mathscr{G}}(\Theta)'\left[\underline{\mathbf{z}}_N - T_{\mathscr{G}}(\Theta)\underline{\mathbf{w}}_{N+L}\right] = \frac{1}{\sigma^2}T_{\mathscr{G}}(\Theta)'\underline{\mathbf{v}}_N. \tag{8.54}$$

To compute the FIM, the following matrices

$$\begin{aligned}
\mathrm{FIM}(\Theta) &= \mathrm{E}\left\{\frac{\partial J(\Theta,\mathbf{w})}{\partial \tilde{\vartheta}}\frac{\partial J(\Theta,\mathbf{w})}{\partial \tilde{\vartheta}'}\right\} = \frac{1}{\sigma^2}\tilde{M}_{\mathbf{w}}\tilde{M}'_{\mathbf{w}},\\
\mathrm{E}\left\{\frac{\partial J(\Theta,\mathbf{w})}{\partial \tilde{\vartheta}}\frac{\partial J(\Theta,\mathbf{w})}{\partial \mathbf{w}'}\right\} &= \frac{1}{\sigma^2}\tilde{M}_{\mathbf{w}}T_{\mathscr{G}}(\Theta),\\
\mathrm{FIM}(\mathbf{w}) &= \mathrm{E}\left\{\frac{\partial J(\Theta,\mathbf{w})}{\partial \mathbf{w}}\frac{\partial J(\Theta,\mathbf{w})}{\partial \mathbf{w}'}\right\} = \frac{1}{\sigma^2}T_{\mathscr{G}}(\Theta)'T_{\mathscr{G}}(\Theta)
\end{aligned}$$

are useful. The CRLB for estimation of Θ can be obtained according to

$$\mathrm{CRB}(\Theta) = \sigma^2\left(\mathrm{FIM}(\Theta) - \tilde{M}_{\mathbf{w}}T_{\mathscr{G}}(\Theta)\left[T_{\mathscr{G}}(\Theta)'T_{\mathscr{G}}(\Theta)\right]^{-1}T_{\mathscr{G}}(\Theta)'\tilde{M}'_{\mathbf{w}}\right)^{-1}. \tag{8.55}$$

The above CRLB can be difficult to compute, if N, the number of time samples, is large, in light of the fact that the blocked Toeplitz matrix $T_{\mathscr{G}}(\Theta)$ has size $(N+1)(p+m)\times(N+L+1)m$. It is left as an exercise to derive an efficient algorithm for computing the CRLB (Problem 8.16).

Example 8.11. Consider the SISO plant model given by

$$P(z) = \frac{1.4496z^{-1} + 0.7248z^{-2}}{1 - 1.5z^{-1} + 0.7z^{-2}}.$$

This plant model has the same poles and zeros as the one in Example 8.3. The difference lies in the gain factor of 4. A total of $N = 3{,}000$ input and output measurements are generated by taking the auxiliary input $\{\mathbf{w}(t)\}$ as white Gauss of variance one. The resulting $\{u(t)\}$ and $\{y(t)\}$ are WSS and admit variance of 5.672 and 4.264 dB, respectively. The measurement error $\{v(t)\}$ is also taken as white Gauss with variance one, implying that the SNR for the input and output signals is 5.672 and 4.264 dB, respectively. Using the graph space method, the estimation errors are plotted against the CRLB similar to that in Example 8.8. A total of 2500 ensemble runs are used to compute the RMSE value of the estimation error. The RMSE is seen to converge to the CRLB, albeit slowly. In fact, a larger number of time samples are used in order to see such a convergence.

