Appendix

A

Least-Squares Method

We briefly review the least-squares method for a linear regression model, together with its relation to the LQ decomposition.

A.1 Linear Regressions

Suppose that there exists a linear relation between the output variable $y(t)$ and the d-dimensional regression vector $\varphi(t) = [\varphi_1(t)\ \varphi_2(t)\ \cdots\ \varphi_d(t)]^{\mathrm{T}}$. We assume that N observations $\{y(t),\ \varphi(t),\ t = 0,\ 1,\ \cdots,\ N-1\}$ are given. Then, it follows that

$$y(t) = \varphi_1(t)\theta_1 + \cdots + \varphi_d(t)\theta_d + e(t), \qquad t = 0,\ 1,\ \cdots,\ N-1 \tag{A.1}$$

where $e(t)$ denotes the measurement noise, or the variation in $y(t)$ that cannot be explained by means of $\varphi_1(t),\ \cdots,\ \varphi_d(t)$. We also assume that $\varphi_1(t),\ \cdots,\ \varphi_d(t)$ have no uncertainties[1].

For simplicity, we define the stacked vectors

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{bmatrix}, \qquad y = \begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(N-1) \end{bmatrix}, \qquad e = \begin{bmatrix} e(0) \\ e(1) \\ \vdots \\ e(N-1) \end{bmatrix}$$

and the matrix

$$\Phi = \begin{bmatrix} \varphi_1(0) & \varphi_2(0) & \cdots & \varphi_d(0) \\ \varphi_1(1) & \varphi_2(1) & \cdots & \varphi_d(1) \\ \vdots & \vdots & & \vdots \\ \varphi_1(N-1) & \varphi_2(N-1) & \cdots & \varphi_d(N-1) \end{bmatrix} = \begin{bmatrix} \varphi^{\mathrm{T}}(0) \\ \varphi^{\mathrm{T}}(1) \\ \vdots \\ \varphi^{\mathrm{T}}(N-1) \end{bmatrix}$$

where $\Phi \in \mathbb{R}^{N \times d}$. Then (A.1) can be written as

[1] If φ are also subject to noises, (A.1) is called an errors-in-variables model.

$$y = \Phi\theta + e$$

This is referred to as a linear regression model. The regression analysis involves the estimation of unknown parameters and the analysis of residuals.

The basic assumptions needed for the least-squares method are listed below.

A1) The error vector e is uncorrelated with Φ and θ.

A2) The error vector is a random vector with mean zero.

A3) The covariance matrix of the error vector is $\sigma^2 I_N$ with $\sigma^2 > 0$.

A4) The column vectors of Φ are linearly independent, *i.e.*, $\mathrm{rank}(\Phi) = d$.

Under the above assumptions, we consider the least-squares problem minimizing the quadratic performance index

$$J(\theta) := \sum_{t=0}^{N-1} [y(t) - \varphi^{\mathrm{T}}(t)\theta]^2 = \|y - \Phi\theta\|^2$$

Setting the gradient of $J(\theta)$ with respect to θ to zero yields

$$\left(\sum_{t=0}^{N-1} \varphi(t)\varphi^{\mathrm{T}}(t) \right) \theta = \sum_{t=0}^{N-1} \varphi(t) y(t) \quad \Rightarrow \quad (\Phi^{\mathrm{T}}\Phi)\theta = \Phi^{\mathrm{T}} y \tag{A.2}$$

This is a well-known normal equation.

From Assumption A4), we see that $\Phi^{\mathrm{T}}\Phi \in \mathbb{R}^{d\times d}$ is nonsingular. Thus, solving (A.2), the least-squares estimate is given by

$$\hat{\theta}_{\mathrm{LS}} := \left(\sum_{t=0}^{N-1} \varphi(t)\varphi^{\mathrm{T}}(t) \right)^{-1} \sum_{t=0}^{N-1} \varphi(t) y(t) = (\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}} y \tag{A.3}$$

Also, from Assumptions A1) and A2),

$$\begin{aligned} E\{\hat{\theta}_{\mathrm{LS}}\} &= E\{(\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}}(\Phi\theta + e)\} \\ &= \theta + (\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}} E\{e\} = \theta \end{aligned} \tag{A.4}$$

so that the least-squares estimate $\hat{\theta}_{\mathrm{LS}}$ is unbiased. It follows from Assumption A3) that the error covariance matrix of the estimate $\hat{\theta}_{\mathrm{LS}}$ is

$$\mathrm{cov}\{\hat{\theta}_{\mathrm{LS}}\} = E\{[\theta - \hat{\theta}_{\mathrm{LS}}][\theta - \hat{\theta}_{\mathrm{LS}}]^{\mathrm{T}}\} = \sigma^2(\Phi^{\mathrm{T}}\Phi)^{-1}$$

Moreover, define the residual vector as $\varepsilon := y - \Phi\hat{\theta}_{\mathrm{LS}}$. Then, it follows that

$$\varepsilon = [I_N - \Phi(\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}}]y = [I_N - \Phi(\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}}]e \tag{A.5}$$

It should be noted that $\Pi := \Phi(\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}}$ satisfies $\Pi^2 = \Pi$ and $\Pi = \Pi^{\mathrm{T}}$, so that Π is an orthogonal projection onto $\mathrm{Im}(\Phi)$. Also, $Q := I_N - \Pi$ is an orthogonal

projection onto the orthogonal complement $(\mathrm{Im}\,\Phi)^{\perp} = \mathrm{Ker}(\Phi^{\mathrm{T}})$. Then, $\|\varepsilon\|^2 = y^{\mathrm{T}}(I_N - \Pi)y$ denotes the square of the minimum distance between the point y and the space $\mathrm{Im}(\Phi^{\mathrm{T}})$.

We compute the variance of the residual. From (A.5),

$$\begin{aligned} E\{\|\varepsilon\|^2\} &= \mathrm{trace} E\{\varepsilon\varepsilon^{\mathrm{T}}\} \\ &= \mathrm{trace}\left([I_N - \Phi(\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}}]E\{ee^{\mathrm{T}}\}[I_N - \Phi(\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}}]\right) \\ &= \sigma^2 \mathrm{trace}[I_N - \Phi(\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}}] \\ &= \sigma^2[\mathrm{trace}(I_N) - \mathrm{trace}(\Phi(\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}})] = \sigma^2(N-d) \end{aligned}$$

Hence, the unbiased estimate of the variance σ^2 is given by

$$s^2 := \frac{1}{N-d}\sum_{t=0}^{N-1}\varepsilon^2(t) = \frac{1}{N-d}\|\varepsilon\|^2$$

In practice, the above assumptions A1) $\sim$ A4) are not completely satisfied. If either A1) or A2) is not satisfied, then a bias arises in the least-squares estimate. In fact, in the computation of (A.4), we have

$$E\{\hat{\theta}_{\mathrm{LS}}\} = \theta + E\{(\Phi^{\mathrm{T}}\Phi)^{-1}\Phi^{\mathrm{T}}e\} \neq \theta$$

Suppose that $E\{ee^{\mathrm{T}}\} = R > 0$, so that A3) does not hold. In this case, we consider a weighted least-squares problem of minimizing

$$J(\theta) = \|y - \Phi\theta\|^2_{R^{-1}} = (y - \Phi\theta)^{\mathrm{T}}R^{-1}(y - \Phi\theta)$$

By using the same technique of deriving the least-squares estimate $\hat{\theta}_{\mathrm{LS}}$, we can show that the optimal estimate is given by

$$\hat{\theta}_{\mathrm{GLS}} := (\Phi^{\mathrm{T}}R^{-1}\Phi)^{-1}\Phi^{\mathrm{T}}R^{-1}y$$

where $\hat{\theta}_{\mathrm{GLS}}$ is called the generalized least-squares estimate. The corresponding error covariance matrix becomes

$$\mathrm{cov}\{\hat{\theta}_{\mathrm{GLS}}\} = (\Phi^{\mathrm{T}}R^{-1}\Phi)^{-1}$$

We now turn to Assumption A4). In real problems, we often encounter the case where there exist some "approximate" linear relations among regression vectors (column vectors of Φ); this is called a multicolinearity problem in econometrics. In this case, one or more eigenvalues of $\Phi^{\mathrm{T}}\Phi$ get closer to zero, so that the condition number $\kappa(\Phi)$ becomes very large, leading to unreliable least-squares estimates. An SVD-based method of solving a least-squares problem under ill-conditioning is introduced in Section 2.7. There are also other methods to solve ill-conditioned least-squares problems, including regularization methods, the ridge regression, *etc*.

Example A.1. Consider the normal equation of (A.2):

$$(\Phi^{\mathrm{T}}\Phi)\theta = \Phi^{\mathrm{T}}y, \qquad \Phi \in \mathbb{R}^{N\times d}, \qquad y \in \mathbb{R}^N \tag{A.6}$$

We show that (A.6) has always a solution for any $y \in \mathbb{R}^N$. It is well known that (A.6) is solvable if and only if the vector $\Phi^{\mathrm{T}}y$ belongs to $\mathrm{Im}(\Phi^{\mathrm{T}}\Phi)$. However, this is easily verified by noting that $\Phi^{\mathrm{T}}y \in \mathrm{Im}(\Phi^{\mathrm{T}}) = \mathrm{Im}(\Phi^{\mathrm{T}}\Phi)$.

By direct manipulation, we can show that $\theta = \Phi^{\dagger}y$ is a solution of the normal equation, where $\Phi^{\dagger}$ is the pseudo-inverse defined in Lemma 2.10. Indeed, we have

$$(\Phi^{\mathrm{T}}\Phi)\Phi^{\dagger}y = \Phi^{\mathrm{T}}\Phi\Phi^{\dagger}y = \Phi^{\mathrm{T}}(\Phi^{\dagger})^{\mathrm{T}}\Phi^{\mathrm{T}}y = \Phi^{\mathrm{T}}(\Phi^{\mathrm{T}})^{\dagger}\Phi^{\mathrm{T}}y = \Phi^{\mathrm{T}}y$$

where the Moore-Penrose condition (iii) is used (see Problem 2.9). Also, the general solution

$$\theta = \Phi^{\dagger}y + (I_N - \Phi^{\dagger}\Phi)z, \qquad \forall\, z \in \mathbb{R}^d$$

satisfies the normal equation. □

Let $\boldsymbol{\Theta}$ be the set of minimizers

$$\boldsymbol{\Theta} := \{\, \theta \mid \|y - \Phi\theta\| = \min \}$$

Then, we can show that

1. If θ is a minimizer, *i.e.* $\theta \in \boldsymbol{\Theta}$, then $\Phi^{\mathrm{T}}(y - \Phi\theta) = 0$, and *vice versa*.
2. If $\mathrm{rank}(\Phi) = d$, then $\boldsymbol{\Theta} = \{\hat{\theta}_{\mathrm{LS}}\}$, a singleton.
3. The set $\boldsymbol{\Theta}$ is convex.
4. The set $\boldsymbol{\Theta}$ has a unique minimum norm solution $\theta = \Phi^{\dagger}y$.

We apply the regression analysis technique to an ARX model[2], leading to a least-squares identification method, which is one of the simplest methods for a realistic identification problem.

Example A.2. Consider an ARX model

$$A(z)y(t) = B(z)u(t) + e(t) \tag{A.7}$$

where the unknown parameters are $\theta := (a_1 \;\; \cdots \;\; a_n \;\; b_1 \;\; \cdots \;\; b_m)^{\mathrm{T}}$ and the noise variance σ^2. This is also called an equation error model, which is most easily identified by using the least-squares method. It should be noted that the ARX model of (A.7) is derived from (1.1) by setting $H(z) = 1/A(z)$.

From (A.7), the prediction error is given by

$$\varepsilon(t,\theta) := A(z,\theta)y(t) - B(z,\theta)u(t) = y(t) - \varphi^{\mathrm{T}}(t)\theta \tag{A.8}$$

where $\varphi(t)$ is the regression vector defined by

[2] ARX = AutoRegressive with eXogenous input.

$$\varphi(t) := [-y(t-1)\ \cdots\ -y(t-n)\ u(t-1)\ \cdots\ u(t-m)]^{\mathrm{T}} \in \mathbb{R}^{d\times 1}$$

Also, the unknown parameter vector is given by

$$\theta := (a_1 \cdots a_n\ b_1 \cdots b_m)^{\mathrm{T}}$$

Thus, it follows from (1.3) and (A.8) that

$$V_N(\theta) = \frac{1}{N}\sum_{t=0}^{N-1}[y(t) - \varphi^{\mathrm{T}}(t)\theta]^2$$

This implies that the PEM as applied to ARX models reduces to the least-squares method, so that the optimal estimate is given by

$$\hat{\theta}_{\mathrm{LS}}(N) = \left(\frac{1}{N}\sum_{t=0}^{N-1}\varphi(t)\varphi^{\mathrm{T}}(t)\right)^{-1}\frac{1}{N}\sum_{t=0}^{N-1}\varphi(t)y(t) \tag{A.9}$$

Suppose that the actual observations are expressed as

$$y(t) = \varphi^{\mathrm{T}}(t)\theta_0 + v_0(t) \tag{A.10}$$

where v_0 is a noise, and θ_0 is the "true" parameter. Substituting the above equation into (A.9) yields

$$\hat{\theta}_{\mathrm{LS}}(N) = \theta_0 + \left(\frac{1}{N}\sum_{t=0}^{N-1}\varphi(t)\varphi^{\mathrm{T}}(t)\right)^{-1}\frac{1}{N}\sum_{t=0}^{N-1}\varphi(t)v_0(t)$$

Suppose that

$$\mathrm{LS1)}\quad \lim_{N\to\infty}\frac{1}{N}\sum_{t=0}^{N-1}\varphi(t)\varphi^{\mathrm{T}}(t) = E\{\varphi(t)\varphi^{\mathrm{T}}(t)\} = \text{nonsingular}$$

$$\mathrm{LS2)}\quad \lim_{N\to\infty}\frac{1}{N}\sum_{t=0}^{N-1}\varphi(t)v_0(t) = E\{\varphi(t)v_0(t)\} = 0$$

hold[3]. Then we can show that

$$\lim_{N\to\infty}\hat{\theta}_{\mathrm{LS}}(N) = \theta_0$$

Thus the least-squares estimate is consistent. □

For convergence results based on laws of large numbers, see [109, 145]. If the above condition LS2) is not satisfied, then the least-squares estimate becomes biased. In order to obtain an unbiased estimate, we can employ a vector sequence correlated with the regressor vector $\varphi(t)$ but uncorrelated with the external noise $v_0(t)$.

[3]The second condition is surely satisfied if v_0 is a filtered white noise and $\varphi(t)$ is a bounded sequence [109].

Example A.3. (IV estimate) Let $\zeta(t) \in \mathbb{R}^d$ be a vector sequence. Pre-multiplying (A.10) by $\zeta(t)$ and summing over $[0, N-1]$ yield

$$\frac{1}{N}\sum_{t=0}^{N-1}\zeta(t)y(t) = \left(\frac{1}{N}\sum_{t=0}^{N-1}\zeta(t)\varphi(t)\right)\theta_0 + \frac{1}{N}\sum_{t=0}^{N-1}\zeta(t)v_0(t)$$

Suppose that a vector $\zeta(t)$ satisfies two conditions

$$\text{IV1)} \quad \lim_{N\to\infty}\frac{1}{N}\sum_{t=0}^{N-1}\zeta(t)\varphi^{\mathrm{T}}(t) = E\{\zeta(t)\varphi^{\mathrm{T}}(t)\} = \text{nonsingular}$$

$$\text{IV2)} \quad \lim_{N\to\infty}\frac{1}{N}\sum_{t=0}^{N-1}\zeta(t)v_0(t) = E\{\zeta(t)v_0(t)\} = 0$$

Then, we obtain a consistent estimate

$$\hat{\theta}_{\mathrm{IV}}(N) = \left(\frac{1}{N}\sum_{t=0}^{N-1}\zeta(t)\varphi^{\mathrm{T}}(t)\right)^{-1}\frac{1}{N}\sum_{t=0}^{N-1}\zeta(t)y(t) \tag{A.11}$$

This estimate is usually called an instrumental variable (IV) estimate, and the vectors $\zeta(t)$ satisfying the conditions IV1) and IV2) are called IV vectors. □

Detailed discussions on the IV estimate, including the best choice of the IV vector and convergence results, are found in [109, 145].

A.2 LQ Decomposition

We consider the relation between the least-squares method and LQ decomposition, which is a key technique in subspace identification methods.

Consider an FIR (finite impulse response) model

$$y(t) = \sum_{i=0}^{k-1} g_i u(t-i) + e(t) \tag{A.12}$$

where e is a white noise with mean zero and variance σ^2. The problem is to identify the impulse responses $\theta := (g_{k-1} \ \cdots \ g_1 \ g_0)^{\mathrm{T}}$ based on the input-output data $\{u(t),\ y(t),\ t = 0,\ 1,\ \cdots,\ N+k-2\}$. We define a data matrix

$$\left[\frac{U_{0|k-1}}{Y_{k-1|k-1}}\right] := \left[\begin{array}{ccc} u(0) & u(1) \cdots & u(N-1) \\ u(1) & u(2) \cdots & u(N) \\ \vdots & \vdots & \vdots \\ \hline u(k-1) & u(k) \cdots & u(N+k-2) \\ y(k-1) & y(k) \cdots & y(N+k-2) \end{array}\right] \in \mathbb{R}^{(k+1)\times N}$$

where we assume that $U_{0|k-1}$ has full row rank, so that $\mathrm{rank}(U_{0|k-1}) = k$.

We temporarily assume that $\sigma^2 = 0$. Then from (A.12), we get

$$[g_{k-1} \;\; g_{k-2} \;\; \cdots \;\; g_0 \;\; -1] \left[\begin{array}{ccc} u(0) & u(1) \cdots & u(N-1) \\ u(1) & u(2) \cdots & u(N) \\ \vdots & \vdots & \vdots \\ u(k-1) & u(k) \cdots & u(N+k-2) \\ \hline y(k-1) & y(k) \cdots & y(N+k-2) \end{array}\right] = 0$$

or this can be simply written as

$$[\theta^{\mathrm{T}} \;\; -1] \begin{bmatrix} U_{0|k-1} \\ Y_{k-1|k-1} \end{bmatrix} = 0 \tag{A.13}$$

As shown in Example 6.2, this problem can be solved by using the SVD of the data matrix. In fact, let $\begin{bmatrix} U_{0|k-1} \\ Y_{k-1|k-1} \end{bmatrix} = USV^{\mathrm{T}}$. Since the last singular value is zero due to (A.13), *i.e.* $\sigma_{k+1} = 0$, the $(k+1)$th left singular vector u_{k+1} satisfies

$$u_{k+1}^{\mathrm{T}} \begin{bmatrix} U_{0|k-1} \\ Y_{k-1|k-1} \end{bmatrix} = 0$$

Thus, normalizing the vector u_{k+1} so that the last element becomes -1, we obtain an estimate of the vector θ.

Now we assume that $\sigma^2 > 0$, where no θ exists satisfying (A.13), so that we must take a different route to estimate the vector θ. The LQ decomposition of the data matrix yields

$$\begin{bmatrix} U_{0|k-1} \\ Y_{k-1|k-1} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} Q_1^{\mathrm{T}} \\ Q_2^{\mathrm{T}} \end{bmatrix} \tag{A.14}$$

where $L_{11} \in \mathbb{R}^{k \times k}$, $L_{22} \in \mathbb{R}^{1 \times 1}$, $L_{21} \in \mathbb{R}^{1 \times k}$, and matrices $Q_1 \in \mathbb{R}^{N \times k}$, $Q_2 \in \mathbb{R}^{N \times 1}$ are orthogonal. By the rank condition for $U_{0|k-1}$, we see that $\det(L_{11}) \neq 0$, so that

$$Y_{k-1|k-1} = L_{21}Q_1^{\mathrm{T}} + L_{22}Q_2^{\mathrm{T}} = L_{21}L_{11}^{-1}U_{0|k} + L_{22}Q_2^{\mathrm{T}}$$

Since $Q_1^{\mathrm{T}}Q_2 = 0$, two terms in the right-hand side of the above equation are uncorrelated. Define

$$(g_{k-1} \;\; \cdots \;\; g_{k-1} \;\; g_0) := L_{21}L_{11}^{-1}$$

and

$$\left[e(k-1) \;\; e(k) \;\; \cdots \;\; e(N+k-2)\right] := L_{22}Q_2^{\mathrm{T}}$$

Then, for $t = k-1, \, k, \, \cdots, \, N+k-2$, we have

$$y(t) = g_0u(t) + g_1u(t-1) + \cdots + g_{k-1}u(t-k+1) + e(t)$$

This is the same FIR model as (A.12), implying that $\theta^{\mathrm{T}} = L_{21}L_{11}^{-1} \in \mathbb{R}^{1 \times k}$ is the least-squares estimates of impulse response parameters.