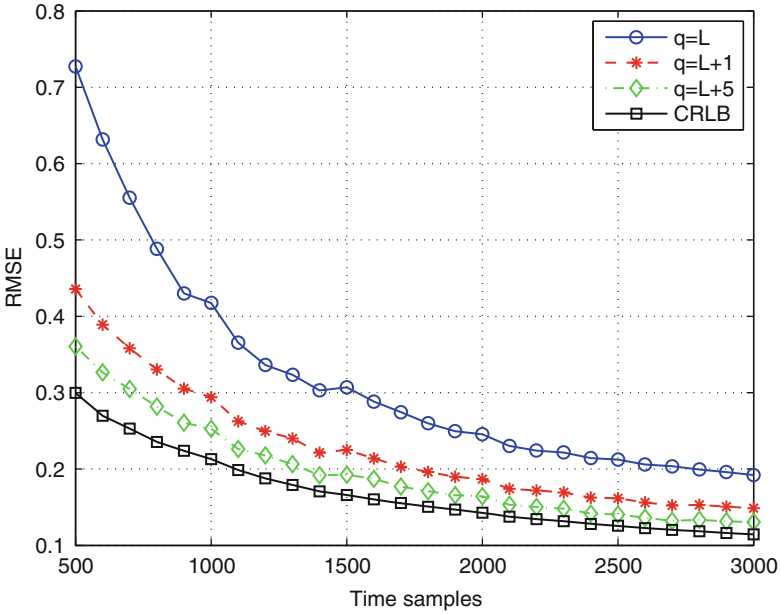


Fig. 8.6 RMSE curves for identification using the graph space method

The simulation results show that as q increases, the RMSE value decreases. However, the largest drop of the RMSE value occurs at $q = L+1$. As N increases beyond $L+1$, the decrease of the RMSE values slows down dramatically that is why the RMSE values are shown for only the cases of $q = L, L+1, L+5$. It needs to be pointed out that the white assumption for auxiliary input $\{\mathbf{w}(t)\}$ is important in order to achieve asymptotic MLE. Figure 8.6 shows the case when the input signal $\{u(t)\}$ is white with variance one, and both $\{y(t)\}$ and $\{w(t)\}$ become colored signals with variances 15.78 and 9.28, respectively. Under the same statistical noise $\{v(t)\}$, the total SNR is greater than the previous case. However, the simulation results in Fig. 8.7 shows that the RMSE values resulted from the graph space method do not converge to the CRLB. In fact, the larger the q, the worse the RMSE performance. The simulation results in this example indicates the importance of the auxiliary input being white, in order to obtain asymptotic MLE, which is consistent with the theoretical result in Theorem 8.10. Since white auxiliary input is not possible to generate prior to system identification, it is suggested to apply the LS algorithm first for system identification. After a reasonably good plant model is identified, white $\{\mathbf{w}(t)\}$ can be generated, and the plant input $\mathbf{u}(t) = \widehat{\mathbf{A}}(q)\mathbf{w}(t)$ can be obtained using the estimate model, which can then be applied as the exogenous input. Once the output measurements are available, the graph space algorithm can be applied to obtain more accurate identification results.

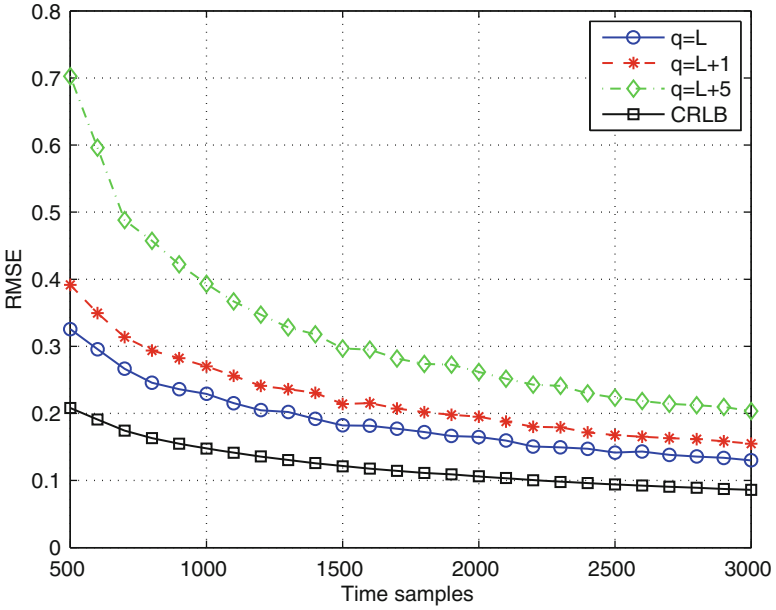


Fig. 8.7 RMSE curves for (graph space) identification with white input

Let σ_u^2 and σ_y^2 be the noise variances of the input and output, respectively. This subsection assumes $\sigma_u = \sigma_y = \sigma$ thus far. In the case when $\sigma_u \neq \sigma_y$,