We show that the above result is also derived by solving the normal equation. The identification problem for the FIR model (A.12) can be cast into a least-squares problem

$$\min J(\theta) = \|Y_{k-1|k-1}^{\mathrm{T}} - U_{0|k-1}^{\mathrm{T}}\theta\|^2$$

Thus, from (A.3) and (A.14), the least-squares estimate is given by

$$\begin{aligned}\hat{\theta} &= (U_{0|k-1}U_{0|k-1}^{\mathrm{T}})^{-1}U_{0|k-1}Y_{k-1|k-1}^{\mathrm{T}} \\ &= (L_{11}L_{11}^{\mathrm{T}})^{-1}[L_{11}Q_1^{\mathrm{T}}(L_{21}Q_1^{\mathrm{T}} + L_{22}Q_2^{\mathrm{T}})^{\mathrm{T}}] = (L_{21}L_{11}^{-1})^{\mathrm{T}}\end{aligned}$$

This is exactly the same as the least-squares estimate of θ obtained above by using the LQ decomposition. Thus we conclude that the least-squares problem can be solved by using the LQ decomposition.

B

Input Signals for System Identification

The selection of input signals has crucial effects on identification results. In this section, several input signals used for system identification are described, including step signals, sinusoids as well as random signals. One of the most important concepts related to input signals is the persistently exciting (PE) condition.

Let $u(t), t = 0,\ 1,\ \cdots$ be a deterministic function. Then we define the mean and auto-covariance function as

$$\mu_u = \lim_{N\to\infty} \frac{1}{N} \sum_{t=0}^{N-1} u(t) \tag{B.1}$$

and for $l = 0,\ \pm 1,\ \cdots,$

$$\Lambda_{uu}(l) = \lim_{N\to\infty} \frac{1}{N} \sum_{t=0}^{N-1} [u(t+l) - \mu_u][u(t) - \mu_u] \tag{B.2}$$

Example B.1. (a) A step function is defined by

$$u(t) = \begin{cases} u_0, & t = 0,\ 1, \cdots \\ 0, & t = -1,\ -2,\ \cdots \end{cases}$$

In this case, we have $\Lambda_{uu}(l) = 0$ for $l = 0, \pm 1, \cdots$.

(b) Consider a sinusoid defined by

$$u(t) = a \sin(\omega t + \phi), \qquad t = 0,\ 1, \cdots \tag{B.3}$$

where $\omega > 0$ denotes the angular frequency, and $a > 0$ and $0 < \phi < \pi$ are the amplitude and phase, respectively. Let

$$S_N = \frac{1}{N} \sum_{t=0}^{N-1} \sin(\omega t + \phi), \qquad C_N = \frac{1}{N} \sum_{t=0}^{N-1} \cos(\omega t + \phi)$$

Since $\lim_{N\to\infty} S_N = 0$ and $\lim_{N\to\infty} C_N = 0$ hold, we have

$$\begin{aligned}\Lambda_{uu}(l) &= \lim_{N\to\infty} \frac{1}{N} \sum_{t=0}^{N-1} a^2 \sin(\omega(t+l)+\phi)\sin(\omega t+\phi) \\ &= \frac{a^2}{2}\cos(\omega l), \qquad l = 0, \pm 1, \cdots\end{aligned}$$

where the formula: $\sin\alpha \sin\beta = [\cos(\alpha-\beta) - \cos(\alpha+\beta)]/2$ is used.

Also, consider a composite sinusoid

$$u(t) = \sum_{j=1}^{p} a_j \sin(\omega_j t + \phi_j), \qquad t = 0,\ 1, \cdots \tag{B.4}$$

where $0 < \omega_1 < \cdots < \omega_p$ denote the angular frequencies, and $\{a_j\}$ and $\{\phi_j\}$ denote the amplitudes and phases, respectively. Then, it can be shown that

$$\Lambda_{uu}(l) = \sum_{j=1}^{p} \frac{a_j^2}{2}\cos(\omega_j l), \qquad l = 0, \pm 1, \cdots$$

(c) In system identification, a pseudo-random binary signal (PRBS) shown in Figure B.1 is often employed as test inputs. The PRBS is a periodic sequence with the maximum period $N = 2^p - 1$ where p is an integer greater than three, and is easily generated by p-stage shift registers. It is shown [145] that the mean and auto-covariance of a PRBS taking values on $\pm b$ are given by

$$\mu_u = \frac{1}{N}\sum_{t=1}^{N} u(t) = \frac{b}{N} \tag{B.5}$$

$$\Lambda_{uu}(l) = \begin{cases} b^2\left(1 - \dfrac{1}{N^2}\right), & l = 0 \pmod N \\ -\dfrac{b^2}{N}\left(1 + \dfrac{1}{N}\right), & l \neq 0 \pmod N \end{cases} \tag{B.6}$$

The auto-covariance function are shown in Figure B.2. □

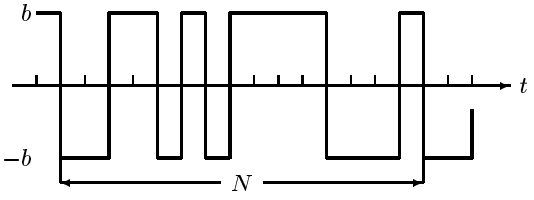

Figure B.1. A PRBS with the maximum period $N = 15$

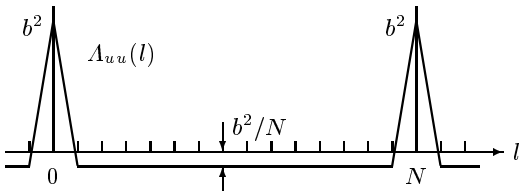

Figure B.2. The auto-covariance sequence of PRBS

In order to explain the PE condition of input signals, we consider the same FIR model as (A.12):

$$y(t) = \sum_{i=0}^{k-1} g_i u(t-i) + e(t) \tag{B.7}$$

where e is a zero mean white noise with variance σ^2. We deal with the identification of the impulse responses $\theta = (g_{k-1} \ \cdots \ g_1 \ g_0)^{\mathrm{T}}$ of the FIR model based on input-output data $\{u(t),\ y(t),\ t = 0,\ 1,\ \cdots,\ N-1\}$. For notational simplicity, we define the stacked vectors

$$y_{N-1} = \begin{bmatrix} y(k-1) \\ y(k) \\ \vdots \\ y(N-1) \end{bmatrix}, \qquad e_{N-1} = \begin{bmatrix} e(k-1) \\ e(k) \\ \vdots \\ e(N-1) \end{bmatrix} \in \mathbb{R}^{(N-k+1)\times 1}$$

and the matrix

$$U_{N-1} = \begin{bmatrix} u(0) & u(1) & \cdots & u(k-1) \\ u(1) & u(2) & \cdots & u(k) \\ \vdots & \vdots & \ddots & \vdots \\ u(N-k) & u(N-k+1) & \cdots & u(N-1) \end{bmatrix} \in \mathbb{R}^{(N-k+1)\times k}$$

Then, from (B.7), we have a linear regression model of the form

$$y_{N-1} = U_{N-1}\theta + e_{N-1} \tag{B.8}$$

The least-squares estimate of θ for (B.8) is obtained by solving

$$\min_{\theta} \|y_{N-1} - U_{N-1}\theta\|$$

Recall that Conditions A1) $\sim$ A4) in Section A.1 are required for solving the least-squares estimation problems. In particular, to get a unique solution, it is necessary to assume that $\mathrm{rank}(U_{N-1}) = k$. This condition is equivalent to the fact that

$$\mathrm{rank} \begin{bmatrix} u(0) & u(1) & \cdots & u(N-k) \\ u(1) & u(2) & \cdots & u(N-k+1) \\ \vdots & \vdots & & \vdots \\ u(k-1) & u(k) & \cdots & u(N-1) \end{bmatrix} = k \tag{B.9}$$

It may be noted that the data length is finite for this case.

Definition B.1. *[165] A deterministic sequence u with length N is PE of order k if (B.9) holds. If the input is a vector process $u \in \mathbb{R}^m$, then the rank condition (B.9) is replaced by* $\mathrm{rank}(U_{N-1}) = km$. □

For a zero mean stationary process $u \in \mathbb{R}^m$, we define the covariance matrix by

$$\begin{aligned}\bar{\Lambda}_{uu}(k) &= \lim_{N\to\infty} \frac{1}{N} U_N^{\mathrm{T}} U_N \\ &= \begin{bmatrix} \Lambda_{uu}(0) & \Lambda_{uu}^{\mathrm{T}}(1) & \cdots & \Lambda_{uu}^{\mathrm{T}}(k-1) \\ \Lambda_{uu}(1) & \Lambda_{uu}(0) & \cdots & \Lambda_{uu}^{\mathrm{T}}(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda_{uu}(k-1) & \Lambda_{uu}(k-2) & \cdots & \Lambda_{uu}(0) \end{bmatrix}\end{aligned} \tag{B.10}$$

Then the PE condition for a stationary stochastic process is defined as follows.

Definition B.2. *[109, 145] If $\bar{\Lambda}_{uu}(k)$ of (B.10) is positive definite, then we say that u has the PE condition of order k.* □

The following example shows that when we deal with finite data, there always exist some ambiguities regarding how we treat boundary data.

Example B.2. Consider the step function treated in Example B.1. It can be shown that the step function is not PE since we have $\mathrm{rank}\,(U_{N-1}) = 1$.

However, in practice, step signals are often used for system identification. To consider this problem, we express (B.7) as

$$\begin{aligned} y(0) &= g_0 u(0) + g_1 u(-1) + \cdots + g_{k-1} u(-k+1) + e(0) \\ y(1) &= g_0 u(1) + g_1 u(0) + \cdots + g_{k-1} u(-k+2) + e(1) \\ &\vdots \\ y(k-1) &= g_0 u(k-1) + g_1 u(k-2) + \cdots + g_{k-1} u(0) + e(k-1) \\ y(k) &= g_0 u(k) + g_1 u(k-1) + \cdots + g_{k-1} u(1) + e(k) \\ &\vdots \end{aligned}$$

Suppose that the system is at rest for $t < 0$. Then we have $u(t) = 0$, $t = -1, -2, \cdots, -k+1$. Rearranging the above equations and assuming that $e(t) = 0$ for $t = 0, 1, \cdots$, we get

$$\begin{aligned}&[y(0)\ y(1)\ \cdots\ y(N-1)] \\ &= [g_{k-1}\ \cdots\ g_0] \begin{bmatrix} 0 & \cdots & 0 & u(0) & \cdots & u(N-k) \\ \vdots & \iddots & \iddots & u(1) & \cdots & u(N-k+1) \\ 0 & \iddots & \iddots & \vdots & & \vdots \\ u(0) & u(1) & \cdots & u(k-1) & \cdots & u(N-1) \end{bmatrix}\end{aligned}$$

Thus if $u(t) = 1,\ t = 0,\ 1, \cdots$, the wide rectangular matrix in the right-hand side of the above equation has rank k. Hence, by using the least-squares method, we can identify the impulse responses $g_0,\ g_1,\ \cdots, g_{k-1}$. However, in this case, it should be understood that the estimate is obtained by using the additional information that $u(t) = 0,\ t = -1,\ -2,\ \cdots,\ -k+1$. □

Example B.3. We consider the order of PE condition for simple signals based on Definition B.2.

(a) Let $u(t)$ be a zero mean white noise with variance σ^2. Then, for all $k > 0$, we see that $\bar{\Lambda}_{uu}(k) = \sigma^2 I_k$ is positive definite. Thus the white noise satisfies the PE condition of order infinity.

(b) Consider a sinusoid $u(t) = A\sin(\lambda_0 t),\ 0 < \lambda_0 < \pi$. Then, the auto-covariance function is given by $\Lambda_{uu}(k) = (A^2/2)\cos(\lambda_0 k)$, so that

$$\bar{\Lambda}_{uu}(2) = \frac{A^2}{2}\begin{bmatrix} 1 & \cos\lambda_0 \\ \cos\lambda_0 & 1 \end{bmatrix}$$

$$\bar{\Lambda}_{uu}(3) = \frac{A^2}{2}\begin{bmatrix} 1 & \cos\lambda_0 & \cos 2\lambda_0 \\ \cos\lambda_0 & 1 & \cos\lambda_0 \\ \cos 2\lambda_0 & \cos\lambda_0 & 1 \end{bmatrix}$$

We see that $\mathrm{rank}[\bar{\Lambda}_{uu}(2)] = 2$, and $\mathrm{rank}[\bar{\Lambda}_{uu}(k)] = 2$ for $k = 3,\ 4,\ \cdots$. Hence the sinusoid has PE condition of order two. This is obvious because a sinusoid has two independent parameters, a magnitude and a phase shift. □

Lemma B.1. *The PE conditions for some familiar stochastic processes are provided.*

(i) ARMA processes have the PE condition of order infinity.

(ii) The composite sinusoid of (B.4) satisfies the PE condition of order $2p$.

Proof. [145] (i) Let u be a zero mean ARMA process with the spectral density function $\Phi_{uu}(\omega)$. Define $h = (h(0),\ h(1),\ \cdots,\ h(l-1))^{\mathrm{T}}$, and

$$H(z) = \sum_{i=0}^{l-1} h(i) z^{-i}$$

Consider a process defined by $y = H(z)u$. Then we easily see that y is a zero mean second-order stationary process, so that the variance of y is given by

$$\sigma_y^2 = E\left\{ \left| \sum_{i=0}^{l-1} h(i)u(t-i) \right|^2 \right\} = \sum_{i,j=0}^{l-1} \Lambda_{uu}(i-j)h(i)h(j) = h^{\mathrm{T}} \bar{\Lambda}_{uu}(l) h$$

It follows from Lemma 4.4 that

$$h^{\mathrm{T}} \bar{\Lambda}_{uu}(l) h = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 \Phi_{uu}(\omega) d\omega \tag{B.11}$$

Suppose that u does not satisfy the PE condition of order l. Then there exists a nonzero vector $h \in \mathbb{R}^l$ such that $h^{\mathrm{T}} \bar{\Lambda}_{uu}(l) h = 0$. Since the integrand of (B.11) is nonnegative, we have $|H(e^{j\omega})|^2 \Phi_{uu}(\omega) = 0$ for all ω[1]. However, from (4.35), the spectral density function of the ARMA process is positive except for at most finite points. It therefore follows that $H(e^{j\omega}) = 0$ (a.e.), and hence $h = 0$. This is a contradiction, implying that the ARMA process satisfies the PE condition of order l. Since l is arbitrary, the ARMA process satisfies the PE condition of infinite order.

(ii) Since, as shown in Example B.3, a sinusoid has the PE condition of order two, the composite sinusoid of (B.4) has the PE condition of order $2p$. □

From Lemma B.1 (i), we can say that for a stationary process u, if

$$\Phi_{uu}(\omega) > 0, \qquad -\pi < \omega < \pi$$

is satisfied, then u is PE of order infinity. This condition has already been mentioned in Chapters 9 and 10.

[1]The equality holds for $\omega \in (-\pi, \pi)$ almost everywhere (a.e.).

C

Overlapping Parametrization

In this section, we derive an overlapping parametrization for a stationary process; see also Example 1.2. From Theorems 4.3 and 4.4 (see Section 4.5), a zero mean regular full rank process $y \in \mathbb{R}^p$ can uniquely be expressed as

$$y(t) = \sum_{i=0}^{\infty} H_i\, e(t-i) = \sum_{i=-\infty}^{t} H_{t-i}\, e(i) \tag{C.1}$$

where e is the innovation process with mean 0 and covariance matrix $R > 0$, and where H_i, $i = 0, 1, \cdots$ are impulse response matrices satisfying

$$\sum_{i=0}^{\infty} \|H_i\|^2 < \infty; \qquad H_0 = I_p$$

Define the transfer matrix by

$$H(z) = \sum_{i=0}^{\infty} H_i z^{-i}$$

Moreover, define

$$\mathcal{Y}_t^- = \overline{\text{span}}\{y(t-1),\, y(t-2),\, \cdots\}$$
$$\mathcal{E}_t^- = \overline{\text{span}}\{e(t-1),\, e(t-2),\, \cdots\}$$

Then, it follows that $\mathcal{Y}_t^- = \mathcal{E}_t^-$, $t = 0, \pm 1, \cdots$. In the following, we assume that both $H(z)$ and $H^{-1}(z)$ are stable.

Let t be the present time. Then, from (C.1),

$$y(t+k) = \sum_{i=t}^{t+k} H_{t+k-i}\, e(i) + \sum_{i=-\infty}^{t-1} H_{t+k-i}\, e(i), \quad k = 0, 1, \cdots \tag{C.2}$$

Thus we see that the first term in the right-hand side of the above equation is a linear combination of the future innovations $e(t), \cdots, e(t+k)$ and that the second

term is a linear combination of the past innovations $e(t-1)$, $e(t-2)$, $\cdots$. Since $\mathcal{Y}_t^- = \mathcal{E}_t^-$, the second term is also expressed as a linear combination of the past outputs $y(t-1)$, $y(t-2)$, $\cdots$, and hence it belongs to $\mathcal{Y}_t^-$. Thus it follows that the optimal predictor for $y(t+k)$ based on the past $\mathcal{Y}_t^-$ is given by (see Example 4.10)

$$\hat{y}(t+k \mid t-1) = \sum_{i=-\infty}^{t-1} H_{t+k-i}\, e(i), \qquad k = 0,\, 1,\, \cdots \tag{C.3}$$

Repeated use of this relation yields

$$\begin{bmatrix} \hat{y}(t \mid t-1) \\ \hat{y}(t+1 \mid t-1) \\ \hat{y}(t+2 \mid t-1) \\ \vdots \end{bmatrix} = \begin{bmatrix} H_1 & H_2 & H_3 & \cdots \\ H_2 & H_3 & H_4 & \cdots \\ H_3 & H_4 & H_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} e(t-1) \\ e(t-2) \\ e(t-3) \\ \vdots \end{bmatrix} \tag{C.4}$$

It should be noted that this is a free response of the system with the initial state resulting from the past inputs e up to time $t-1$ (see also Section 6.2).

Let the block Hankel operator be

$$H = \begin{bmatrix} H_1 & H_2 & H_3 & \cdots \\ H_2 & H_3 & H_4 & \cdots \\ H_3 & H_4 & H_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

where it is assumed that $\operatorname{rank}(H) = n < \infty$. As shown in Section 8.3, the predictor space is defined by

$$\mathcal{X}_t^{+/-} := \hat{E}\{\mathcal{Y}_t^+ \mid \mathcal{Y}_t^-\} = \overline{\operatorname{span}}\{\hat{y}(t+k \mid t-1) \mid k = 0,\, 1,\, \cdots\}$$

Thus, we can find n independent vectors from the infinite components

$$\{\hat{y}_i(t+k \mid t-1), \quad i = 1, \cdots, p, \quad k = 0,\, 1, \cdots\} \tag{C.5}$$

where the n independent vectors form a basis of the predictor space $\mathcal{X}_t^{+/-}$.

Suppose that $\hat{y}(t \mid t-1)$ has full rank, *i.e.* $\operatorname{cov}\{\hat{y}(t \mid t-1)\} > 0$. Then,

$$\hat{y}_1(t \mid t-1),\ \hat{y}_2(t \mid t-1),\ \cdots,\ \hat{y}_p(t \mid t-1) \tag{C.6}$$

are linearly independent, and hence we see that the first p rows of H are linearly independent.

Let $\overline{n} = (n_1, \cdots, n_p)$ be a set of p positive integers such that $n_1 + \cdots + n_p = n$. We pick n elements including the p components of (C.6) from the infinite components defined by (C.5). Let such vectors be given by

$$\begin{matrix}
\hat{y}_1(t \mid t-1),\ \hat{y}_1(t+1 \mid t-1),\ \cdots,\ \hat{y}_1(t+n_1-1 \mid t-1) \\
\hat{y}_2(t \mid t-1),\ \hat{y}_2(t+1 \mid t-1),\ \cdots,\ \hat{y}_2(t+n_2-1 \mid t-1) \\
\vdots \qquad\qquad\qquad\qquad \vdots \\
\hat{y}_p(t \mid t-1),\ \hat{y}_p(t+1 \mid t-1),\ \cdots,\ \hat{y}_p(t+n_p-1 \mid t-1)
\end{matrix}$$

Note that, for example, if $n_1 = 1$, then only $\hat{y}_1(t \mid t-1)$ is selected from the first row. If these n vectors are linearly independent, we call $\overline{n} = (n_1, \cdots, n_p)$ a multi-index; see [54, 68, 109] for more details.

By using the above linearly independent components, we define a state vector of the system by

$$x(t) := \begin{bmatrix}
\hat{y}_1(t \mid t-1) \\
\vdots \\
\hat{y}_1(t+n_1-1 \mid t-1) \\
\vdots \\
\hat{y}_p(t \mid t-1) \\
\vdots \\
\hat{y}_p(t+n_p-1 \mid t-1)
\end{bmatrix} \in \mathbb{R}^n \tag{C.7}$$

From (C.3), we get

$$\begin{aligned}
\hat{y}(t+k \mid t) &= \sum_{i=-\infty}^{t} H_{t+k-i}\, e(i) = \sum_{i=-\infty}^{t-1} H_{t+k-i}\, e(i) + H_k\, e(t) \\
&= \hat{y}(t+k \mid t-1) + H_k\, e(t), \qquad k = 0,\, 1,\, \cdots
\end{aligned}$$

In terms of the components, this can be written as

$$\hat{y}_i(t+k \mid t) = \hat{y}_i(t+k \mid t-1) + h_{ik} e(t), \qquad k = 0,\, 1,\, \cdots \tag{C.8}$$

where $i = 1, \cdots, p$, and $h_{ik} = [h_{ik}(1) \ \cdots \ h_{ik}(p)] \in \mathbb{R}^{1\times p}$ is the ith row of H_k. Also, from (C.7), the state vector at $t+1$ is expressed as

$$x(t+1) = \begin{bmatrix}
\hat{y}_1(t+1 \mid t) \\
\vdots \\
\hat{y}_1(t+n_1 \mid t) \\
\vdots \\
\hat{y}_p(t+1 \mid t) \\
\vdots \\
\hat{y}_p(t+n_p \mid t)
\end{bmatrix} \in \mathbb{R}^n$$

Thus from (C.8), we can verify that

$$x(t+1) = \begin{bmatrix} \hat{y}_1(t+1 \mid t-1) \\ \vdots \\ \hat{y}_1(t+n_1 \mid t-1) \\ \vdots \\ \hat{y}_p(t+1 \mid t-1) \\ \vdots \\ \hat{y}_p(t+n_p \mid t-1) \end{bmatrix} + \begin{bmatrix} h_{11}(1) & \cdots & h_{11}(p) \\ \vdots & & \vdots \\ h_{1n_1}(1) & \cdots & h_{1n_1}(p) \\ \vdots & & \vdots \\ h_{p1}(1) & \cdots & h_{p1}(p) \\ \vdots & & \vdots \\ h_{pn_p}(1) & \cdots & h_{pn_p}(p) \end{bmatrix} e(t) \tag{C.9}$$

Note that the first term in the right-hand side of (C.9) belongs to the space $\mathcal{X}_t^{+/-}$. In particular, we see that $\hat{y}_i(t+n_i \mid t-1), i = 1, \cdots, p$ are expressed in terms of a linear combination of the components of the basis vector $x(t)$. Thus, we have

$$\hat{y}_i(t+n_i \mid t-1) = \sum_{j=1}^{p} \sum_{k=1}^{n_j} \alpha_{ij}^k \hat{y}_j(t+k-1 \mid t-1), \quad i = 1, \cdots, p \tag{C.10}$$

Other components $\hat{y}_i(t+l \mid t-1), l = 1, \cdots, n_i - 1$ are already contained in the vector $x(t)$ as its elements, so that they are expressed in terms of shift operations. Moreover, putting $k = 0$ in (C.8) and noting that $\hat{y}(t \mid t) = y(t)$ and $H_0 = I_p$ yield

$$y_i(t) = \hat{y}_i(t \mid t-1) + e_i(t), \qquad i = 1, \cdots, p \tag{C.11}$$

where $\hat{y}_i(t \mid t-1)$ belongs to $\mathcal{X}_t^{+/-}$.

For simplicity, we consider a 3-dimensional process y with 9-dimensional state vector, and assume that $n_1 = 3, n_2 = 4, n_3 = 2$ with $n_1 + n_2 + n_3 = 9$. Then, by using (C.8) and (C.11), we have the following A- and C-matrix:

$$A = \left[\begin{array}{ccc|cccc|cc} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{11}^1 & \alpha_{11}^2 & \alpha_{11}^3 & \alpha_{12}^1 & \alpha_{12}^2 & \alpha_{12}^3 & \alpha_{12}^4 & \alpha_{13}^1 & \alpha_{13}^2 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \alpha_{21}^1 & \alpha_{21}^2 & \alpha_{21}^3 & \alpha_{22}^1 & \alpha_{22}^2 & \alpha_{22}^3 & \alpha_{22}^4 & \alpha_{23}^1 & \alpha_{23}^2 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \alpha_{31}^1 & \alpha_{31}^2 & \alpha_{31}^3 & \alpha_{32}^1 & \alpha_{32}^2 & \alpha_{32}^3 & \alpha_{32}^4 & \alpha_{33}^1 & \alpha_{33}^2 \end{array}\right]$$

$$(p = 3, n = 9, n_1 = 3, n_2 = 4, n_3 = 2)$$

and

$$C = \left[\begin{array}{ccc|cccc|cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right]$$

We can easily infer the forms of A- and C-matrix for general cases. Thus, we have the following state space equation

$$x(t+1) = Ax(t) + Ke(t) \tag{C.12a}$$

$$y(t) = Cx(t) + e(t) \tag{C.12b}$$

where $K \in \mathbb{R}^{n\times p}$ is the coefficient matrix for $e(t)$ of (C.9). We see that the number of unknown parameters in this Markov model is $2np$, since K has no particular structure.

From the property of block Hankel matrix, we have the following lemma.

Lemma C.1. *[54, 109] Any n-dimensional stochastic LTI state space system can be expressed by means of a state space model (C.12) with a particular multi-index $\overline{n}$. In other words, the state space model (C.12) with a particular multi-index $\overline{n}$ can describe almost all n-dimensional stochastic LTI systems.* □

More precisely, let $M_{\overline{n}}(p)$ be the model structure of (C.12) with a multi-index $\overline{n}$. Also, let the sum of $M_{\overline{n}}(p)$ over possible multi-indices be

$$\overline{\mathcal{M}}(p) = \bigcup_{\overline{n}} \operatorname{Im} M_{\overline{n}}(p)$$

Then, the set $\overline{\mathcal{M}}(p)$ denotes the set of all n-dimensional linear stochastic system with p outputs. Of course, $M_{\overline{n}}(p)$ may overlap, but $\overline{\mathcal{M}}(p)$ contains all the n-dimensional linear systems $M_{\overline{n}}(p)$.

The state space model of (C.12) is called an overlapping parametrization with $2np$ independent parameters. Thus, we can employ the PEM to identify the $2np$ unknown parameters, but we need some complicated algorithms for switching from a particular $\bar{n}^1$ to another $\bar{n}^2$ during the parameter identification, since we do not know the multi-index $\bar{n}$ prior to identification.

In general, a p-dimensional process y with state dimension n is called generic if the state vector x is formed as in (C.7) using some multi-index $\bar{n} = (n_1, \cdots, n_p)$. The next example shows that there exist non-generic processes.

Example C.1. [54] Let $p = 2$ and $n = 3$, and consider the following matrices:

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \qquad A = \begin{bmatrix} \alpha & \beta & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \qquad K = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

where $\alpha\beta \neq 0$. Then, since $H_j = CA^{j-1}K$, $j = 1, \cdots$, we get

$$H_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad H_2 = \begin{bmatrix} \beta & 0 \\ 0 & 1 \end{bmatrix}, \quad H_3 = \begin{bmatrix} \alpha\beta & \beta \\ 1 & 0 \end{bmatrix}, \quad H_4 = \begin{bmatrix} \alpha^2\beta + \beta & \alpha\beta \\ 0 & 1 \end{bmatrix}, \cdots$$

Thus the first 3×2 block submatrix of H is given by

$$H_{3,2} := \begin{bmatrix} H_1 & H_2 \\ H_2 & H_3 \\ H_3 & H_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \beta & 0 \\ 1 & 0 & 0 & 1 \\ \beta & 0 & \alpha\beta & \beta \\ 0 & 1 & 1 & 0 \\ \alpha\beta & \beta & \alpha^2\beta + \beta & \alpha\beta \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

It is easy to see that the first two rows of the block Hankel matrix are linearly independent, but the 3rd row is linearly dependent on the first two rows. Thus we observe that the selection $\bar{n} = (2, 1)$ $(n_1 = 2, n_2 = 1)$ does not yield a basis. Actually, in this case, we should pick the first two rows and the fourth row to form a basis. □

D

List of Programs

In Appendix D, some of MATLAB® programs used in this book are included.

D.1 Deterministic Realization Algorithm

Table D.1 displays a program for the Ho-Kalman's algorithm of Lemma 6.1, where it is assumed that k, $l > n := \text{rank}(H)$.

Table D.1. Ho-Kalman's algorithm

```
% Function  zeiger.m
% Lemma 6.1
function[A,B,C] = zeiger(H,p,m,n)
% p = dim(y); m = dim(u); n = dim(x)
% (p, m) are known
% kp × lm Hankel matrix
% k, l > n; H must be finite rank
kp = size(H,1); lm = size(H',1);
[U,S,V] = svd(H);                                   % Eq. (6.14)
n=rank(S); % if n is known, this is redundant.
S1 = sqrtm(S(1:n,1:n));
% T = identity matrix                               % Eq. (6.15)
Ok = U(:,1:n)*S1;
Cl = S1*V(:,1:n)';
A = Ok(1:kp-p,:)\Ok(p+1:kp,:);                      % Eq. (6.16)
B = Cl(:,1:m);
C = Ok(1:p,:);
```

Table D.2. MOESP method

```
% Function moeps.m
% Lemma 6.6
% m = dim(u),  p = dim(y),  n = dim(x);  k = number of block rows
% U = km x N input data matrix
% Y = kp x N output data matrix
function [A,B,C,D] = moesp(U,Y,m,p,n,k)
km = size(U,1);  kp = size(Y,1);
L = triu(qr([U;Y]'))';  % LQ decomposition
L11 = L(1:km,1:km);
L21 = L(km+1:km+kp,1:km);
L22 = L(km+1:km+kp,km+1:km+kp);
[UU,SS,VV] = svd(L22);                                  % Eq. (6.39)
U1 = UU(:,1:n); % n is known
Ok = U1*sqrtm(SS(1:n,1:n));
% Matrices A and C
C = Ok(1:p,1:n);                                        % Eq. (6.41)
A = pinv(Ok(1:p*(k-1),1:n))*Ok(p+1:p*k,1:n);            % Eq. (6.42)
% Matrices B and D
U2 = UU(:,n+1:size(UU',1));
Z = U2'*L21/L11;
XX = []; RR = [];
for j = 1:k
XX = [XX; Z(:,m*(j-1)+1:m*j)];
   Okj = Ok(1:p*(k-j),:);
   Rj = [zeros(p*(j-1),p)  zeros(p*(j-1),n);
      eye(p)  zeros(p,n); zeros(p*(k-j),p)  Okj];
RR = [RR; U2'*Rj];
end
DB = pinv(RR)*XX;                                       % Eq. (6.44)
D = DB(1:p,:);
B = DB(p+1:size(DB,1),:);
```

D.2 MOESP Algorithm

Table D.2 displays a program for the basic MOESP method developed in [172, 173]. A formation of data matrices is omitted in this program, but Table D.3 contains a related method of constructing data matrices.

It should be noted that way of computing matrices A and C is different in each method, but the computing method of B and D in the MOESP method in Table D.2 is commonly used in many other subspace identification methods (not always). Thus we can say that differences in algorithms of subspace system identification methods are attributed to the way of computing A and C, or the image of extended observability matrix $\mathrm{Im}(\mathcal{O}_k)$.

D.3 Stochastic Realization Algorithms

We show two algorithms of stochastic realization based on Lemma 7.9 in Section 7.7 and Algorithm A in Section 8.7. It will be instructive to understand the difference between the two stochastic realization algorithms.

Table D.3. Stochastic realization algorithm

```
% Function   stochastic.m
% Lemma 7.9
% function [A,C,Cb,K,R] = stochastic(y,n,k)
% y = [y(1),y(2),...,y(Ndat)];  p × Ndat   matrix
% n = dim(x);  k = number of block rows
function [A,C,Cb,K,R] = stochastic(y,n,k)
[p,Ndat] = size(y);  N = Ndat-2*k;
ii = 0;
for i = 1:p:2*k*p-p+1
   ii = ii+1;
   Y(i:i+p-1,:) = y(:,ii:ii+N-1);
end;
% Data matrix
Ypp = Y(1:k*p,:);
for i = 1:k
   j = (k-i)*p+1;
   Yp(j:j+p-1,:) = Ypp((i-1)*p+1:i*p,:); % Yp := Y check
end
Yf = Y(k*p+1:2*k*p,:);
Rfp = (Yf*Yp')/N;  % Covariance matrix
[U,S,V] = svd(Rfp);                                  % Eq. (7.81)
S2 = sqrtm(S(1:n,1:n));
Ok = U(:,1:n)*S2;                                    % Eq. (7.82)
Ck = S2*V(:,1:n)';
A = Ok(1:k*p-p,:)\ Ok(p+1:k*p,:);                    % Eq. (7.83)
C = Ok(1:p,:);
Cb = Ck(1:n,1:p)';
RR = (Yf*Yf')/N;
R0 = RR(1:p,1:p);  % Variance of output
[P,L,G,Rept] = dare(A',C',zeros(n,n),-R0,-Cb');      % ARE (7.84)
K = G'
R = R-C*P*C';
```

Table D.3 displays the stochastic realization algorithm of Lemma 7.9, in which ARE is solved by using the function **dare**. This function **dare** can solve the ARE appearing in stochastic realization as well as the one appearing in Kalman filtering. For details, see the manual of the function **dare**.

Table D.4. Balanced stochastic realization – Algorithm A

```
% Function  stocha_bal.m
% Algorithm A in Section 8.7
% y = [y(1),y(2),...,y(Ndat)];  p × Ndat matrix
% n = dim(x);  k = number of block rows
function [A,C,Cb,K,R] = stocha_bal (y,n,k)
[p,Ndat] = size(y);   N = Ndat-2*k;
ii = 0;
for i = 1:p:2*k*p-p+1
   ii = ii+1;     Y(i:i+p-1,:) = y(:,ii:ii+N-1);
end
Yp = Y(1:k*p,:);  Yf = Y(k*p+1:2*k*p,:);
% LQ decomposition
H = [Yp; Yf];  [Q,L] = qr(H',0);  L = L'/sqrt(N);                % Eq. (8.76)
L11 = L(1:k*p,1:k*p);  L21 = L(k*p+1:2*k*p,1:k*p);
L22 = L(k*p+1:2*k*p,k*p+1:2*k*p);
% Covariance matrices
Rff = (L21*L21'+L22*L22');
Rfp = L21*L11';   Rpp = L11*L11';
% Square roots & inverses
[Uf,Sf,Vf] = svd(Rff);  [Up,Sp,Vp] = svd(Rpp);
Sf = sqrtm(Sf);  Sp = sqrtm(Sp);
L = Uf*Sf*Vf';  M = Up*Sp*Vp';                                   % Eq. (8.77)
Sfi = inv(Sf);  Spi = inv(Sp);
Linv = Vf*Sfi*Uf';  Minv = Vp*Spi*Up';
OC = Linv*Rfp*Minv';
[UU,SS,VV] = svd(OC);                                            % Eq. (8.78)
Lambda = Rpp(1:p,1:p); % Covariance matrix of output
S = SS(1:n,1:n);
Ok = L*UU(:,1:n)*sqrtm(S);                                       % Eq. (8.79)
Ck = sqrtm(S)*VV(:,1:n)'*M';
A = Ok(1:k*p-p,:)\Ok(p+1:k*p,:);                                 % Eq. (8.80)
C = Ok(1:p,:);     Cb = Ck(:,(k-1)*p+1:k*p)';
R = Lambda-C*S*C';  K = (Cb'-A*S*C')/R;                          % Eq. (8.81)
```

Table D.4 shows a program for Algorithm A of Section 8.7. The form of data matrix Y_p in Table D.4 is slightly different from Y_p in Table D.3, since in Table D.3, after generating Y_p, we formed $\check{Y}_p$ by re-ordering the elements. Thus a way of computing $\bar{C}^{\mathrm{T}}$ in Table D.4 is different from that in Table D.3. There is no theoretical difference, but numerical results may be slightly different.

The program of Table D.4 is very simple since the solution of ARE is not employed, but there are possibilities that $A - BK$ is unstable. Also, it should be noted that we compute L^{-1} and M^{-1} by using pseudo-inverses. For, if the function **chol** is used for computing the matrix square roots, the program stops unless Σ_{ff} and Σ_{pp} are positive definite, but these matrices may be rank deficient.

D.4 Subspace Identification Algorithms

The programs for the ORT and CCA methods derived in Sections 9.7 and 10.6 are displayed in Tables D.5 and D.6, respectively. Also, a program of the PO-MOESP is included in Table D.7. Comparing the programs in Tables D.5 and D.7, we can easily understand the difference in algorithms of the ORT and PO-MOESP; both use the same LQ decomposition, but the way of utilizing L factors is different. For identifying B and D, the ORT uses the same method as the PO-MOESP.

Table D.5. Subspace identification of deterministic subsystem – ORT

```
% Function  ort_pk.m
% Subsection 9.7.1
function [A,B,C,D] = ort_pk(U,Y,m,p,n,k);
% ORT method by Picci and Katayama
km = size(U,1)/2;  kp = size(Y,1)/2;
% LQ decomposition                                        % Eq. (9.48)
L = triu(qr([U;Y]'))';
L11 = L(1:km,1:km);
L41 = L(2*km+kp+1:2*km+2*kp,1:km);
L42 = L(2*km+kp+1:2*km+2*kp,km+1:2*km);
% SVD                                                     % Eq. (9.52)
[UU,SS,VV] = svd(L42);
U1 = UU(:,1:n);
Ok = U1*sqrtm(SS(1:n,1:n));
C = Ok(1:p,1:n);
A = pinv(Ok(1:p*(k-1),1:n))*Ok(p+1:k*p,1:n);            % Eq. (9.53)
% Matrices B and D
U2 = UU(:,n+1:size(UU',1));
Z = U2'*L41/L11;                                        % Eq. (9.54)
% The program for computing B and D is the same
% as that of MOESP of Table D.2.
XX = [];
RR = [];
for j = 1:k
XX = [XX; Z(:,m*(j-1)+1:m*j)];
Okj = Ok(1:p*(k-j),:);
Rj = [zeros(p*(j-1),p),zeros(p*(j-1),n);
eye(p), zeros(p,n);
zeros(p*(k-j),p),Okj];
RR = [RR;U2'*Rj];
end
DB = pinv(RR)*XX;
D = DB(1:p,:);
B = DB(p+1:size(DB,1),:);
```

Table D.6. Stochastic subspace identification – CCA

```
% Function  cca.m
% Section 10.6 CCA  Algorithm B
% y = [y(1),y(2),...,y(Ndat)];  p×Ndat  matrix
% u = [u(1),u(2),...,u(Ndat)];  m×Ndat  matrix
% n = dim(x);  k = number of block rows
% Written by H. Kawauchi; modified by T. Katayama
function [A,B,C,D,K] = cca(y,u,n,k)
[p,Ndat] = size(y);  [m,Ndat] = size(u);  N = Ndat-2*k;
ii = 0;
for i = 1:m:2*k*m-m+1
   ii = ii+1;    U(i:i+m-1,:) = u(:,ii:ii+N-1);  % Data matrix
end
ii = 0;
for i = 1:p:2*k*p-p+1
   ii = ii+1;
   Y(i:i+p-1,:) = y(:,ii:ii+N-1);  % Data matrix
end
Uf = U(k*m+1:2*k*m,:);  Yf = Y(k*p+1:2*k*p,:);
Up = U(1:k*m,:);  Yp = Y(1:k*p,:);  Wp = [Up; Yp];
H = [Uf; Up; Yp; Yf];
[Q,L] = qr(H',0);  L = L';  % LQ decomposition
L22 = L(k*m+1:k*(2*m+p),k*m+1:k*(2*m+p));
L32 = L(k*(2*m+p)+1:2*k*(m+p),k*m+1:k*(2*m+p));
L33 = L(k*(2*m+p)+1:2*k*(m+p),k*(2*m+p)+1:2*k*(m+p));
Rff = L32*L32'+L33*L33';  Rpp = L22*L22';  Rfp = L32*L22';
[Uf,Sf,Vf] = svd(Rff);  [Up,Sp,Vp] = svd(Rpp);
Sf = sqrtm(Sf);  Sfi = inv(Sf);  Sp = sqrtm(Sp);  Spi = inv(Sp);
Lfi = Vf*Sfi*Uf';  Lpi = Vp*Spi*Up';   % Lf = Uf*Sf*Vf';  Lp = Up*Sp*Vp'
OC = Lfi*Rfp*Lpi';
[UU,SS,VV] = svd(OC);  % Normalized SVD
S1 = SS(1:n,1:n);  U1 = UU(:,1:n);  V1 = VV(:,1:n);
X = sqrtm(S1)*V1'*Lpi*Wp;  XX = X(:,2:N);  X = X(:,1:N-1);
U = Uf(1:m,1:N-1);  Y = Yf(1:p,1:N-1);
ABCD = [XX;Y]/[X;U];  % System matrices
A = ABCD(1:n,1:n);  B = ABCD(1:n,n+1:n+m);
C = ABCD(n+1:n+p,1:n);  D = ABCD(n+1:n+p,n+1:n+m);
W = XX-A*X-B*U;    E = Y-C*X-D*U;
SigWE = [W;E]*[W;E]'/(N-1);
QQ = SigWE(1:n,1:n);  RR = SigWE(n+1:n+p,n+1:n+p);
SS = SigWE(1:n,n+1:n+p);
[P,L,G,Rept] = dare(A',C',QQ,RR,SS);  % Kalman filter ARE
K = G';  % Kalman gain
```

The CCA method – Algorithm B – in Table D.6 is based on the use of estimates of state vectors. It may be noted that the LQ decomposition in the above table is

different from the one defined by (10.46); in fact, in the above program, the past input-output data $\begin{bmatrix} U_p \\ Y_p \end{bmatrix}$ is employed for $\check{W}_{0|k-1}$, since the row spaces of both data matrices are the same.