$$\mathbf{z}_r(t) = \begin{bmatrix} r\mathbf{u}(t) \\ \mathbf{y}(t) \end{bmatrix}, \quad r = \frac{\sigma_y}{\sigma_u},$$

can be employed to replace $\mathbf{z}(t)$ in (8.45). The noise vectors associated with $\{\mathbf{z}_r(t)\}$ admit covariance $\sigma_y^2 I$, and thus, the graph space method can be applied to $\{\mathbf{z}_r(t)\}$ which estimates $r\mathbf{A}(z)$ and $\mathbf{B}(z)$. A more serious issue is how to estimate the variances σ_u^2 and σ_y^2. Methods have been developed in the research literature, and are not pursued in this book.

Notes and References

LS algorithm is presented in almost every textbook on system identification. See for instance [78,102]. The RLS algorithm can not only be found in system identification books but also in adaptive control books [7, 38]. The TLS algorithm has a shorter history. A good source is [35, 113]. It is basically the same as the bias elimination LS [49, 103, 104]. Blind channel estimation based on subspace in [87] is connected to TLS. See also [51, 110] for blind channel estimation. The graph space method in this chapter is generalized from the subspace method in [87].

Exercises

8.1. Let A, B, and X be real matrices with compatible dimensions. Show that

$$\frac{\partial \mathrm{Tr}\{AXB\}}{\partial X} = A'B', \quad \frac{\partial \mathrm{Tr}\{AX'B\}}{\partial X} = BA,$$
$$\frac{\partial \mathrm{Tr}\{AXBX'\}}{\partial X} = A'XB' + AXB.$$

8.2. Suppose that $p = m = 1$, and thus, (8.2) is reduced to $y(t) = \underline{\phi}(t)'\theta$ where $\Theta = \theta'$. Show that the RLS algorithm in Sect. 8.1.1 can be derived with Kalman filter. (*Hint:* Use $\mathbf{x}(t) = \theta$ as the state vector, and thus,

$$\mathbf{x}(t+1) = A_t\mathbf{x}(t), \quad y(t) = \mathbf{c}_t\mathbf{x}(t) + v(t)$$

with $A_t = I$, $\mathbf{c}_t = \underline{\phi}(t)'$, and $\sigma_t^2 = \mathrm{E}\left\{|v(t)|^2\right\}$.)

8.3. Let P_t be updated in (8.8). Show that the RLS algorithm in the previous problem can be obtained from minimizing

$$J_\theta = \left(\widehat{\theta}_{t+1} - \widehat{\theta}_t\right)' P_t^{-1} \left(\widehat{\theta}_{t+1} - \widehat{\theta}_t\right) + \sigma_t^{-2}\left[y(t) - \underline{\phi}(t)'\widehat{\theta}_{t+1}\right]^2.$$

8.4. Show that $(D - CA^{-1}B)^{-1} = D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1}$. (*Hint:* Recall the inverse of the square transfer matrix:

$$[D + C(zI - A)^{-1}B]^{-1} = D^{-1} - D^{-1}C(zI - A + BD^{-1}C)^{-1}BD^{-1}$$

and then set $z = 0$.)

8.5. For the LS solution $\Theta_{\mathrm{LS}} = Y_{0,f}\Phi_{0,f}'(\Phi_{0,f}\Phi_{0,f}')^{-1}$ in Theorem 8.1, denote $\underline{\theta}_i$ and $\widehat{\underline{\theta}}_i$ as the ith row of Θ and Θ_{LS}, respectively. Assume that $\Phi_{0,f}$ is noise-free. Show that

$$\text{(i) } \mathrm{E}\left\{\widehat{\underline{\theta}}_i\right\} = \underline{\theta}_i, \quad \text{(ii) } \mathrm{E}\left\{\left(\widehat{\underline{\theta}}_i - \underline{\theta}_i\right)'\left(\widehat{\underline{\theta}}_i - \widehat{\underline{\theta}}_i\right)\right\} = \sigma_i^2(\Phi_{0,f}\Phi_{0,f}')^{-1}.$$