The following table shows a program of the PO-MOESP algorithm [171].

Table D.7. PO-MOESP algorithm

```
% Function  po_moesp.m
function [A,B,C,D] = po_moesp(U,Y,m,p,n,k);
% cf. Remark 9.3
% m=dim(u),  p=dim(y),  n=dim(x)
% k=number of block rows;  U=2km x N matrix;  Y=2kp x N matrix
km=k*m;
kp=k*p;
% LQ decomposition
L = triu(qr([U;Y]'))';
L11 = L(1:km,1:km);
L21 = L(km+1:2*km,1:km);
L22 = L(km+1:2*km,km+1:2*km);
L31 = L(2*km+1:2*km+kp,1:km);
L32 = L(2*km+1:2*km+kp,km+1:2*km);
L41 = L(2*km+kp+1:2*km+2*kp,1:km);
L42 = L(2*km+kp+1:2*km+2*kp,km+1:2*km);
L43 = L(2*km+kp+1:2*km+2*kp,2*km+1:2*km+kp);
[UU,SS,VV]=svd([L42 L43]);
U1 = UU(:,1:n);
Ok = U1*sqrtm(SS(1:n,1:n));
C = Ok(1:p,1:n);
A = pinv(Ok(1:p*(k-1),1:n))*Ok(p+1:k*p,1:n);
% Matrices B and D
U2 = UU(:,n+1:size(UU',1));
Z = U2'*[L31 L32 L41]/[L21 L22 L11];
% The rest is the same as that of MOESP of Table D.2.
% The subsequent part is omitted.
```

E

Solutions to Problems

Chapter 2

2.1 (a) Suppose that $\mathrm{rank}(A) = r$. Let $A = U\Sigma V^{\mathrm{T}}$, where $\Sigma = \mathrm{diag}(\Sigma_r, 0)$, and $\Sigma_r \in \mathbb{R}^{r\times r} > 0$. Also, partition $U = [U_r \;\; \tilde{U}_r]$ and $V = [V_r \;\; \tilde{V}_r]$. From Lemma 2.9 (i), we see that $\mathrm{Im}(A) = \mathrm{Im}(U_r)$, $\mathrm{Ker}(A^{\mathrm{T}}) = \mathrm{Im}(\tilde{U}_r)$, and $\mathrm{Im}(A^{\mathrm{T}}) = \mathrm{Im}(V_r)$, $\mathrm{Ker}(A) = \mathrm{Im}(\tilde{V}_r)$. Item (a) is proved by using

$$\mathrm{Im}(U_r) \oplus \mathrm{Im}(\tilde{U}_r) = \mathbb{R}^m, \qquad \mathrm{Im}(U_r) \perp \mathrm{Im}(\tilde{U}_r)$$

$$\mathrm{Im}(V_r) \oplus \mathrm{Im}(\tilde{V}_r) = \mathbb{R}^n, \qquad \mathrm{Im}(V_r) \perp \mathrm{Im}(\tilde{V}_r)$$

(b) These are the restatement of the relations in (a).

(c) We can prove the first relation of (c) as

$$\mathrm{Im}(AA^{\mathrm{T}}) = \mathrm{Im}(U_r\Sigma_r^2 U_r^{\mathrm{T}}) = \mathrm{Im}(U_r) = \mathrm{Im}(A)$$

Also, the second relation is proved as follows:

$$A\mathrm{Im}(B) = \{Ax \mid x = B\eta,\ \eta \in \mathbb{R}^p\} = \{AB\eta \mid \eta \in \mathbb{R}^p\} = \mathrm{Im}(AB)$$

2.2 Compute the product of three matrices in the right-hand side.

2.3 (a) It suffices to compute the determinant of both sides of the relations in Problem 2.2. (b) This is obvious from (a). (c) Pre-multiplying the right-hand side of the formula by $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ yields the identity. (d) Comparing the $(1,1)$-blocks of the formula in (c) gives

$$[A - BD^{-1}C]^{-1} = A^{-1} + A^{-1}B[D - CA^{-1}B]^{-1}CA^{-1}$$

By changing the sign of D, we get the desired result.

2.4 Let $Px = \lambda x$, $x \neq 0$. Then, $P(Px) = P(\lambda x) = \lambda^2 x$ holds. Hence, from $P^2 = P$, we have $Px = \lambda^2 x$. It thus follows that $\lambda x = \lambda^2 x$ for $x \neq 0$, implying that $\lambda = 0$ or $\lambda = 1$. Suppose that $\lambda_1 = \cdots = \lambda_r = 1$ and $\lambda_{r+1} = \cdots = \lambda_n = 0$.

We see that $\mathrm{trace}(P) = \sum_{i=1}^{n} \lambda_i = r$. Let the SVD of P be given by $P = U\Sigma V^{\mathrm{T}}$, where $\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_n)$. Since, in this case, $\sigma_i = \lambda_i$, we see that $\mathrm{rank}(P) = \mathrm{rank}(\Sigma) = r$.

2.5 Suppose that $P^2 = P$ holds. Then, from Lemma 2.4 and Corollary 2.1, we have (2.17) and (2.18). Thus (a) implies (b).

We show (b) $\rightarrow$ (c). As in the proof of Lemma 2.5, we define $\mathcal{V} = \mathrm{Im}(P)$ and $\mathcal{W} = \mathrm{Im}(I_n - P)$. Note that for the dimensions of subspaces, we have

$$\dim(\mathcal{V} \vee \mathcal{W}) = \dim(\mathcal{V}) + \dim(\mathcal{W}) - \dim(\mathcal{V} \cap \mathcal{W})$$

Since $x = Px + (I_n - P)x$, it follows that $\mathbb{R}^n = \mathcal{V} \vee \mathcal{W}$ and $n = \dim(\mathcal{V} \vee \mathcal{W})$. Also, from (b), we get $\dim(\mathcal{V}) + \dim(\mathcal{W}) = n$, and hence $\dim(\mathcal{V} \cap \mathcal{W}) = 0$. This implies that $\mathcal{V} \cap \mathcal{W} = \{0\}$, so that (c) holds.

Finally, we show (c) $\rightarrow$ (a). Post-multiplying $I_n = P + (I_n - P)$ by P yields $P = P^2 + (I_n - P)P$, so that we have $P(I_n - P) = (I_n - P)P$. Thus

$$\mathrm{Im}\, P(I_n - P) \subset \mathrm{Im}(P), \qquad \mathrm{Im}\,(I_n - P)P \subset \mathrm{Im}(I_n - P)$$

hold. If (c) holds, we get $\mathrm{Im}(P) \cap \mathrm{Im}(I_n - P) = \{0\}$, implying that $\mathrm{Im}[P(I_n - P)] = \{0\}$ follows. Hence, we have $P^2 = P$. This completes the proof.

2.6 Since $LT = I_r$, we get $P^2 = TLTL = TL = P$. Also, T and L are of full rank, so that $\mathrm{Im}(P) = \mathrm{Im}(TL) = \mathrm{Im}(T)$ and $\mathrm{Ker}(P) = \mathrm{Ker}(TL) = \mathrm{Ker}(L)$. This implies that P is the oblique projection on $\mathrm{Im}(T)$ along $\mathrm{Ker}(L)$. Similarly, we can prove that Q is a projection.

2.7 Define $L = [L_1 \;\; L_2]$ and $V = [V_1 \;\; V_2]$. Since $\begin{bmatrix} L \\ V \end{bmatrix} [T \;\; U] = \begin{bmatrix} I_r & 0 \\ 0 & I_{n-r} \end{bmatrix}$, we have

$$\begin{bmatrix} L_1 & L_2 \\ V_1 & V_2 \end{bmatrix} \begin{bmatrix} I_r & -X \\ 0 & I_{n-r} \end{bmatrix} = \begin{bmatrix} I_r & 0 \\ 0 & I_{n-r} \end{bmatrix}$$

This implies that $L_1 = I_r$, $L_2 = X$, $V_1 = 0$, $V_2 = I_{n-r}$, and hence

$$P = \begin{bmatrix} I_r & X \\ 0 & 0 \end{bmatrix}$$

2.8 (a) Let $P = V_r V_r^{\mathrm{T}}$. Then, $P^2 = P$ and $P^{\mathrm{T}} = P$ hold, so that P is an orthogonal projection. Also, from Lemma 2.9, we have $\mathrm{Im}(A^{\mathrm{T}}) = \mathrm{Im}(V_r) = \mathrm{Im}(V_r V_r^{\mathrm{T}})$. Similarly, we can prove (b), (c), (d).

2.9 Let $A = U\Sigma V^{\mathrm{T}}$, where $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$ are orthogonal and $\Sigma = \begin{bmatrix} \Sigma_s & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{m \times n}$ with $\Sigma_s \in \mathbb{R}^{r \times r}$ diagonal, where $r := \mathrm{rank}(A)$. Then, we get

$$(AA^{\mathrm{T}})^{\dagger} = (U\Sigma\Sigma^{\mathrm{T}}U^{\mathrm{T}})^{\dagger} = U(\Sigma\Sigma^{\mathrm{T}})^{\dagger}U^{\mathrm{T}}$$

where

$$(\Sigma\Sigma^{\mathrm{T}})^{\dagger} = \begin{bmatrix} \Sigma_s^{-2} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{m \times m}$$

Thus

$$A^{\mathrm{T}}(AA^{\mathrm{T}})^{\dagger} = V \begin{bmatrix} \Sigma_s & 0 \\ 0 & 0 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \Sigma_s^{-2} & 0 \\ 0 & 0 \end{bmatrix} U^{\mathrm{T}} = V \begin{bmatrix} \Sigma_s^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^{\mathrm{T}} = A^{\dagger}$$

That $(A^{\mathrm{T}}A)^{\dagger}A^{\mathrm{T}} = A^{\dagger}$ is proved similarly.

2.10 Let $A = U\Sigma V^{\mathrm{T}}$, where $\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_n)$, and $U \in \mathbb{R}^{m\times n}$, $V \in \mathbb{R}^{n\times n}$. Then, we have $Q = UV^{\mathrm{T}}$ and $\Pi = V\Sigma V^{\mathrm{T}}$. Note that $VV^{\mathrm{T}} = V^{\mathrm{T}}V = I_n$.

Chapter 3

3.1 Since $|g_k| = 1/k$, $k = 1, 2, \cdots$, we have

$$\sum_{k=1}^{\infty} |g_k| = \sum_{k=1}^{\infty} \frac{1}{k} = \infty$$

This implies that the system is not stable.

3.2 To apply the Routh-Hurwitz test for a continuous-time LTI system to a discrete-time LTI system, let $z = (s+1)/(s-1)$. Then, we see that $|z| < 1 \;\Leftrightarrow\; \mathfrak{Re}[s] < 0$. From $f((s+1)/(s-1)) = 0$, we get

$$(1 + a_1 + a_2)s^2 + 2(1 - a_2)s + (1 - a_1 + a_2) = 0$$

Thus the stability condition for $z^2 + a_1 z + z_2$ is given by

$$1 + a_1 + a_2 > 0, \qquad 1 - a_1 + a_2 > 0, \qquad 1 - a_2 > 0 \tag{E.1}$$

3.3 From a diagonal system of Figure 3.3, A and B are given by

$$A = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}, \qquad B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Thus, from Theorem 3.4 (ii), it suffices to find the condition such that

$$\mathrm{rank} \begin{bmatrix} \lambda_1 - z & 0 & 0 & b_1 \\ 0 & \lambda_2 - z & 0 & b_2 \\ 0 & 0 & \lambda_3 - z & b_3 \end{bmatrix} = 3, \qquad z = \lambda_1,\ \lambda_2,\ \lambda_3$$

holds. Hence, the reachability condition becomes

$$b_1 b_2 b_3 \neq 0, \qquad (\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)(\lambda_3 - \lambda_1) \neq 0$$

3.4 Note that $\mathcal{C} = [b \;\; Ab \;\; \cdots \;\; A^{n-1}b]$. From (2.3), we have

$$A^n = -(\alpha_1 A^{n-1} + \cdots + \alpha_{n-1}A + \alpha_n I)$$

Hence,

$$
\begin{aligned}
A\mathcal{C} &= [Ab \;\; A^2b \;\; \cdots \;\; A^nb] \\
&= [Ab \;\; A^2b \;\; \cdots \;\; -(\alpha_1 A^{n-1} + \cdots + \alpha_{n-1}A + \alpha_n I)b] \\
&= [b \;\; Ab \;\; \cdots \;\; A^{n-1}b] \begin{bmatrix} 0 & & -\alpha_n \\ 1 & & -\alpha_{n-1} \\ & \ddots & \vdots \\ & 1 & -\alpha_1 \end{bmatrix} = \mathcal{C}\bar{A}
\end{aligned}
$$

and

$$
b = [b \;\; Ab \;\; \cdots \;\; A^{n-1}b] \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathcal{C}\bar{b}
$$

3.5 We can show that

$$
\operatorname{rank}[A + BK - \lambda I \;\; B] = n \quad \Leftrightarrow \quad \operatorname{rank}[A - \lambda I \;\; B] = n
$$

and

$$
\operatorname{rank} \begin{bmatrix} A + LC - \lambda I \\ C \end{bmatrix} = n \quad \Leftrightarrow \quad \operatorname{rank} \begin{bmatrix} A - \lambda I \\ C \end{bmatrix} = n
$$

The results follow from Theorems 3.4 $\sim$ 3.9.

3.6 Define $\bar{A} := A/(\rho(A) + \varepsilon)$. Then, the spectral radius of $\bar{A}$ is strictly less than 1, and hence $\bar{A}^k \to 0$ as $k \to \infty$. Thus, in particular, the elements of the sequence $\{\bar{A}^k,\ k = 1, 2, \cdots\}$ are bounded, so that we have $\left|(\bar{A}^k)_{ij}\right| \le C$, $C > 0$ for $k = 1, 2, \cdots$ and $i, j = 1, \cdots, n$. Since $(\bar{A}^k)_{ij} = (A^k)_{ij}/(\rho(A) + \varepsilon)^k$, we get the desired result.

3.7 Before proving this assertion, it will be helpful to look at the proof of a basic convergence result for the Césaro sum in Problem 4.3 (a).

The solution $x(t)$ is given by

$$
x(t) = A^t x(0) + \sum_{k=0}^{t-1} A^{t-k-1} f(k)
$$

By assumption, $\rho(A) < 1$. Thus we can take $\varepsilon > 0$ such that $\rho(A) + \varepsilon =: a < 1$. From Problem 3.6,

$$
\left|(A^k)_{ij}\right| \le C(\rho(A) + \varepsilon)^k \quad \Rightarrow \quad \|A^k\|_\alpha \le C_1 a^k, \quad k = 1, 2, \cdots
$$

where $C_1 > 0$, and $\|\cdot\|_\alpha$ is a matrix norm (see Section 2.3). By using the above estimate,

$$
\|x(t)\| \le C_1 a^t \|x(0)\| + C_1 \sum_{k=0}^{t-1} a^{t-k-1} \|f(k)\|
$$

Since the first term tends to zero as $t \to \infty$, it suffices to show that the second term tends to zero as $t \to \infty$. Let the second term be $g(t)$. Then, we get

$$g(t) = C_1 a^{t-1} \sum_{k=0}^{t-1} a^{-k} \beta(k), \qquad \beta(k) = \|f(k)\|$$

By hypothesis, $\lim_{k\to\infty} \beta(k) = 0$, so that for any $\varepsilon_1 > 0$, there exists $N_0 > 0$ such that $\beta(k) < \varepsilon_1(1-a)/(aC_1)$ for all $k > N_0$. Thus, for a sufficiently large t,

$$\begin{aligned} g(t) &= C_1 a^{t-1} \left[\sum_{k=0}^{N_0} a^{-k} \beta(k) + \sum_{k=N_0+1}^{t-1} a^{-k} \beta(k) \right] \\ &\le C_1 a^{t-1} \left[\sum_{k=0}^{N_0} a^{-k} \beta(k) + \frac{\varepsilon(1-a)}{aC_1} \sum_{k=N_0+1}^{t-1} a^{-k} \right] \\ &= C_1 a^{t-1} \left[\sum_{k=0}^{N_0} a^{-k} \beta(k) + \frac{\varepsilon_1(1-a)}{aC_1} \left(\frac{a^{-N_0-1} - a^{-t+1}}{1-a^{-1}} \right) \right] \\ &\le C_1 a^{t-1} \left[\sum_{k=0}^{N_0} a^{-k} \beta(k) \right] + \varepsilon_1 \left[1 - a^{t-N_0-2} \right] \end{aligned}$$

The first term in the right-hand side of the above inequality tends to zero as $t \to \infty$, while the second term is smaller than ε_1. This completes the proof.

3.8 It can be shown that

$$\begin{bmatrix} zI - A & B \\ -C & D \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ -C(zI-A)^{-1} & I_p \end{bmatrix} \begin{bmatrix} zI - A & B \\ 0 & G(z) \end{bmatrix}$$

where $G(z) = C(zI-A)^{-1}B$. Since $\operatorname{rank} \begin{bmatrix} I_n & 0 \\ -C(zI-A)^{-1} & I_p \end{bmatrix} = n + p$, we get

$$\operatorname{rank}_z S(z) = \operatorname{rank}_z (zI - A) + \operatorname{rank}_z G(z) = n + \operatorname{rank}_z G(z)$$

3.9 ([51], vol. 2, pp. 206–207) Suppose that $R(z) = b(z)/a(z)$ is rational, and that the series expansion

$$\frac{b(z)}{a(z)} = \frac{h_1}{z} + \frac{h_2}{z^2} + \cdots \tag{E.2}$$

converges for $|z| > \rho$ for some $\rho > 0$. Suppose that polynomials $a(z)$ and $b(z)$ are given by

$$a(z) = z^m + a_1 z^{m-1} + \cdots + a_m, \quad b(z) = b_1 z^{m-1} + b_2 z^{m-2} + \cdots + b_m$$

Multiplying (E.2) by $a(z)$ yields

$$b_1 z^{m-1} + b_2 z^{m-2} + \cdots + b_m = (z^m + a_1 z^{m-1} + \cdots + a_m) \left(\frac{h_1}{z} + \frac{h_2}{z^2} + \cdots \right)$$

Equating the coefficients of equal powers of z on both sides, we obtain

$$\begin{aligned} b_1 &= h_1 \\ b_2 &= h_2 + a_1 h_1 \\ &\vdots \\ b_m &= h_m + a_1 h_{m-1} + \cdots + a_{m-1} h_1 \end{aligned}$$

and for $j = m+1, \cdots,$

$$0 = h_j + a_1 h_{j-1} + \cdots + a_m h_{j-m}$$

This implies that (2.40) holds with $r = m$, so that the Hankel matrix (2.35) has finite rank.

Conversely, if H has finite rank, then (2.40) holds from Lemma 2.14. Hence, by using $a_1, \cdots, a_r$ of (2.40) and the above relations, we can define $b_1, \cdots, b_r$. Thus we see that $b(z)/a(z)$ is a desired rational function, which equals $R(z) = h_1/z + h_2/z^2 + \cdots$.

3.10 Note that the following power series expansion:

$$\log(1+z^{-1}) = z^{-1} - \frac{1}{2}z^{-2} + \frac{1}{3}z^{-3} - \cdots, \qquad |z| > 1$$

Thus the right-hand side converges to a non-rational transfer function, implying that the impulse response cannot be realized by a state space model.

Chapter 4

4.1 Putting $i - j = k$, we change variables from (i, j) to (j, k). Then, k is bounded by $-N+1 \le k \le N-1$, and j is bounded by $1 \le j \le N-k$ if $k \ge 0$ and by $-k+1 \le j \le N$ if $k < 0$. Thus we get

$$\begin{aligned} \sum_{i=1}^{N}\sum_{j=1}^{N} \phi(i-j) &= \sum_{k=0}^{N-1}\sum_{j=1}^{N-k} \phi(k) + \sum_{k=-N+1}^{-1}\sum_{j=-k+1}^{N} \phi(k) \\ &= \sum_{k=0}^{N-1} (N-k)\phi(k) + \sum_{k=-N+1}^{-1} (N+k)\phi(k) \\ &= \sum_{k=-N+1}^{N-1} (N-|k|)\phi(k) \end{aligned}$$

4.2 Define $k = t - s$. Then, applying the formula in Problem 4.1, we have

$$\sum_{i=-N}^{N}\sum_{j=-N}^{N} \Lambda(t-s) = \sum_{k=-2N}^{2N} (2N+1-|k|)\Lambda(k)$$

Thus dividing the above equation by $2N+1$ gives (4.13).

4.3 (a) Let $\varepsilon > 0$ be a small number. From the assumption, there exists an integer $p > 0$ such that $|a_k| < \varepsilon$ for $k > p$. Let $M = \max\{|a_1|, \cdots, |a_p|\}$. Then,

$$\left|\frac{a_1 + \cdots + a_n}{n}\right| < \frac{pM + \varepsilon(n-p)}{n} < \frac{pM}{n} + \varepsilon$$

Taking the limit $n \to \infty$, we have $pM/n \to 0$. Since $\varepsilon > 0$ is arbitrary, the assertion is proved.

(b) Define $B_n = (1/n)\sum_{k=1}^{n} a_k$ with $B_0 = 0$. Then, $\lim_{n\to\infty} |B_n| = 0$ by hypothesis. Noting that

$$ka_k = k^2 B_k - (k-1)^2 B_{k-1} - (k-1)B_{k-1}$$

we have

$$I_n := \frac{1}{n}\sum_{k=1}^{n}\left(1 - \frac{k}{n}\right)a_k = B_n - \frac{1}{n^2}\sum_{k=1}^{n} ka_k = \frac{1}{n^2}\sum_{k=1}^{n}(k-1)B_{k-1}$$

Thus

$$|I_n| \le \frac{1}{n}\sum_{k=1}^{n}\left|\frac{k-1}{n}\right||B_{k-1}| \le \frac{1}{n}\sum_{k=1}^{n}|B_{k-1}| \to 0$$

since $\lim_{k\to\infty}|B_k| = 0$.

(c) Define $C_n = \sum_{k=n}^{\infty} a_k$. By assumption, $\lim_{n\to\infty} C_n = 0$. It can be shown that

$$\left|\sum_{k=1}^{\infty} a_k - \sum_{k=1}^{n}\left(1 - \frac{k}{n}\right)a_k\right| \le \left|\sum_{k=n+1}^{\infty} a_k\right| + \frac{1}{n}\left|\sum_{k=1}^{n} ka_k\right|$$

$$= |C_{n+1}| + \frac{1}{n}\left|\sum_{k=1}^{n} ka_k\right|$$

Since the first term in the right-hand side of the above equation converges to zero, it remains to prove the convergence of the second term. By the definition of C_n,

$$\frac{1}{n}\sum_{k=1}^{n} ka_k = \frac{1}{n}\sum_{k=1}^{n} k(C_k - C_{k+1}) = \frac{1}{n}\sum_{k=1}^{n}\big([kC_k - (k+1)C_{k+1}] + C_{k+1}\big)$$

$$= \frac{C_1}{n} - \frac{n+1}{n}C_{n+1} + \frac{1}{n}\sum_{k=1}^{n} C_{k+1}$$

We see that the first and second terms of the right-hand side of the above equation converge to zero, and the third term also converges to zero by (a).

4.4 For zero mean Gaussian random variables $a,\ b,\ c,\ d$, we have (see *e.g.* [145])

$$E\{abcd\} = E\{ab\}E\{cd\} + E\{ac\}E\{bd\} + E\{ad\}E\{bc\} \tag{E.3}$$

By using (E.3), it follows from (4.17) that

$$\begin{aligned}\Lambda_{\xi\xi}(k) &= E\{x(t+l+k)x(t+k)x(t+l)x(t)\} - \mu_\xi^2 \\ &= \Lambda_{xx}(l)\Lambda_{xx}(l) + \Lambda_{xx}(k)\Lambda_{xx}(k) + \Lambda_{xx}(l+k)\Lambda_{xx}(l-k) - \mu_\xi^2 \\ &= \Lambda_{xx}^2(k) + \Lambda_{xx}(l+k)\Lambda_{xx}(l-k)\end{aligned}$$

By the Schwartz inequality,

$$\begin{aligned}\left|\sum_{k=0}^{N}\Lambda_{\xi\xi}(k)\right| &\le \sum_{k=0}^{N}\Lambda_{xx}^2(k) + \left|\sum_{k=0}^{N}\Lambda_{xx}(k+l)\Lambda_{xx}(k-l)\right| \\ &\le \sum_{k=0}^{N}\Lambda_{xx}^2(k) + \left(\sum_{k=0}^{N}\Lambda_{xx}^2(k+l)\sum_{k=0}^{N}\Lambda_{xx}^2(k-l)\right)^{1/2}\end{aligned} \tag{E.4}$$

Since (4.20) holds, it follows that

$$\lim_{N\to\infty}\frac{1}{N+1}\sum_{k=0}^{N}\Lambda_{xx}^2(k\pm l) = 0, \qquad l = 0,\ 1,\ \cdots$$

and hence from (E.4)

$$\lim_{N\to\infty}\frac{1}{N+1}\sum_{k=0}^{N}\Lambda_{\xi\xi}(k) = 0$$

This implies that (4.19) holds from Problem 4.3 (c).

4.5 Similarly to the calculation in Problem 4.2, we have

$$\begin{aligned}I_N(\omega) &= \frac{1}{2N+1}\sum_{l=-N}^{N}\sum_{k=-N}^{N}E\{x(l)x(k)\}e^{-j\omega(l-k)} \\ &= \sum_{\tau=-2N}^{2N}\left(1 - \frac{|\tau|}{2N+1}\right)\Lambda(\tau)e^{-j\omega\tau}\end{aligned} \tag{E.5}$$

Note that

$$\lim_{N\to\infty}\sum_{\tau=-2N}^{2N}\Lambda(\tau)e^{-j\omega\tau} = \Phi(\omega)$$

exists. It therefore from Problem 4.3 (c) that the limit of the right-hand side of (E.5) converges to $\Phi(\omega)$.

4.6 A proof is similar to that of Lemma 4.4. Post-multiplying

$$y(t) = \sum_{k=0}^{\infty} g_k u(t-k)$$

by $u(t-l)$ and taking the expectation yield

$$\Lambda_{yu}(l) = \sum_{k=0}^{\infty} g_k E\{u(t-k)u(t-l)\} = \sum_{k=0}^{\infty} g_k \Lambda_{uu}(l-k)$$

Post-multiplying the above equation by $e^{-j\omega l}$ and summing up with respect to l yield

$$\begin{aligned}\Phi_{yu}(\omega) &= \sum_{l=-\infty}^{\infty} \sum_{k=0}^{\infty} g_k e^{-j\omega k} \Lambda_{uu}(l-k) e^{-j\omega(l-k)} \\ &= \sum_{k=0}^{\infty} g_k e^{-j\omega k} \Phi_{uu}(\omega) = G(e^{j\omega}) \Phi_{uu}(\omega)\end{aligned}$$

4.7 Since $\Phi_{yy}(\omega) = 2 - 2\cos\omega = 4\sin^2(\omega/2)$,

$$\begin{aligned}\int_{-\pi}^{\pi} \log \Phi_{yy}(\omega) d\omega &= 2 \int_0^{\pi} \log\left[4\sin^2(\omega/2)\right] d\omega \\ &= 4\pi \log 2 + 4 \int_0^{\pi} \log \sin(\omega/2) d\omega \\ &= 4\pi \log 2 + 8 \int_0^{\pi/2} \log \sin\theta \, d\theta = 0 > -\infty\end{aligned}$$

where $\int_0^{\pi/2} \log\sin\theta d\theta = -(\pi/2)\log 2$ (Euler) is used.

4.8 The form of $\Phi(\omega)$ implies that y is a one-dimensional ARMA process, so that

$$y(t) + ay(t-1) = e(t) + ce(t-1)$$

Thus from (4.35), the spectral density function of y becomes

$$\Phi(\omega) = \sigma^2 \left| \frac{1 + ce^{j\omega}}{1 + ae^{j\omega}} \right|^2 = \sigma^2 \frac{1 + c^2 + 2c\cos\omega}{1 + a^2 + 2a\cos\omega}$$

Comparing the coefficients, we have $a = -0.9$, $c = 0.5$.

4.9 Since $H(z) = (z+c)/(z+a)$, we have

$$z^m H(z) = \frac{z^m(z+c)}{z+a} = (z^m + cz^{m-1})(1 + (-a)z^{-1} + (-a)^2 z^{-2} + \cdots)$$

Computing $[z^m H(z)]_+$ yields

$$\begin{aligned}[z^m H(z)]_+ &= (-a)^m + (-a)^{m+1} z^{-1} + \cdots + c[(-a)^{m-1} + (-a)^m z^{-1} + \cdots] \\ &= (-a)^{m-1}(c-a)(1 + (-a)z^{-1} + (-a)^2 z^{-1} + \cdots) \\ &= \frac{(-a)^{m-1}(c-a)z}{z+a}\end{aligned}$$

Thus from (4.57), the optimal predictor is given by

$$G(z) = \frac{(-a)^{m-1}(c-a)z}{z+a}\,\frac{z+a}{z+c} = \frac{(-a)^{m-1}(c-a)z}{z+c}$$

4.10 From (4.58),

$$\begin{bmatrix} x(t+1) \\ y(t+1) \end{bmatrix} = \begin{bmatrix} A(t) & 0 \\ C(t)A(t) & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} + \begin{bmatrix} B(t) & 0 \\ C(t)B(t) & I \end{bmatrix} \begin{bmatrix} w(t) \\ v(t+1) \end{bmatrix}$$

This is a state space equation, implying that the joint process $(x,\ y)$ is Markov.

4.11 A proof is by direct substitution.

4.12 By definition,

$$\Phi_{yy}(z) = \sum_{l=-\infty}^{-1} \bar{C}(A^{\mathrm{T}})^{-l-1}C^{\mathrm{T}}z^{-l} + \Lambda_{yy}(0) + \sum_{l=1}^{\infty} CA^{l-1}\bar{C}^{\mathrm{T}}z^{-l}$$

Since $\bar{C}^{\mathrm{T}} = A\Pi C^{\mathrm{T}} + S$, we compute the terms that include S. Thus,

$$\begin{aligned} I_S :&= \sum_{l=-\infty}^{-1} S^{\mathrm{T}}(A^{\mathrm{T}})^{-l-1}C^{\mathrm{T}}z^{-l} + \sum_{l=1}^{\infty} CA^{l-1}Sz^{-l} \\ &= S^{\mathrm{T}}\left(\sum_{l=1}^{\infty}(A^{\mathrm{T}})^{l-1}z^{l}\right)C^{\mathrm{T}} + C\left(\sum_{l=1}^{\infty}A^{l-1}z^{-l}\right)S \\ &= S^{\mathrm{T}}(z^{-1}I - A^{\mathrm{T}})^{-1}C^{\mathrm{T}} + C(zI-A)^{-1}S \\ &= S^{\mathrm{T}}W^{\mathrm{T}}(z^{-1}) + W(z)S \end{aligned}$$

Adding I_S to the right-hand side of (4.80) yields (4.81).

Chapter 5

5.1 This is a special case of Lemma 5.1.

5.2 Let $K_\alpha(t)$ and $P_\alpha(t)$ respectively be the Kalman gain and the error covariance matrices corresponding to $\alpha Q(t)$, $\alpha S(t)$, $\alpha R(t)$, $\alpha P(0)$. We use the algorithm of Theorem 5.1. For $t = 0$, it follows from (5.41a) that

$$\begin{aligned} K_\alpha(0) &= [A(0)\alpha P(0)C^{\mathrm{T}}(0) + B(0)\alpha S(0)][C(0)\alpha P(0)C^{\mathrm{T}}(0) + \alpha R(0)]^{-1} \\ &= K(0) \end{aligned}$$

Also, from (5.42a),

$$\begin{aligned} P_\alpha(1) = A(0)\alpha P(0)A^{\mathrm{T}}(0) &- K_\alpha(0)[C(0)\alpha P(0)C^{\mathrm{T}}(0) + \alpha R(0)]K_\alpha^{\mathrm{T}}(0) \\ &+ B(0)\alpha Q(0)B^{\mathrm{T}}(0) = \alpha P(1) \end{aligned}$$

Similarly, for $t = 1$, we have $K_\alpha(1) = K(1)$, $P_\alpha(2) = \alpha P(2)$, and hence inductively $K_\alpha(t) = K(t)$, $t = 2,\ 3,\ \cdots$.

5.3 It follows that

$$x(t) - \mu_x(t) = \tilde{x}(t \mid t-1) + (\hat{x}(t \mid t-1) - \mu_x(t))$$

where $\tilde{x}(t \mid t-1) \perp \hat{x}(t \mid t-1) - \mu_x(t)$. Thus we have

$$\Pi(t) = P(t \mid t-1) + \Sigma(t)$$

Since $\Sigma(t) \geq 0$ and $P(t \mid t-1) \geq P(t \mid t) \geq 0$, we get

$$\Pi(t) \geq P(t \mid t-1) \geq P(t \mid t) \geq 0, \qquad \Pi(t) \geq \Sigma(t)$$

5.4 Follow the hint.

5.5 The derivation is straightforward.

5.6 Substituting $A = \Phi + SR^{-1}C$ into (5.68), we have

$$\begin{aligned} K &= [(\Phi + SR^{-1}C)PC^{\mathrm{T}} + S](CPC^{\mathrm{T}} + R)^{-1} \\ &= \Phi PC^{\mathrm{T}}(CPC^{\mathrm{T}} + R)^{-1} + SR^{-1} = \Gamma + SR^{-1} \end{aligned}$$

Thus we get $A - KC = \Phi - \Gamma C$.

It follows from (5.67) that

$$\begin{aligned} P &= APA^{\mathrm{T}} - K(CPC^{\mathrm{T}} + R)K^{\mathrm{T}} + Q \\ &= APA^{\mathrm{T}} - (\Gamma + SR^{-1})(CPC^{\mathrm{T}} + R)(\Gamma + SR^{-1})^{\mathrm{T}} + Q \\ &= (\Phi + SR^{-1}C)P(\Phi + SR^{-1}C)^{\mathrm{T}} \\ &\quad - (\Gamma + SR^{-1})(CPC^{\mathrm{T}} + R)(\Gamma + SR^{-1})^{\mathrm{T}} + Q \end{aligned}$$

From the definition of Γ,

$$\begin{aligned} P &= \Phi P\Phi^{\mathrm{T}} - \Gamma(CPC^{\mathrm{T}} + R)\Gamma^{\mathrm{T}} + Q + SR^{-1}CPF^{\mathrm{T}} \\ &\quad + \Phi PC^{\mathrm{T}}R^{-1}S^{\mathrm{T}} + SR^{-1}CPC^{\mathrm{T}}R^{-1}S^{\mathrm{T}} \\ &\quad - \Gamma(CPC^{\mathrm{T}} + R)R^{-1}S^{\mathrm{T}} - SR^{-1}(CPC^{\mathrm{T}} + R)\Gamma^{\mathrm{T}} \\ &\quad - SR^{-1}(CPC^{\mathrm{T}} + R)R^{-1}S^{\mathrm{T}} \\ &= \Phi P\Phi^{\mathrm{T}} - \Gamma(CPC^{\mathrm{T}} + R)\Gamma^{\mathrm{T}} + (Q - SR^{-1}S^{\mathrm{T}}) \end{aligned}$$

This proves (5.70) since $M = Q - SR^{-1}S^{\mathrm{T}}$.

5.7 Equation (5.90) is given by

$$\Sigma = A\Sigma A^{\mathrm{T}} + (\bar{C}^{\mathrm{T}} - A\Sigma C^{\mathrm{T}})(\Lambda(0) - C\Sigma C^{\mathrm{T}})^{-1}(\bar{C} - C\Sigma A^{\mathrm{T}}) \tag{E.6}$$

Using $A = F + \bar{C}^{\mathrm{T}}\Lambda^{-1}(0)C$, the first term in the right-hand side of (E.6) is

$$\begin{aligned} I_1 :&= (F + \bar{C}^{\mathrm{T}}\Lambda^{-1}(0)C)\Sigma(F + \bar{C}^{\mathrm{T}}\Lambda^{-1}(0)C)^{\mathrm{T}} \\ &= F\Sigma F^{\mathrm{T}} + \bar{C}^{\mathrm{T}}\Lambda^{-1}(0)C\Sigma F^{\mathrm{T}} + F\Sigma C^{\mathrm{T}}\Lambda^{-1}(0)\bar{C} \\ &\quad + \bar{C}^{\mathrm{T}}\Lambda^{-1}(0)C\Sigma C^{\mathrm{T}}\Lambda^{-1}(0)\bar{C} \end{aligned}$$

Also, we have

$$\bar{C}^{\mathrm{T}} - A\Sigma C^{\mathrm{T}} = -F\Sigma C^{\mathrm{T}} + \bar{C}^{\mathrm{T}}\Lambda^{-1}(0)(\Lambda(0) - C\Sigma C^{\mathrm{T}})$$

so that the second term in the right-hand side of (E.6) becomes

$$\begin{aligned} I_2 :&= (-F\Sigma C^{\mathrm{T}} + \bar{C}^{\mathrm{T}}\Lambda^{-1}(0)[\Lambda(0) - C\Sigma C^{\mathrm{T}}])(\Lambda(0) - C\Sigma C^{\mathrm{T}})^{-1} \\ &\quad \times (-F\Sigma C^{\mathrm{T}} + \bar{C}^{\mathrm{T}}\Lambda^{-1}(0)[\Lambda(0) - C\Sigma C^{\mathrm{T}}])^{\mathrm{T}} \\ &= F\Sigma C^{\mathrm{T}}(\Lambda(0) - C\Sigma C^{\mathrm{T}})^{-1}C\Sigma F^{\mathrm{T}} - F\Sigma C^{\mathrm{T}}\Lambda^{-1}(0)\bar{C} \\ &\quad - \bar{C}^{\mathrm{T}}\Lambda^{-1}(0)C\Sigma F^{\mathrm{T}} + \bar{C}^{\mathrm{T}}\Lambda^{-1}(0)(\Lambda(0) - C\Sigma C^{\mathrm{T}})\Lambda^{-1}(0)\bar{C} \end{aligned}$$

Computing $I_1 + I_2$, we get (5.91).

5.8 Since

$$N = \begin{bmatrix} F^{\mathrm{T}} & 0 \\ -\bar{C}^{\mathrm{T}}\Lambda^{-1}(0)\bar{C} & I_n \end{bmatrix}, \qquad L = \begin{bmatrix} I_n & -C^{\mathrm{T}}\Lambda^{-1}(0)C \\ 0 & F \end{bmatrix}$$

we have

$$\begin{aligned} L\hat{J}L^{\mathrm{T}} &= \begin{bmatrix} I_n & -C^{\mathrm{T}}\Lambda^{-1}(0)C \\ 0 & F \end{bmatrix} \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix} \begin{bmatrix} I_n & -C^{\mathrm{T}}\Lambda^{-1}(0)C \\ 0 & F \end{bmatrix}^{\mathrm{T}} \\ &= \begin{bmatrix} 0 & F^{\mathrm{T}} \\ -F & 0 \end{bmatrix} = N\hat{J}N^{\mathrm{T}} \end{aligned}$$

Consider the following two eigenvalue problems:

$$\text{(A)} \qquad Nx = \lambda Lx \qquad\qquad \text{(B)} \qquad L^{\mathrm{T}}x = \mu N^{\mathrm{T}}x$$

Let $\lambda \neq 0$ be an eigenvalue of Problem (A). Since

$$\det(L^{\mathrm{T}} - \mu N^{\mathrm{T}}) = \det(L - \mu N) = 0$$

we see that $\mu = 1/\lambda$ is an eigenvalue of Problem (B). Also, pre-multiplying $L^{\mathrm{T}}x = \mu N^{\mathrm{T}}x$ by $N\hat{J}$ yields

$$N\hat{J}L^{\mathrm{T}}x = \mu N\hat{J}N^{\mathrm{T}}x = \mu L\hat{J}L^{\mathrm{T}}x \quad \Rightarrow \quad Nz = \mu Lz, \quad z = \hat{J}L^{\mathrm{T}}x$$

Thus $\mu = 1/\lambda$ is also an eigenvalue of Problem (A).

Chapter 6

6.1 (a) Since $g_k = k,\ k = 1, \cdots, g_0 = 0$, we have

$$H_{44} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \end{bmatrix}, \qquad \text{rank}\, H_{44} = 2$$

By using the MATLAB® program in Table D.1, we get

$$A = \begin{bmatrix} 1.2182 & -0.2182 \\ 0.2182 & 0.7818 \end{bmatrix}, \quad B = \begin{bmatrix} -1.3039 \\ 0.8368 \end{bmatrix}, \quad C = [-1.3039 \;\; -0.8368]$$

Thus the transfer function is given by $G(z) = z/(z-1)^2$.

(b) In this case, the Hankel matrix becomes

$$H_{66} = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 & 1 & 0 \end{bmatrix}, \qquad \text{rank}\, H_{66} = 4$$

so that we have

$$A = \begin{bmatrix} 0.1450 & 0.8808 & -0.3327 & -0.3239 \\ -0.8808 & 0.3551 & 0.3533 & -0.0115 \\ 0.3327 & -0.3533 & -0.6187 & -0.5087 \\ 0.3239 & -0.0115 & 0.5087 & -0.8814 \end{bmatrix}, \qquad B = \begin{bmatrix} -1.0016 \\ 0.1151 \\ -0.2418 \\ 0.2200 \end{bmatrix}$$

$$C = [-1.0016 \;\; -0.1151 \;\; -0.2418 \;\; -0.2200]$$

Thus the transfer function is given by $G(z) = (z^3 + z^2)/(z^4 + z^3 + z^2 + z + 1)$.