8.6. Prove the RLS algorithm with the forgetting factor in (8.10) and (8.11).

8.7. Program the RLS algorithm, and the RLS algorithm with forgetting factor. Use the following transfer function as a testing example:

$$P(z) = \frac{1.4496z^{-1} + b_t z^{-2}}{1 - 1.5z^{-1} + a_t z^{-2}}$$

where both a_t and b_t lie in the interval of $[0.6,\ 0.8]$ and change slowly. The forgetting factor can be generated via $\gamma_t = \gamma_0\gamma_{t-1} + (1 - \gamma_0)$ with $\gamma_0 \in [0.95,\ 0.99]$.

8.8. Consider the input/output measurement model in Fig. 8.2. Show that the LS solution to system identification is MLE, if the plant model has an FIR structure and input is noise-free. What happens when the plant input is not noise-free?

8.9. Consider the input/output measurement model in Fig. 8.2. (i) If the plant input is noise-free, show that the LS solution is asymptotically unbiased, and (ii) if the plant input involves noise, show that the corresponding RMSE depends on the system parameters.

8.10. For the signal model in (8.2) for $t \in [t_0,\ t_f]$, arising from the system input/output description in (8.1) in which the observation noises corrupt both input and output signals, show that

1. The matrix $\Phi_{0,f}$ in $Y_{0,f} = \Theta\Phi_{0,f} + V_{0,f}$ involves observation noises if $M_k \neq \mathbf{0}$ for at least one $k > 0$.
2. The LS solution can be written as

$$\Theta_{\text{LS}} = \Theta + V_{0,f}\Phi'_{0,f}(\Phi_{0,f}\Phi'_{0,f})^{-1}.$$

3. Show that Θ_{LS} is a biased estimate of Θ, if $M_k \neq \mathbf{0}$ for at least one $k > 0$.

(*Hint:* $\text{E}\{V_{0,f}\Phi'_{0,f}(\Phi_{0,f}\Phi'_{0,f})^{-1}\} \neq \mathbf{0}$, because the noise components of $\{\mathbf{v}(t)\}$ corrupted to $\{\mathbf{y}(t)\}$ cannot be removed from $\Phi_{0,f}$, if $M_k \neq \mathbf{0}$ for at least one $k > 0$.)

8.11. Consider partition of the eigenvector matrix G and eigenvalue matrix Λ in (8.21).

(i) Show that

$$G'_{12}G_{11} + G'_{22}G_{21} = \mathbf{0}.$$

(ii) Show that G_{11} is nonsingular, if and only if G_{22} is nonsingular.
(iii) Show that Θ_{TLS} in (8.22) is indeed the TLS solution.

8.12. Show that the optimal solution to (8.30) is the generalized TLS solution.

8.13. Consider minimization of $J_{A,B}$ in (8.19). Show that $\min J_{A,B} = \text{Tr}\{\Lambda_2\}$ where Λ_2 of dimension $p \times p$ is defined in (8.21).

8.14. For the case $p = 1$, the bias-eliminating LS estimate is given by

$$\widehat{\Theta} = Y_{0,f}\Phi'_{0,f}(\Phi_{0,f}\Phi'_{0,f} - \widehat{\lambda}^2_{\min}I)^{-1}$$

where $\widehat{\lambda}_{\min}$ is the minimum eigenvalue of

$$\hat{\Sigma}_z = \frac{1}{N}\sum_{t=0}^{N-1}\mathbf{z}(t)\mathbf{z}(t)'$$

with $\mathbf{z}(t)$ the same as in (8.43). Show that the bias-eliminating LS estimate is the same as the TLS solution.