6.2 Let P be the reachability Gramian. Substituting $A = SA^{\mathrm{T}}S$, $B = SC^{\mathrm{T}}$ into (3.34) yields

$$P = APA^{\mathrm{T}} + BB^{\mathrm{T}} = SA^{\mathrm{T}}SPSAS + SC^{\mathrm{T}}CS$$

Since $SS = I$, we get $SPS = A^{\mathrm{T}}(SPS)A + C^{\mathrm{T}}C$. Thus the observability Gramian is expressed as $Q = SPS$. Though (A, B, C) are not balanced, both Gramians have the same eigenvalues. Note that Σ_s (with $T = I$) is diagonal, *i.e.*,

$$\Sigma_s = \mathcal{C}_k \mathcal{C}_k^{\mathrm{T}} = \sum_{i=0}^{k-1} A^i BB^{\mathrm{T}} (A^{\mathrm{T}})^i \quad \left(\neq \sum_{i=0}^{\infty} A^i BB^{\mathrm{T}} (A^{\mathrm{T}})^i = P \right)$$

6.3 Since the orthogonal projection is expressed as $\hat{E}\{A \mid B\} = KB$, $K \in \mathbb{R}^{p \times q}$, the optimality condition is reduced to $A - KB \perp B$. Hence we have

$$(A - KB)B^{\mathrm{T}} = 0 \quad \Rightarrow \quad K = (AB^{\mathrm{T}})(BB^{\mathrm{T}})^{\dagger}$$

showing that $\hat{E}\{A \mid B\} = (AB^{\mathrm{T}})(BB^{\mathrm{T}})^{\dagger}B$.

6.4 Since $Q_1^{\mathrm{T}}Q_2 = 0$, two terms in the right-hand side of $A = L_{21}Q_1^{\mathrm{T}} + L_{22}Q_2^{\mathrm{T}}$ are orthogonal. From $B = L_{11}Q_1^{\mathrm{T}}$ with B full row rank, we see that L_{11} is nonsingular and Q_1^{T} forms a basis of the space spanned by the row vectors of B. It therefore follows that $\hat{E}\{A \mid B\} = L_{21}Q_1^{\mathrm{T}} = L_{21}L_{11}^{-1}B$. Also, from $AQ_1 = L_{21}$,

we get $\hat{E}\{A \mid B\} = A(Q_1 Q_1^{\mathrm{T}})$. Since $L_{22}Q_2^{\mathrm{T}}$ is orthogonal to the row space of B, it follows that $\hat{E}\{A \mid B^{\perp}\} = L_{22}Q_2^{\mathrm{T}}$.

6.5 Let $D = \begin{bmatrix} B \\ C \end{bmatrix}$. Then, D has full row rank. Thus from Problem 6.3,

$$\hat{E}\{A \mid D\} = A[B^{\mathrm{T}} \;\; C^{\mathrm{T}}] \begin{bmatrix} BB^{\mathrm{T}} & BC^{\mathrm{T}} \\ CB^{\mathrm{T}} & CC^{\mathrm{T}} \end{bmatrix}^{-1} \begin{bmatrix} B \\ C \end{bmatrix}$$

We see that the above equation is expressed as

$$\begin{aligned} \hat{E}\{A \mid D\} = {} & A[B^{\mathrm{T}} \;\; C^{\mathrm{T}}] \begin{bmatrix} BB^{\mathrm{T}} & BC^{\mathrm{T}} \\ CB^{\mathrm{T}} & CC^{\mathrm{T}} \end{bmatrix}^{-1} \begin{bmatrix} B \\ 0 \end{bmatrix} \\ & + A[B^{\mathrm{T}} \;\; C^{\mathrm{T}}] \begin{bmatrix} BB^{\mathrm{T}} & BC^{\mathrm{T}} \\ CB^{\mathrm{T}} & CC^{\mathrm{T}} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ C \end{bmatrix} \end{aligned}$$

Since $\mathrm{span}\{B\} \cap \mathrm{span}\{C\} = \{0\}$, the first term of the right-hand side of the above equation is the oblique projection of the row vectors of A onto the space spanned by the row vectors of B along the row vectors of C. Thus we have

$$\hat{E}_{\parallel C}\{A \mid B\} = A[B^{\mathrm{T}} \;\; C^{\mathrm{T}}] \begin{bmatrix} BB^{\mathrm{T}} & BC^{\mathrm{T}} \\ CB^{\mathrm{T}} & CC^{\mathrm{T}} \end{bmatrix}^{-1} \begin{bmatrix} B \\ 0 \end{bmatrix}$$

6.6 Note that $R_{22} = \begin{bmatrix} L_{22} & 0 \\ L_{32} & L_{33} \end{bmatrix}$ and $R_{32} = [L_{42} \;\; L_{43}]$. Let $\begin{bmatrix} \eta \\ \xi \end{bmatrix} \in \mathrm{Ker}(R_{22})$. Then, $L_{22}\eta = 0$ and $L_{32}\eta + L_{33}\xi = 0$ hold. However, since L_{22} is nonsingular, we have $\eta = 0$, so that $L_{33}\xi = 0$. Thus it suffices to show that $L_{33}\xi = 0$ implies $L_{43}\xi = 0$. Consider the following vectors

$$\begin{bmatrix} L_{13} \\ L_{23} \\ L_{33} \\ L_{43} \end{bmatrix} \xi = \begin{bmatrix} 0 \\ 0 \\ L_{33}\xi \\ L_{43}\xi \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ L_{43}\xi \end{bmatrix}$$

Lemmas 6.4 and 6.5 show that the above vector is also an input-output pair. However, since the past input-output and future inputs are zero, the future outputs must be zero, implying that $L_{43}\xi = 0$. This completes the proof.

Chapter 7

7.1 Let $Z(z) = B(z)/A(z)$. Let $z = e^{j\omega}$. Then, we have

$$Z(e^{j\omega}) = \frac{B(e^{j\omega})}{A(e^{j\omega})} = \frac{c(\omega) + jd(\omega)}{a(\omega) + jb(\omega)}$$

It thus follows that

$$\mathfrak{Re} Z(e^{j\omega}) = \frac{a(\omega)c(\omega) + b(\omega)d(\omega)}{a(\omega)^2 + b(\omega)^2} \geq 0, \qquad -\pi < \omega \leq \pi$$

Hence $Z(z)$ is positive real, if $A(z)$ is stable and

$$a(\omega)c(\omega) + b(\omega)d(\omega) \geq 0, \qquad -\pi < \omega \leq \pi \tag{E.7}$$

From the given first-order transfer function, we have

$$Z(e^{j\omega}) = \frac{c + b\cos\omega + jb\sin\omega}{a + \cos\omega + j\sin\omega}$$

Thus from (E.7), the positivity is satisfied if $z + a$ is stable and if

$$ac + b + (ab + c)\cos\omega \geq 0, \qquad -\pi < \omega \leq \pi$$

It therefore follows that $|ab + c| \leq ac + b$ and $ac + b > 0$. Hence, we have

$$|a| < 1, \qquad |c| \leq b, \qquad b \geq 0$$

7.2 It can be shown that

$$\begin{aligned} \mathfrak{Re}[A(e^{j\omega})] &= 1 + a_1\cos\omega + a_2\cos 2\omega \\ &= 2a_2\cos^2\omega + a_1\cos\omega - a_2 + 1 \end{aligned} \tag{E.8}$$

For $a_2 = 0$, we see that the positive real condition is reduced to

$$a_1\cos\omega + 1 \geq 0, \qquad -\pi < \omega \leq \pi$$

This is satisfied if and only if $-1 \leq a_1 \leq 1$.

In the following, we assume that $a_2 \neq 0$, and define

$$f(x) := 2x^2 + (a_1/a_2)x + 1/a_2 - 1, \quad -1 \leq x \leq 1$$

1. Suppose that $a_2 < 0$. Then, from (E.8), the positive real condition becomes

$$f(x) \leq 0, \qquad -1 \leq x \leq 1 \tag{E.9}$$

Since $f(0) = 1/a_2 - 1 < 0$, (E.9) is satisfied if and only if $f(-1) \leq 0$ and $f(1) \leq 0$. Thus we have

$$a_2 + a_1 + 1 \geq 0, \qquad a_2 - a_1 + 1 \geq 0, \qquad a_2 \leq 0$$

2. Suppose that $a_2 > 0$. In this case, the positive real condition becomes

$$f(x) \geq 0, \qquad -1 \leq x \leq 1$$

Let $x_1 = -a_1/4a_2$. According to the location of x_1, we have three cases:

a) If $x_1 \leq -1$, then $f(-1) \geq 0$. This implies that

$$0 < a_2 \leq a_1/4, \qquad a_2 \geq a_1 - 1$$

b) If $-1 \le x_1 \le 1$, then $f(x_1) \ge 0$, so that

$$a_2 \ge \frac{a_1}{4}, \qquad a_2 \ge -\frac{a_1}{4}, \qquad \frac{a_1^2}{2} + 4\left(a_2 - \frac{1}{2}\right)^2 \le 1 \tag{E.10}$$

c) If $x_1 \ge 1$, then $f(1) \ge 0$. Hence, we have

$$0 < a_2 \le -a_1/4, \qquad a_2 \ge -a_1 - 1,$$

Thus, the region $D = \{(a_1, a_2) \mid \mathfrak{Re}[A(e^{j\omega})] \ge 0\}$ is a convex set enclosed by the two lines $a_2 = \pm a_1 - 1$ and a portion of the ellipsoid in (E.10) [see Figure E.1].

$$a_2 \ge a_1 - 1, \qquad a_2 \ge -a_1 - 1, \qquad \frac{a_1^2}{2} + 4\left(a_2 - \frac{1}{2}\right)^2 \le 1 \tag{E.11}$$

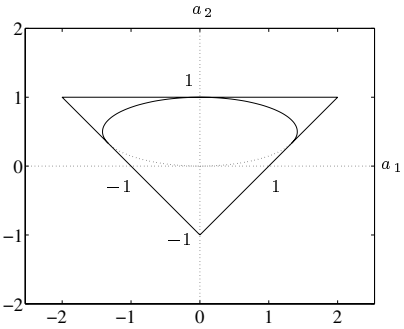

Figure E.1. Region of positive realness in (a_1, a_2)-plane

7.3 It is easy to see that $Z(z)$ is positive real if and only if

$$f(x) := 1 - a_1^2 - a_2^2 + 2a_1a_2x \ge 0, \qquad -1 \le x \le 1$$

Thus the condition is given by

$$|a_1 - a_2| \le 1, \qquad |a_1 + a_2| \le 1 \tag{E.12}$$

Remark E.1. It will be instructive to compare the positive real conditions (E.11) and (E.12) above and the stability condition (E.1). □

7.4 Using the Frobenius norm, we have

$$\begin{aligned}
\left\| \begin{bmatrix} A & B \\ C & D \end{bmatrix} \right\|_F^2 &= \text{trace}\left(\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{\mathrm{T}} \right) \\
&= \text{trace}(AA^{\mathrm{T}} + BB^{\mathrm{T}} + CC^{\mathrm{T}} + DD^{\mathrm{T}}) \\
&= \|A\|_F^2 + \|B\|_F^2 + \|C\|_F^2 + \|D\|_F^2 \\
&\le (\|A\|_F + \|B\|_F + \|C\|_F + \|D\|_F)^2
\end{aligned}$$

Taking the square root of the above relation, we get the desired result.

For the 2-norm, we define $X = \begin{bmatrix} A \\ C \end{bmatrix}$, $Y = \begin{bmatrix} B \\ D \end{bmatrix}$. By the definition of 2-norm,

$$\begin{aligned} \|X\|_2^2 &= \overline{\sigma}(X)^2 = \max \lambda \left(A^{\mathrm{T}} A + C^{\mathrm{T}} C \right) \\ &\leq \max \lambda (A^{\mathrm{T}} A) + \max \lambda (C^{\mathrm{T}} C) = \|A\|_2^2 + \|C\|_2^2 \leq (\|A\|_2 + \|C\|_2)^2 \end{aligned}$$

Thus we get $\|X\|_2 \leq \|A\|_2 + \|C\|_2$. Similarly, we get $\|Y\|_2 \leq \|B\|_2 + \|D\|_2$. Thus combining these results,

$$\begin{aligned} \|[X \;\; Y]\|_2^2 &= \max \lambda (X X^{\mathrm{T}} + Y Y^{\mathrm{T}}) \\ &\leq \max \lambda (X X^{\mathrm{T}}) + \max \lambda (Y Y^{\mathrm{T}}) \\ &= \|X\|_2^2 + \|Y\|_2^2 \leq (\|X\|_2 + \|Y\|_2)^2 \end{aligned}$$

Hence we have

$$\|[X \;\; Y]\|_2 \leq \|X\|_2 + \|Y\|_2 \leq \|A\|_2 + \|B\|_2 + \|C\|_2 + \|D\|_2$$

7.5 $M(\Pi)$ is easily derived. Let $\Pi = 3$. Since $M(\Pi) = \begin{bmatrix} 8/3 & 0 \\ 0 & 0 \end{bmatrix} \geq 0$, we see that $\Pi = 3$ satisfies the LMI. Now suppose that $\Pi < 3$. It then follows from (7.35) that

$$\frac{8}{9}\Pi - \frac{1}{3}\left(1 - \frac{1}{3}\Pi\right) \geq 0 \quad \Rightarrow \quad \Pi \geq 1/3$$

Hence we have $\Pi_* = 1/3$ and $\Pi^* = 3$, implying that the solutions of LMI satisfy $1/3 \leq \Pi \leq 3$. Note that in this case $F := A - \bar{C}^{\mathrm{T}} \Lambda^{-1}(0) C = 0$; see (5.91).

7.6 By the definition of $\mathcal{C}_k$ and $T_-(k)$,

$$\begin{aligned} \Omega_{k+1} &= \mathcal{C}_{k+1} T_-^{-1}(k+1) \mathcal{C}_{k+1}^{\mathrm{T}} \\ &= [\bar{C}^{\mathrm{T}} \;\; A\mathcal{C}_k] \begin{bmatrix} \Lambda(0) & C\mathcal{C}_k \\ \mathcal{C}_k^{\mathrm{T}} C^{\mathrm{T}} & T_-(k) \end{bmatrix}^{-1} \begin{bmatrix} \bar{C} \\ \mathcal{C}_k^{\mathrm{T}} A^{\mathrm{T}} \end{bmatrix} \end{aligned}$$

Note that this equation has the same form as (7.59). It is easy to see that Ω_k satisfies (7.62) by the following correspondence in (7.60).

$$\bar{\Omega}_k \;\leftrightarrow\; \Omega_k, \qquad A \;\leftrightarrow\; A^{\mathrm{T}}, \qquad \bar{C} \;\leftrightarrow C$$

7.7 First from (7.64), we note that

$$K(\Lambda(0) - C\Pi C^{\mathrm{T}}) + A\Pi C^{\mathrm{T}} = \bar{C}^{\mathrm{T}} \tag{E.13}$$

Substituting $A = A_K + KC$ into (7.63) yields

$$\begin{aligned} \Pi &= (A_K + KC)\Pi(A_K + KC)^{\mathrm{T}} + K(\Lambda(0) - C\Pi C^{\mathrm{T}})K^{\mathrm{T}} \\ &= A_K \Pi A_K^{\mathrm{T}} + KC\Pi C^{\mathrm{T}} K^{\mathrm{T}} + KC\Pi A_K^{\mathrm{T}} + A_K \Pi C^{\mathrm{T}} K^{\mathrm{T}} \\ &\quad + K(\Lambda(0) - C\Pi C^{\mathrm{T}})K^{\mathrm{T}} \end{aligned}$$

Again by using $A_K = A - KC$, it follows that

$$\begin{aligned}\varPi &= A_K \varPi A_K^{\mathrm{T}} + KC\varPi C^{\mathrm{T}}K^{\mathrm{T}} + KC\varPi(A-KC)^{\mathrm{T}} + (A-KC)\varPi C^{\mathrm{T}}K^{\mathrm{T}} \\ &\quad + K(\varLambda(0) - C\varPi C^{\mathrm{T}})K^{\mathrm{T}} \\ &= A_K \varPi A_K^{\mathrm{T}} + KC\varPi A^{\mathrm{T}} + A\varPi C^{\mathrm{T}}K^{\mathrm{T}} - KC\varPi C^{\mathrm{T}}K^{\mathrm{T}} \\ &\quad + K(\varLambda(0) - C\varPi C^{\mathrm{T}})K^{\mathrm{T}} + (K\varLambda(0)K^{\mathrm{T}} - K\varLambda(0)K^{\mathrm{T}})\end{aligned}$$

Using (E.13) in the right-hand side of the above equation, we readily obtain the desired result, *i.e.*, (7.65).

7.8 In view of Subsection 7.4.2, the constraint is given by $\xi = \mathcal{O}^{\mathrm{T}}\boldsymbol{u}$, so that the Lagrangian becomes

$$\mathcal{L} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} + \lambda^{\mathrm{T}}(\xi - \mathcal{O}^{\mathrm{T}}\boldsymbol{u})$$

Differentiating $\mathcal{L}$ with respect to $\boldsymbol{u}$ yields $2\boldsymbol{u} - \mathcal{O}\lambda = 0$. Thus, from the constraint, we have

$$\xi - \mathcal{O}^{\mathrm{T}}\mathcal{O}\lambda/2 = 0 \quad \Rightarrow \quad \lambda = 2(\mathcal{O}^{\mathrm{T}}\mathcal{O})^{-1}\xi$$

so that $\boldsymbol{u} = \mathcal{O}(\mathcal{O}^{\mathrm{T}}\mathcal{O})^{-1}\xi$ holds. Hence we have $\min_{\boldsymbol{u}}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}) = \xi^{\mathrm{T}}(\mathcal{O}^{\mathrm{T}}\mathcal{O})^{-1}\xi$.

Chapter 8

8.1 It is easy to show that

$$\begin{bmatrix} L^{\mathrm{T}} & 0 \\ 0 & M^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \varSigma_{xx} & \varSigma_{xy} \\ \varSigma_{yx} & \varSigma_{yy} \end{bmatrix} \begin{bmatrix} L & 0 \\ 0 & M \end{bmatrix} = \begin{bmatrix} L^{\mathrm{T}}\varSigma_{xx}L & L^{\mathrm{T}}\varSigma_{xy}M \\ M^{\mathrm{T}}\varSigma_{yx}L & M^{\mathrm{T}}\varSigma_{yy}M \end{bmatrix}$$

Thus from (8.9) and (8.10), the result follows. Also, the computation of the determinant is immediate.

8.2 Though this can be proved by using the orthogonality condition $a - Kb \perp b$, we give a different proof. See also Problem 6.3.

Since $I := \|a - Kb\|_{\mathcal{H}}^2 = \mathrm{trace}E\{(a-Kb)(a-Kb)^{\mathrm{T}}\}$,

$$I = \mathrm{trace}\Big(E\{aa^{\mathrm{T}}\} - E\{ab^{\mathrm{T}}\}K^{\mathrm{T}} - KE\{ba^{\mathrm{T}}\} + KE\{bb^{\mathrm{T}}\}K^{\mathrm{T}}\Big)$$

We see that the right-hand side is a quadratic form in $K = (k_{ij})$.

Recall the formulas for the differentiation of trace (*e.g.* see [185]):

$$\frac{\partial}{\partial X}\mathrm{trace}(AX) = A^{\mathrm{T}}, \qquad \frac{\partial}{\partial X}\mathrm{trace}(AX^{\mathrm{T}}) = A$$
$$\frac{\partial}{\partial X}\mathrm{trace}(AXBX^{\mathrm{T}}) = A^{\mathrm{T}}XB^{\mathrm{T}} + AXB$$

Thus it follows that

$$\frac{\partial I}{\partial K} = -2E\{ab^{\mathrm{T}}\} + 2KE\{bb^{\mathrm{T}}\} = 0 \quad \Rightarrow \quad K = E\{ab^{\mathrm{T}}\}(E\{bb^{\mathrm{T}}\})^{-1}$$

8.3 Applying Lemma 4.11 to (8.48), we get

$$\begin{aligned}\Lambda_{yy}(l) &= CA^l\Sigma C^{\mathrm{T}} + CA^{l-1}KR \\ &= CA^{l-1}(A\Sigma C^{\mathrm{T}} + KR) = CA^{l-1}\bar{C}^{\mathrm{T}}\end{aligned}$$

for $l = 1, 2, \cdots$. For $l = 0$, we have $\Lambda_{yy}(0) = C\Sigma C^{\mathrm{T}} + R$, so that we have the desired result. We can show that Theorem 8.5 and Lemma 8.5 give the same result.

8.4 Since y is scalar, we note that $T_+ = T_-$, $L = M$, $H^T = H$ hold. Thus $\bar{H} := L^{-1}HL^{-\mathrm{T}} = U\Sigma V^{\mathrm{T}}$ is symmetric, so that $\bar{H} = U\Sigma V^{\mathrm{T}} = V\Sigma U^{\mathrm{T}}$. Since $\mathrm{Im}(\bar{H}) = \mathrm{Im}(U) = \mathrm{Im}(V)$ holds, there exists a nonsingular matrix $S \in \mathbb{R}^{n\times n}$ such that $U = VS$. Since $I_n = U^{\mathrm{T}}U = S^{\mathrm{T}}V^{\mathrm{T}}VS = S^{\mathrm{T}}S$, we see that $S = V^{\mathrm{T}}U$ is orthogonal. From $U\Sigma V^{\mathrm{T}} = V\Sigma U^{\mathrm{T}}$, we have $\Sigma S^{\mathrm{T}} = S\Sigma$, so that similarly to the proof of Lemma 5.2, we can show that $S = \mathrm{diag}(\pm 1, \cdots, \pm)$ holds. By using $\mathcal{O} = LU\Sigma^{1/2}$, $U = VS$, $\Sigma = S\Sigma S$, we see that

$$\mathcal{C} = \Sigma^{1/2}V^{\mathrm{T}}M^{\mathrm{T}} = \Sigma^{1/2}SU^{\mathrm{T}}L^{\mathrm{T}} = S\Sigma^{1/2}U^{\mathrm{T}}L^{\mathrm{T}} = S\mathcal{O}^{\mathrm{T}}$$

holds, where we used the fact that S and $\Sigma^{1/2}$ are diagonal. Thus,

$$\mathcal{C}^{\leftarrow} = S(\mathcal{O}^{\uparrow})^{\mathrm{T}}, \quad \mathcal{C}^{\dagger} = (\mathcal{O}^{\dagger})^{\mathrm{T}}S$$

Hence, from (8.50), we get

$$A = \mathcal{C}^{\leftarrow}\mathcal{C}^{\dagger} = S(\mathcal{O}^{\uparrow})^{\mathrm{T}}(\mathcal{O}^{\dagger})^{\mathrm{T}}S = S(\mathcal{O}^{\dagger}\mathcal{O}^{\uparrow})^{\mathrm{T}}S = SA^{\mathrm{T}}S$$

Also, from (8.52),

$$\bar{C}^{\mathrm{T}} = \mathcal{C}(1:n, 1:p) = SC^{\mathrm{T}}$$

implying that $\bar{C} = CS$ holds.

8.5 Since $H = L_1U_1\Sigma V_1^{\mathrm{T}}M_1^{\mathrm{T}} = L_2U_2\Sigma V_2^{\mathrm{T}}M_2^{\mathrm{T}}$, and since $\mathrm{Im}(L_1U_1) = \mathrm{Im}(L_2U_2)$, there exists a nonsingular $S \in \mathbb{R}^{n\times n}$ such that $L_2U_2 = L_1U_1S$. Note that $L_1L_1^{\mathrm{T}} = L_2L_2^{\mathrm{T}}$ holds. Thus $Z = L_1^{-1}L_2$ becomes an orthogonal matrix with $ZZ^{\mathrm{T}} = Z^{\mathrm{T}}Z = I$. This implies that ZU_2 becomes an orthogonal matrix, and hence $S = U_1^{\mathrm{T}}(ZU_2)$ becomes an orthogonal matrix. Again, using $L_1U_1\Sigma V_1^{\mathrm{T}}M_1^{\mathrm{T}} = L_2U_2\Sigma V_2^{\mathrm{T}}M_2^{\mathrm{T}}$, and noting that $M_1M_1^{\mathrm{T}} = M_2M_2^{\mathrm{T}}$, it follows that

$$U_1\Sigma V_1^{\mathrm{T}} = L_1^{-1}L_2U_2\Sigma V_2^{\mathrm{T}}M_2^{\mathrm{T}}M_1^{-\mathrm{T}}$$

so that

$$U_1\Sigma^2U_1^{\mathrm{T}} = L_1^{-1}L_2U_2\Sigma^2U_2^{\mathrm{T}}L_2^{\mathrm{T}}L_1^{-\mathrm{T}} = U_1S\Sigma^2S^{\mathrm{T}}U_1^{\mathrm{T}}$$

and hence

$$\Sigma^2 = S\Sigma^2S^{\mathrm{T}} \tag{E.14}$$

It should be noted that Σ^2 is a diagonal matrix with different elements and that S is orthogonal. Thus similarly to Lemma 5.2, we have $S = (\pm 1, \cdots, \pm 1)$. In fact, suppose that

$$S = \begin{bmatrix} S_{n-1} & a \\ b^{\mathrm{T}} & c \end{bmatrix}, \quad S_{n-1} \in \mathbb{R}^{(n-1)\times(n-1)}, \quad a,\, b \in \mathbb{R}^{n-1}, \quad c \in \mathbb{R}$$

Then, from $S^{\mathrm{T}}S = I$, we have $\|a\|^2 + c^2 = 1$ and from (E.14)

$$\Sigma_{n-1}^2 S_{n-1}^{\mathrm{T}} = S_{n-1}\Sigma_{n-1}^2, \qquad \Sigma_{n-1}^2 a = \sigma_n^2 a, \qquad b^{\mathrm{T}}\Sigma_{n-1} = \sigma_n^2 b^{\mathrm{T}}$$

Since σ_n^2 is not an eigenvalue of Σ_{n-1}^2, we see that $a = 0$, $b = 0$, so that $c^2 = 1$.

By using $L_2U_2 = L_1U_1S$ and $U_1\Sigma V_1^{\mathrm{T}} = L_1^{-1}L_2U_2\Sigma V_2^{\mathrm{T}}M_2^{\mathrm{T}}M_1^{-\mathrm{T}}$,

$$\Sigma = S\Sigma V_2^{\mathrm{T}}M_2^{\mathrm{T}}M_1^{-\mathrm{T}}V_1 = \Sigma S V_2^{\mathrm{T}}M_2^{\mathrm{T}}M_1^{-\mathrm{T}}V_1 \quad \Rightarrow \quad V_1^{\mathrm{T}}M_1^{-1}M_2V_2S = I_n$$

where we used the fact that $\det \Sigma \neq 0$. Since the right-inverse of $V_1^{\mathrm{T}}M_1^{-1}$ is M_1V_1, we get $M_2V_2S = M_1V_1$, so that $M_2V_2 = M_1V_1S$. Also, from (8.41),

$$\begin{aligned} \mathcal{O}_2 &= L_2U_2\Sigma^{1/2} = L_1U_1S\Sigma^{1/2} = \mathcal{O}_1S \\ \mathcal{C}_2 &= \Sigma^{1/2}V_2^{\mathrm{T}}M_2^{\mathrm{T}} = \Sigma^{1/2}SV_1^{\mathrm{T}}M_1^{\mathrm{T}} = S\mathcal{C}_1 \end{aligned}$$

It thus follows from (8.50) that

$$A_2 = \mathcal{C}_2^{\leftarrow}\mathcal{C}_2^{\dagger} = S\mathcal{C}_1^{\leftarrow}\mathcal{C}_1^{\dagger}S = SA_1S$$

Moreover, from (8.51) and (8.52),

$$\begin{aligned} C_2 &= \mathcal{O}_2(1:p, 1:n) = \mathcal{O}_1(1:p, 1:n)S = C_1S \\ \bar{C}_2 &= \mathcal{C}_2(1:p, 1:n)^{\mathrm{T}} = \mathcal{C}_1(1:p, 1:n)^{\mathrm{T}}S = \bar{C}_1S \end{aligned}$$

From (8.53) and (8.54), we have

$$\begin{aligned} R_2 &= \Lambda(0) - C_2\Sigma C_2^{\mathrm{T}} = \Lambda(0) - C_1S\Sigma SC_1^{\mathrm{T}} = \Lambda(0) - C_1\Sigma C_1^{\mathrm{T}} = R_1 \\ K_2 &= (\bar{C}_2^{\mathrm{T}} - A_2\Sigma C_2^{\mathrm{T}})R_2^{-1} = (S\bar{C}_1^{\mathrm{T}} - SA_1S\Sigma SC_1^{\mathrm{T}})R_1^{-1} \\ &= S(\bar{C}_1^{\mathrm{T}} - A_1\Sigma C_1^{\mathrm{T}})R_1^{-1} = SK_1 \end{aligned}$$

This completes the proof.

Glossary

Notation

$\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}$	real numbers, complex numbers, integers
$\mathbb{R}^n$, $\mathbb{C}^n$	n-dimensional real vectors, complex vectors
$\mathbb{R}^{m\times n}$	$(m \times n)$-dimensional real matrices
$\mathbb{C}^{m\times n}$	$(m \times n)$-dimensional complex matrices
$\dim(x)$	dimension of vector x
$\dim(\mathcal{V})$	dimension of subspace $\mathcal{V}$
$\mathcal{V} \vee \mathcal{W}$	vector sum of subspaces $\mathcal{V}$ and $\mathcal{W}$
$\mathcal{V} + \mathcal{W}$	direct sum of subspaces $\mathcal{V}$ and $\mathcal{W}$
$\mathrm{span}\{v, w, x\}$	subspace generated by vectors v, w, x
A^{T}, A^{H}	transpose of $A \in \mathbb{R}^{m\times n}$, conjugate transpose of $A \in \mathbb{C}^{m\times n}$
A^{-1}, $A^{-\mathrm{T}}$	inverse and transpose of the inverse of A
$A^{\dagger}$	pseudo-inverse of A
$A \geq 0$	symmetric, nonnegative definite
$A > 0$	symmetric, positive definite
$A^{1/2}$	square root of A
$\det(A)$	determinant of A
$\mathrm{trace}(A)$	trace of A
$\mathrm{rank}(A)$	rank of A
$\lambda(A)$, $\lambda_i(A)$	eigenvalue, ith eigenvalue of A
$\rho(A)$	spectral radius, *i.e.*, $\max_i \lvert\lambda_i(A)\rvert$
$\sigma(A)$, $\sigma_i(A)$	singular value, ith singular value of A
$\underline{\sigma}(A)$, $\overline{\sigma}(A)$	minimum singular value, maximum singular value of A
$\mathrm{Im}(A)$	image (or range) of A
$\mathrm{Ker}(A)$	kernel (or null space) of A
$\lVert x\rVert_2$, $\lVert x\rVert_\infty$	2-norm, ∞-norm of x
$\lVert A\rVert_2$, $\lVert A\rVert_F$	2-norm, Frobenius norm of A
$\left[\begin{array}{c\vert c} A & B \\ \hline C & D \end{array}\right]$	transfer matrix $G(z) = D + C(zI - A)^{-1}B$

$E\{x\}$	mathematical expectation of random vector x
$\mathrm{cov}\{x, y\}$	(cross-) covariance matrix of random vectors x and y
$\mathcal{N}(\mu, \Sigma)$	Gaussian (normal) distribution with mean μ and covariance matrix Σ
$E\{x \mid y\}$	conditional expectation of x given y
$(x, y)_{\mathcal{H}}$	inner product of x and y in Hilbert space $\mathcal{H}$
$\|x\|_{\mathcal{H}}$	norm of x in Hilbert space $\mathcal{H}$
$\overline{\mathrm{span}}\{\cdots\}$	closed Hilbert subspace generated by infinite elements $\{\cdots\}$
$\hat{E}\{x \mid \mathcal{Y}\}$	orthogonal projection of x onto subspace $\mathcal{Y}$
$\hat{E}_{\parallel \mathcal{Z}}\{x \mid \mathcal{Y}\}$	oblique projection of x onto $\mathcal{Y}$ along $\mathcal{Z}$
$a := b$	a is defined by b
$a =: b$	b is defined by a
$\mathfrak{Z}$	z-transform operator
z	complex variable, shift operator $zf(t) := f(t+1)$
$\mathfrak{Re}$	real part
$\mathrm{Ric}(\cdot)$	Riccati operator; (7.34)

Abbreviations

AIC	Akaike Information Criterion; see Section 1.1
AR	AutoRegressive; (4.33)
ARMA	AutoRegressive Moving Average; (4.34)
ARMAX	AutoRegressive Moving Average with eXogenous input; (1.4)
ARX	AutoRegressive with eXogenous input; (A.7)
ARE	Algebraic Riccati Equation; (5.67)
ARI	Algebraic Riccati Inequality; (7.35)
BIBO	Bounded-Input, Bounded-Output; see Section 3.2
CCA	Canonical Correlation Analysis; see Section 8.1
CVA	Canonical Variate Analysis; see Section 10.8
FIR	Finite Impulse Response; (A.12)
IV	Instrumental Variable; see Section A.1
LMI	Linear Matrix Inequality; see (7.26)
LTI	Linear Time-Invariant; see Section 3.2
MA	Moving Average; (4.44)
MIMO	Multi-Input, Multi-Output; see Section 1.3
ML	Maximum Likelihood; see Section 1.1
MOESP	Multivariable Output Error State sPace; see Section 6.5
N4SID	Numerical algorithms for Subspace State Space System IDentification; see Section 6.6
ORT	ORThogonal decomposition based; see Section 9.7
PE	Persistently Exciting; see Sections 6.3 and Appendix B
PEM	Prediction Error Method; see Sections 1.2 and 1.3
PO-MOESP	Past Output MOESP; see Section 6.6
SISO	Single-Input, Single-Output; see Section 3.2
SVD	Singular Value Decomposition; see (2.26)

References

1. H. Akaike, "Stochastic theory of minimal realization," *IEEE Trans. Automatic Control*, vol. AC-19, no. 6, pp. 667–674, 1974.
2. H. Akaike, "Markovian representation of stochastic processes by canonical variables," *SIAM J. Control*, vol. 13, no. 1, pp. 162–173, 1975.
3. H. Akaike, "Canonical correlation analysis of time series and the use of an information criterion," In *System Identification: Advances and Case Studies* (R. Mehra and D. Lainiotis, eds.), Academic, pp. 27–96, 1976.
4. H. Akaike, "Comments on 'On model structure testing in system identification'," *Int. J. Control*, vol. 27, no. 2, pp. 323–324, 1978.
5. U. M. Al-Saggaf and G. F. Franklin, "An error bound for a discrete reduced order model of a linear multivariable system," *IEEE Trans. Automatic Control*, vol. 32, no. 9, pp. 815–819, 1987.
6. U. M. Al-Saggaf and G. F. Franklin, "Model reduction via balanced realizations: An extension and frequency weighting techniques," *IEEE Trans. Automatic Control*, vol. 33, no. 7, pp. 687–692, 1988.
7. B. D. O. Anderson, "A system theory criterion for positive real matrices," *SIAM J. Control*, vol. 5, no. 2, pp. 171–182, 1967.
8. B. D. O. Anderson, "An algebraic solution to the spectral factorization problem," *IEEE Trans. Automatic Control*, vol. AC-12, no. 4, pp. 410–414, 1967.
9. B. D. O. Anderson, "The inverse problem of stationary covariance generation," *J. Statistical Physics*, vol. 1, no. 1, pp. 133–147, 1969.
10. B. D. O. Anderson, K. L. Hitz and N. D. Diem, "Recursive algorithm for spectral factorization," *IEEE Trans. Circuits Systems*, vol. CAS-21, no. 6, pp. 742–750, 1974.
11. B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, 1979.
12. B. D. O. Anderson and M. R. Gevers, "Identifiability of linear stochastic systems operating under linear feedback," *Automatica*, vol. 18, no. 2, pp. 195–213, 1982.
13. T. W. Anderson, *The Statistical Analysis of Time Series*, Wiley, 1971.
14. T. W. Anderson, *An Introduction to Multivariable Statistical Analysis* (2nd ed.), Wiley, 1984.
15. M. Aoki, *State Space Modeling of Time Series* (2nd ed.), Springer, 1990.
16. M. Aoki and A. M. Havenner (eds.), *Applications of Computer Aided Time Series Modeling* (Lecture Notes in Statistics 119), Springer, 1997.
17. K. J. Åström and T. Bohlin, "Numerical identification of linear dynamic systems for normal operating records," *Proc. 2nd IFAC Symp. Theory of Self-Adaptive Systems*, Teddington, pp. 96–111, 1965.

18. K. S. Arun and S. Y. Kung, "Balanced approximation of stochastic systems," *SIAM J. Matrix Analysis and Applications*, vol. 11, no. 1, pp. 42–68, 1990.
19. D. Bauer and M. Jansson, "Analysis of the asymptotic properties of the MOESP type of subspace algorithms," *Automatica*, vol. 36, no. 4, pp. 497–509, 2000.
20. S. Bittanti, A. J. Laub and J. C. Willems (eds.), *The Riccati Equation*, Springer, 1991.
21. A. Björck and G. H. Golub, "Numerical methods for computing angles between linear subspaces," *Mathematics of Computation*, vol. 27, pp. 579–594, July 1973.
22. G. E. P. Box and G. M. Jenkins, *Time Series Analysis - Forecasting and Control*, Holden-Day, 1970.
23. R. S. Bucy, *Lectures on Discrete Time Filtering*, Springer, 1994.
24. P. E. Caines and C. W. Chan, "Feedback between stationary stochastic processes," *IEEE Trans. Automatic Control*, vol. AC-20, no. 4, pp. 498–508, 1975.
25. P. E. Caines and C. W. Chan, "Estimation, identification and feedback," In *System Identification: Advances and Case Studies* (R. Mehra and D. Lainiotis, eds.), Academic, pp. 349–405, 1976.
26. P. E. Caines, "Weak and strong feedback free processes," *IEEE Trans. Automatic Control*, vol. AC-21, no. 5, pp. 737–739, 1976.
27. C. T. Chen, *Linear System Theory and Design*, Holt-Saunders, 1984.
28. H. -F. Chen, P. K. Kumar and J. H. van Schuppen, "On Kalman filtering for conditionally Gaussian systems with random matrices," *Systems and Control Letters*, vol. 13, pp. 397–404, 1989.
29. A. Chiuso and G. Picci, "Subspace identification by orthogonal decomposition," *Proc. 14th IFAC World Congress*, Beijing, vol. H, pp. 241–246, 1999.
30. A. Chiuso and G. Picci, "Some algorithmic aspects of subspace identification with inputs," *Int. J. Applied Math. and Computer Science*, vol. 11, no. 1, pp. 55–75, 2001.
31. A. Chiuso and G. Picci, "On the ill-conditioning of subspace identification with inputs," *Automatica*, vol. 40, no. 4, pp. 575–589, 2004.
32. A. Chiuso and G. Picci, "Numerical conditioning and asymptotic variance of subspace estimates," *Automatica*, vol. 40, no. 4, pp. 677–683, 2004.
33. T. Chonavel, *Statistical Signal Processing*, Springer, 2002.
34. C. T. Chou and M. Verhaegen, "Closed-loop identification using canonical correlation analysis," *Proc. 5th European Control Conference*, Karlsruhe, F-162, 1999.
35. N. L. C. Chui and J. M. Maciejowski, "Realization of stable models with subspace methods," *Automatica*, vol. 32, no. 11, pp. 1587–1595, 1996.
36. R. V. Churchill and J. W. Brown, *Complex Variables and Applications* (4th ed.), McGraw-Hill, 1984.
37. B. Codrons, B. D. O. Anderson and M. Gevers, "Closed-loop identification with an unstable or nonminimum phase controller," *Automatica*, vol. 38, no. 12, pp. 2127–2137, 2002.
38. A. Dahlén, A. Lindquist and J. Mari, "Experimental evidence showing that stochastic subspace identification methods may fail," *Systems and Control Letters*, vol. 34, no. 2, pp. 302–312, 1998.
39. K. De Cock, *Principal Angles in System Theory, Information Theory and Signal Processing*, Ph.D. Thesis, Katholieke Universiteit Lueven, Belgium, 2002.
40. B. De Moor, "The singular value decomposition and long and short spaces of noisy matrices," *IEEE Trans. Signal Processing*, vol. 41, no. 9, pp. 2826–2838, 1993.
41. B. De Moor, M. Moonen, L. Vandenberghe and J. Vandewalle, "Identification of linear state space models with SVD using canonical correlation analysis," In *Singular Value Decomposition and Signal Processing* (E. Deprettere, ed.), North-Holland, pp. 161–169, 1988.