8.15. Show that the TLS solution to $Y_{0,f} \approx \Theta \Phi_{0,f}$ minimizes

$$J_{\mathrm{TLS}}(T) := \mathrm{Tr}\left\{ \left(Y_{0,f} - \Theta\Phi_{0,f}\right)' \left(I + \Theta\Theta'\right)^{-1} \left(Y_{0,f} - \Theta\Phi_{0,f}\right) \right\}.$$

8.16. The CRLB in (8.55) is difficult to compute, if the time horizon $[t_0,\ t_f]$ is large. This exercise provides a guideline on an efficient algorithm for computing the CRLB in (8.55) in the case of large time horizon:

1. Show that $T_{\mathscr{G}}(\Theta)' T_{\mathscr{G}}(\Theta) = \tilde{T}_{\mathscr{G}}(\Theta)' \tilde{T}_{\mathscr{G}}(\Theta) - T_G' T_G$ where

$$\tilde{T}_{\mathscr{G}}(\Theta) = \begin{bmatrix} G_L & & \\ \vdots & \ddots & \\ \vdots & & G_L \\ G_0 & & \vdots \\ & \ddots & \vdots \\ & & G_0 \end{bmatrix},$$

$$T_G = \begin{bmatrix} G_L & & & & \\ \vdots & \ddots & & & \\ G_1 & \cdots & G_L & & \\ & & & G_0 \cdots & G_{L-1} \\ & & & \ddots & \vdots \\ & & & & \tilde{g}_0 \end{bmatrix}$$

   are of dimension $(t_f - t_0 + 2L + 1)(p+m) \times (t_f - t_0 + L + 1)m$ and $2(p+m)L \times (t_f - t_0 + L + 1)m$, respectively.
2. Denote $\Psi = \tilde{T}_{\mathscr{G}}(\Theta)' \tilde{T}_{\mathscr{G}}(\Theta)$. Show that

$$\begin{aligned} [T_{\mathscr{G}}(\Theta)' T_{\mathscr{G}}(\Theta)]^{-1} &= (\Psi - T_G' T_G)^{-1} \\ &= \Psi^{-1} + \Psi^{-1} T_G' (I_{2(p+m)L} - T_G \Psi^{-1} T_G')^{-1} T_G \Psi^{-1}. \end{aligned}$$

3. Let $\Psi^{-1} = \tilde{T}_{-1}' \tilde{T}_{-1}$ be Cholesky factorization such that $\tilde{T}_{-1}$ is block lower triangular with block size $m \times m$. Denote $\Omega = T_G \tilde{T}_{-1}'$ and $\Gamma = \tilde{M}_w T_{\mathscr{G}}(\Theta) \tilde{T}_{-1}'$. Show that

$$\mathrm{CRB}(\Theta) = \sigma^2 \left(\mathrm{FIM}(\Theta) - \Gamma\Gamma' - \Gamma\Omega' (I - \Omega\Omega')^{-1} \Omega\Gamma' \right)^{-1}.$$

   The efficient computation of $\mathrm{CRB}(\Theta)$ is hinged to the Cholesky factorization of $\Psi^{-1} = \tilde{T}_{-1}' \tilde{T}_{-1}$ that will be worked out in the next problem.

8.17. (i) For $\{\mathbf{A}(z), \mathbf{B}(z)\}$ in (8.42) which are right coprime, show that the existence of the spectral factorization

$$\mathbf{A}(z)^{\sim}\mathbf{A}(z) + \mathbf{B}(z)^{\sim}\mathbf{B}(z) = \mathbf{C}(z)^{\sim}\mathbf{C}(z)$$

where $\mathbf{C}(z)$ is the right spectral factor with size $m \times m$ and given by

$$\mathbf{C}(z) = \sum_{k=0}^{L} C_k z^{-k}.$$

(ii) Show that the Toeplitz matrix $\Psi = T_{\mathscr{G}}(\Theta)' T_{\mathscr{G}}(\Theta)$ in the previous problem corresponds to spectral factorization of (Sect. C.3 in Appendix C)

$$\left[z^{-L}\mathbf{G}\left(z^{-1}\right)\right]^{\sim}\left[z^{-L}\mathbf{G}\left(z^{-1}\right)\right] = \left[z^{-L}\mathbf{C}\left(z^{-1}\right)\right]^{\sim}\left[z^{-L}\mathbf{C}\left(z^{-1}\right)\right].$$

(iii) Show that $\tilde{T}_{-1}$ in the previous problem is lower block Toeplitz, and consists of the impulse response of $[z^{-L}\mathbf{C}(z^{-1})]^{-1}$.