42. U. B. Desai and D. Pal, "A transformation approach to stochastic model reduction," *IEEE Trans. Automatic Control*, vol. AC-29, no. 12, pp. 1097–1100, 1984.
43. U. B. Desai, D. Pal and R. D. Kirkpatrick, "A realization approach to stochastic model reduction," *Int. J. Control*, vol. 42, no. 4, pp. 821–838, 1985.
44. J. L. Doob, *Stochastic Processes*, Wiley, 1953.
45. P. Faurre, "Réalisations markoviennes de processus stationnaires," Technical Report No. 13, INRIA, March 1973.
46. P. L. Faurre, "Stochastic realization algorithms," In *System Identification: Advances and Case Studies* (R. Mehra and D. Lainiotis, eds.), Academic, pp. 1–25, 1976.
47. P. Faurre, M. Clerget and F. Germain, *Opérateurs Rationnels Positifs*, Dunod, 1979.
48. U. Forssell and L. Ljung, "Closed-loop identification revisited," *Automatica*, vol. 35, no. 7, pp. 1215–1241, 1999.
49. U. Forssell and L. Ljung, "A projection method for closed-loop identification," *IEEE Trans. Automatic Control*, vol. 45, no. 11, pp. 2101–2106, 2000.
50. S. Fujishige, H. Nagai and Y. Sawaragi, "System-theoretic approach to model reduction and system-order determination," *Int. J. Control*, vol. 22, no. 6, pp. 807–819, 1975.
51. F. R. Gantmacher, *The Theory of Matrices* (2 vols.), Chelsea, 1959.
52. I. M. Gel'fand and A. M. Yaglom, "Calculation of the amount of information about a random function contained in another such function," *American Math. Soc. Transl.*, vol. 12, pp. 199–246, 1959.
53. M. Gevers and B. D. O. Anderson, "On joint stationary feedback-free stochastic processes," *IEEE Trans. Automatic Control*, vol. AC-27, no. 2, pp. 431–436, 1982.
54. M. Gevers and V. Wertz, "Uniquely identifiable state-space and ARMA parametrizations for multivariable linear systems," *Automatica*, vol. 20, no. 3, pp. 333–347, 1984.
55. M. Gevers, "A personal view on the development of system identification," *Proc. 13th IFAC Symp. System Identification*, Rotterdam, pp. 773–784, 2003.
56. E. G. Gilbert, "Controllability and observability in multivariable control systems," *SIAM J. Control*, vol. 1, no. 2, pp. 128–151, 1963.
57. M. Glover and J. C. Willems, "Parametrizations of linear dynamical systems: Canonical forms and identifiability," *IEEE Trans. Automatic Control*, vol. AC-19, no. 6, pp. 640–646, 1974.
58. I. Goethals, T. Van Gestel, J. Suykens, P. Van Dooren and B. De Moor, "Identification of positive real models in subspace identification by using regularization," *IEEE Trans. Automatic Control*, vol. 48, no. 10, pp. 1843–1847, 2003.
59. G. H. Golub and C. F. van Loan, *Matrix Computations* (3rd ed.), The Johns Hopkins University Press, 1996.
60. G. H. Golub and H. A. van der Vorst, "Eigenvalue computation in the 20th century," *J. Computational and Applied Mathematics*, vol.123, pp. 35–65, 2000.
61. G. C. Goodwin and R. L. Payne, *Dynamic System Identification: Experiment Design and Data Analysis*, Academic, 1977.
62. B. Gopinath, "On the identification of linear time-invariant systems from input-output data," *Bell Systems Technical Journal*, vol. 48, no. 5, pp. 1101–1113, 1969.
63. C. W. J. Granger, "Economic processes involving feedback," *Information and Control*, vol. 6, pp. 28–48, 1963.
64. M. Green, "All-pass matrices, the positive real lemma, and unit canonical correlations between future and past," *J. Multivariate Analysis*, vol. 24, pp. 143–154, 1988.
65. U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*, Chelsea, 1958.
66. M. S. Grewal and A. P. Andrews, *Kalman Filtering – Theory and Practice*, Prentice-Hall, 1993.

67. R. Guidorzi, "Canonical structures in the identification of multivariable systems," *Automatica*, vol. 11, no. 4, pp. 361–374, 1975.
68. E. J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*, Wiley, 1988.
69. E. J. Hannan and D. S. Poskit, "Unit canonical correlations between future and past," *Annals of Statistics*, vol. 16, pp. 784–790, 1988.
70. F. R. Hansen, G. F. Franklin and R. Kosut, "Closed-loop identification via the fractional representation: Experiment design," *Proc. 1989 American Control Conference*, Pittsburgh, pp. 1422–1427, 1989.
71. D. Hinrichsen and A. J. Prichard, "An improved error estimate for reduced-order models of discrete-time systems," *IEEE Trans. Automatic Control*, vol. AC-35, no. 3, pp. 317–320, 1990.
72. B. L. Ho and R. E. Kalman, "Effective construction of linear state-variable models from input/output functions," *Regelungstechnik*, vol. 14, no. 12, pp. 545–548, 1966.
73. R. A. Horn and C. A. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
74. H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educational Psychology*, vol. 24, pp. 417–441 and pp. 498–520, 1933.
75. H. Hotelling, "Relations between two sets of variates," *Biometrica*, vol. 28, pp. 321–377, 1936.
76. M. Jansson and B. Wahlberg, "On consistency of subspace methods for system identification," *Automatica*, vol. 34, no. 12, pp. 1507–1519, 1998.
77. E. A. Jonckheere and J. W. Helton, "Power spectrum reduction by optimal Hankel norm approximation of the phase of the outer spectral factor," *IEEE Trans. Automatic Control*, vol. AC-30, no. 12, pp. 1192–1201, 1985.
78. T. Kailath, "A view of three decades of linear filtering theory," *IEEE Trans. Information Theory*, vol. IT-20, no. 2, pp. 146–181, 1974.
79. T. Kailath, *Lectures on Linear Least-Squares Estimation*, Springer, 1976.
80. T. Kailath, *Linear Systems*, Prentice-Hall, 1980
81. R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, J. Basic Engineering*, vol. 82D, no.1, pp. 34–45, 1960.
82. R. E. Kalman, "Mathematical description of linear dynamical systems," *SIAM J. Control*, vol. 1, no. 2, pp. 152–192, 1963.
83. R. E. Kalman, "Algebraic aspects of the generalized inverse of a rectangular matrix," In *Generalized Inverses and Applications* (Z. Nashed, ed.), pp. 111–121, Academic, 1976.
84. R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Trans. ASME, J. Basic Engineering*, vol. 83D, no.1, pp. 95–108, 1961.
85. R. E. Kalman, P. L. Falb and M. A. Arbib, *Topics in Mathematical System Theory*, McGraw-Hill, 1969.
86. T. Katayama, "Subspace-based system identification – A view from realization theory," *J. Systems, Control and Information Engineers*, vol. 41, no. 9, pp. 380–387, 1997 (in Japanese).
87. T. Katayama, H. Kawauchi and G. Picci, "A subspace identification of deterministic part of state space model operating in closed loop," *Proc. 6th European Control Conference*, Porto, pp. 2505–2510, 2001.
88. T. Katayama, H. Kawauchi and G. Picci, "Subspace identification of closed loop systems by stochastic realization," *CD-ROM Preprints 15th IFAC World Congress*, Barcelona, Paper # T-Mo-M02-2, 2002.
89. T. Katayama, H. Kawauchi and G. Picci, "Subspace identification of closed loop systems by the orthogonal decomposition method," *Automatica*, vol. 41, no. 5, pp. 863–872, 2005.

90. T. Katayama and G. Picci, "Realization of stochastic systems with exogenous inputs and subspace identification methods," *Automatica*, vol. 35, no. 10, pp. 1635–1652, 1999.
91. T. Katayama and S. Sugimoto (eds.), *Statistical Methods in Control and Signal Processing*, Marcel Dekker, 1997.
92. T. Katayama, H. Tanaka and T. Enomoto, "A simple subspace identification method of closed-loop systems using orthogonal decomposition," *Preprints 16th IFAC World Congress*, Prague, July 2005.
93. H. Kawauchi, A. Chiuso, T. Katayama and G. Picci, "Comparison of two subspace identification methods for combined deterministic-stochastic systems," *Proc. 31st ISCIE Symp. Stochastic Systems Theory and Applications,* Yokohama, pp. 7–12, 1999.
94. V. Klema and A. J. Laub, "The singular value decomposition: Its computation and some applications," *IEEE Trans. Automatic Control*, vol. AC-25, no. 2, pp. 164–176, 1980.
95. L. H. Koopmans, *The Spectral Analysis of Time Series*, Academic, 1974.
96. T. C. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*, Wiley, 1950.
97. V. Kucèra, "The discrete Riccati equation of optimal control," *Kybernetika*, vol. 8, no. 5, pp. 430–447, 1972.
98. H. Kwakernaak and R. Sivan, *Modern Signals and Systems*, Prentice-Hall, 1991.
99. P. Lancaster and L. Rodman, *Algebraic Riccati Equations*, Oxford Science Publications, 1995.
100. W. E. Larimore, "System identification, reduced-order filtering and modeling via canonical variate analysis," *Proc. 1983 American Control Conference*, San Francisco, pp. 445–451, 1983.
101. W. E. Larimore, "Canonical variate analysis in identification, filtering, and adaptive control," *Proc. 29th IEEE Conference on Decision and Control*, Honolulu, pp. 596–604, 1990.
102. W. E. Larimore, "The ADAPTx software for automated and real-time multivariable system identification," *Proc. 13th IFAC Symp. System Identification*, Rotterdam, pp. 1496–1501, 2003.
103. A. J. Laub, "A Schur method for solving algebraic Riccati equations," *IEEE Trans. Automatic Control*, vol. AC-24, no. 6, pp. 913–921, 1979.
104. M. J. Levin, "Estimation of a system pulse transfer function in the presence of noise," *IEEE Trans. Automatic Control*, vol. 9, no. 2, pp. 229–235, 1964.
105. A. Lindquist and G. Picci, "A geometric approach to modelling and estimation of linear stochastic systems," *J. Math. Systems, Estimation and Control*, vol. 1, no. 3, pp. 241–333, 1991.
106. A. Lindquist and G. Picci, "Canonical correlation analysis, approximate covariance extension, and identification of stationary time series," *Automatica*, vol. 32, no. 5, pp. 709–733, 1996.
107. A. Lindquist and G. Picci, "Geometric methods for state space identification," In *Identification, Adaptation, Learning: The Science of Learning Models from Data* (S. Bittanti and G. Picci, eds.), Springer, pp. 1–69, 1996.
108. Y. Liu and B. D. O. Anderson, "Singular perturbation approximation of balanced systems," *Int. J. Control*, vol. 50, no. 4, pp. 1379–1405, 1989.
109. L. Ljung, *System Identification - Theory for The User* (2nd ed.), Prentice-Hall, 1999.
110. L. Ljung and T. McKelvey, "Subspace identification from closed loop data," *Signal Processing*, vol. 52, no. 2, pp. 209–216, 1996.
111. D. Luenberger, *Optimization by Vector Space Methods*, Wiley, 1969.
112. J. M. Maciejowski, "Parameter estimation of multivariable systems using balanced realizations," In *Identification, Adaptation, Learning: The Science of Learning Models from Data* (S. Bittanti and G. Picci, eds.), Springer, pp. 70–119, 1996.

113. T. McKelvey, H. Ackay and L. Ljung, "Subspace-based multivariable identification from frequency response data," *IEEE Trans. Automatic Control*, vol. AC-41, no. 7, pp. 960–979, 1996.
114. T. McKelvey, A. Helmersson and T. Ribarits, "Data driven local coordinates for multivariable linear systems and their application to system identification," *Automatica*, vol. 40, no. 9, pp. 1629–1635, 2004.
115. P. Masani, "The prediction theory of multivariable stochastic processes - III," *Acta Mathematica*, vol. 104, pp. 141–162, 1960.
116. R. K. Mehra and C. T. Leondes (eds.), *System Identification - Advances and Case Studies*, Academic, 1976.
117. B. P. Molinari, "The stabilizing solution of the discrete algebraic Riccati equation," *IEEE Trans. Automatic Control*, vol. AC-20, no. 3, pp. 396–399, 1975.
118. M. Moonen, B. De Moor, L. Vandenberghe and J. Vandewalle, "On- and off-line identification of linear state-space models," *Int. J. Control*, vol. 49, no. 1, pp. 219–232, 1989.
119. M. Moonen and J. Vandewalle, "QSVD approach to on- and off-line state-space identification," *Int. J. Control*, vol. 51, no. 5, pp. 1133–1146, 1990.
120. A. Ohsumi, K. Kameyama and K. Yamaguchi, "Subspace identification for continuous-time stochastic systems via distribution-based approach," *Automatica*, vol. 38, no. 1, pp. 63–79, 2002.
121. A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, Prentice-Hall, 1975.
122. C. C. Paige, "Properties of numerical algorithms related to computing controllability," *IEEE Trans. Automatic Control*, vol. AC-26, no. 1, pp. 130–138, 1981.
123. A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes* (4th ed.), McGraw-Hill, 2002.
124. T. Pappas, A. J. Laub and N. R. Sandell, Jr., "On the numerical solution of the discrete time algebraic Riccati equation," *IEEE Trans. Automatic Control*, vol. AC-25, no. 4, pp. 631–641, 1980.
125. R. Patel, A. J. Laub and P. M. Van Dooren (eds.), *Numerical Linear Algebra Techniques for Systems and Control*, IEEE Press, 1994.
126. M. Pavon, "Canonical correlations of past inputs and future outputs for linear stochastic systems," *Systems and Control Letters*, vol. 4, pp. 209–215, 1984.
127. L. Pernebo and L. M. Silverman, "Model reduction via balanced state space representations," *IEEE Trans. Automatic Control*, vol. AC-27, no. 2, pp. 382–387, 1982.
128. K. Peternell, W. Scherrer and M. Deistler, "Statistical analysis of novel subspace identification methods," *Signal Processing*, vol. 52, no. 2, pp. 161–177, 1996.
129. G. Picci, "Stochastic realization and system identification," In *Statistical Methods in Control and Signal Processing* (T. Katayama and S. Sugimoto, eds.), Marcel Dekker, pp. 1–36, 1997.
130. G. Picci and T. Katayama, "Stochastic realization with exogenous inputs and 'subspace methods' identification," *Signal Processing*, vol. 52, no. 2, pp. 145–160, 1996.
131. G. Picci and T. Katayama, "A simple 'subspace' identification method with exogenous inputs," *Proc. IFAC 13th World Congress*, San Francisco, Vol. I, pp. 175–180, 1996.
132. R. Pintelon and J. Schoukens, *System Identification – A Frequency Domain Approach*, IEEE Press, 2001.
133. V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophysical J. Royal Astronomical Society*, vol. 33, pp. 347–366, 1973.
134. M. B. Priestley, *Spectral Analysis and Time Series, vol. 1: Univariate Series; vol. 2: Multivariate Series, Prediction and Control*, Academic, 1981.
135. A. Rantzer, "On the Kalman-Yakubovich-Popov lemma," *Systems and Control Letters*, vol. 28, no. 1, pp. 7–10, 1996.

136. C. R. Rao, *Linear Statistical Inference and Its Applications* (2nd ed.), Wiley, 1973.
137. J. Rissanen, "Recursive identification of linear systems," *SIAM J. Control*, vol. 9, no. 3, pp. 420–430, 1971.
138. Yu. A. Rozanov, *Stationary Random Processes*, Holden-Day, 1967.
139. A. H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley, 2003.
140. R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.
141. R. Schrama, "Control oriented approximate closed loop identification via fractional representation," *Proc. 1991 American Control Conference*, Boston, pp. 719–720, 1991.
142. R. H. Shumway, *Applied Statistical Time Series Analysis*, Prentice-Hall, 1988.
143. L. M. Silverman, "Realization of linear dynamical systems," *IEEE Trans. Automatic Control*, vol. AC-16, no. 6, pp. 554–567, 1971.
144. T. Söderström, *Discrete-time Stochastic Systems*, Springer, 2002.
145. T. Söderström and P. Stoica, *System Identification*, Prentice-Hall, 1989.
146. L. H. Son and B. D. O. Anderson, "Design of Kalman filters using signal-model output statistics," *Proc. IEE*, vol. 120, no. 2, pp. 312–318, 1973.
147. E. D. Sontag, *Mathematical Control Theory*, Springer, 1990.
148. G. W. Stewart, "On the early history of the singular value decomposition," *SIAM Review*, vol. 35, no. 4, pp. 551–566, 1993.
149. S. M. Stigler, *The History of Statistics: The Measurement of Uncertainty before 1900*, The Belknap Press of Harvard University Press, 1986.
150. P. Stoica and R. Moses, *Introduction to Spectral Analysis*, Prentice-Hall, 1997.
151. H. Tanaka and T. Katayama, "A stochastic realization in a Hilbert space based on 'LQ decomposition' with application to subspace identification," *Proc. 13th IFAC Symp. System Identification*, Rotterdam, pp. 899–904, 2003.
152. H. Tanaka and T. Katayama, "Stochastic realization on a finite interval via 'LQ decomposition' in Hilbert space," *Proc. 7th European Control Conference*, Cambridge, Paper No. SI-2-2, 2003.
153. H. Tanaka and T. Katayama, "A stochastically balanced realization on a finite interval," *Proc. 16th Int. Symp. on Mathematical Theory of Network and Systems*, THP8 (Identification), Leuven, 2004.
154. H. Tanaka and T. Katayama, "Stochastic subspace identification algorithm guarnteeing stability and minimum phase," *Preprints 16th IFAC World Congress*, Prague, July 2005.
155. A. J. Tether, "Construction of minimal linear state-variable models from finite input-output data," *IEEE Trans. Automatic Control*, vol. 15, no. 4, pp. 427–436, 1970.
156. L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, 1998.
157. L. N. Trefethen and D. Bau, III, *Numerical Linear Algebra*, SIAM, 1997.
158. P. M. J. Van den Hof, "Closed-loop issues in system identification," *Proc. 11th IFAC Symp. System Identification*, Kitakyushu, Japan, pp. 1651–1664, 1997.
159. P. M. J. Van den Hof and R. A. de Callafon, "Multivariable closed-loop identification: From indirect identification to dual-Youla parametrization," *Proc. 35th IEEE Conference on Decision and Control*, Kobe, Japan, pp. 1397–1402, 1996.
160. P. M. J. Van den Hof and R. J. P. Schrama, "An indirect method for transfer function estimation from closed loop data," *Automatica*, vol. 29, no. 6, pp. 1523–1527, 1993.
161. A. C. Van der Klauw, M. Verhaegen and P. P. J. Van den Bosch, "State space identification of closed loop systems," *Proc. 30th IEEE Conference on Decision and Control*, Brighton, pp. 1327–1332, 1991.

162. A. J. Van der Veen, E. F. Deprettere and A. Swindlehurst, "Subspace-based signal analysis using singular value decomposition," *Proc. IEEE*, vol. 81, no. 9, pp. 1277–1308, 1993.
163. P. Van Overschee and B. De Moor, "Subspace algorithms for the stochastic identification problem," *Automatica*, vol. 29, no. 3, pp. 649–660, 1993.
164. P. Van Overschee and B. De Moor, "N4SID - Subspace algorithms for the identification of combined deterministic - stochastic systems," *Automatica*, vol. 30, no. 1, pp. 75–93, 1994.
165. P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems*, Kluwer Academic Pub., 1996.
166. P. Van Overschee, B. De Moor, W. Dehandschutter and J. Swevers, "A subspace algorithm for the identification of discrete-time frequency domain power spectra," *Automatica*, vol. 33, no. 12, pp. 2147–2157, 1997.
167. P. Van Overschee and B. De Moor, "Closed loop subspace systems identification," *Proc. 36th IEEE Conference on Decision and Control*, San Diego, pp. 1848–1853, 1997.
168. A. Varga, "On balancing and order reduction of unstable periodic systems," *Preprints 1st IFAC Workshop on Periodic Control Systems*, Como, pp. 177–182, 2001.
169. M. Verhaegen, "Subspace model identification, Part 3: Analysis of the ordinary output-error state-space model identification algorithm," *Int. J. Control*, vol. 58, no. 3, pp. 555–586, 1993.
170. M. Verhaegen, "Application of a subspace model identification technique to identify LTI systems operating in closed loop," *Automatica*, vol. 29, no. 4, pp. 1027–1040, 1993.
171. M. Verhaegen, "Identification of the deterministic part of MIMO state space models given in innovations form from input-output data," *Automatica*, vol. 30, no. 1, pp. 61–74, 1994.
172. M. Verhaegen and P. Dewilde, "Subspace model identification, Part 1: The output-error state-space model identification class of algorithms," *Int. J. Control*, vol. 56, no. 5, pp. 1187–1210, 1992.
173. M. Verhaegen and P. Dewilde, "Subspace model identification, Part 2: Analysis of the elementary output-error state space model identification algorithm," *Int. J. Control*, vol. 56, no. 5, pp. 1211–1241, 1992.
174. M. Viberg, B. Ottersten, B. Wahlberg and L. Ljung, "A statistical perspective on state-space modeling using subspace methods," *Proc. 30th IEEE Conference on Decision and Control*, Brighton, pp. 1337–1342, 1991.
175. M. Viberg, "Subspace-based methods for identification of linear time-invariant systems," *Automatica*, vol. 31, no. 12, pp. 1835–1851, 1995.
176. M. Viberg and P. Stoica, "Editorial note," *Signal Processing*, Special Issue on Subspace Methods, Part I: Array Processing and Subspace Computations, vol. 50, nos. 1-2, pp. 1–3, 1996; Part II: System Identification, vol. 52, no. 2, pp. 99–101, 1996.
177. K. Wada, "What's the subspace identification methods?" *J. Society Instrument and Control Engineers*, vol. 36, no. 8, pp. 569–574, 1997 (in Japanese).
178. P. Whittle, *Prediction and Regulation by Linear Least-Square Methods* (2nd ed.), Basil Blackwell Publisher, 1984.
179. N. Wiener and P. Masani, "The prediction theory of multivariable stochastic processes - I," *Acta Mathematica*, vol. 98, pp. 111–150, 1957.
180. N. Wiener and P. Masani, "The prediction theory of multivariable stochastic processes - II," *Acta Mathematica*, vol. 99, pp. 93–137, 1958.
181. J. Willems, I. Markovsky, P. Rapisarda and B. De Moor, "A note on persistency of excitation," Technical Report 04-54, Department of Electrical Engineering, Katholieke Universiteit Lueven, 2004.

182. W. M. Wonham, "On a matrix Riccati equation of stochastic control," *SIAM J. Control*, vol. 6, no. 4, pp. 681–698, 1968.
183. N. Young, *An Introduction to Hilbert Space*, Cambridge University Press, 1988.
184. H. P. Zeiger and A. J. McEwen, "Approximate linear realization of given dimension via Ho's algorithm," *IEEE Trans. Automatic Control*, vol. AC-19, no. 2, p. 153, 1974.
185. K. Zhou, J. C. Doyle and K. Glover, *Robust Optimal Control*, Prentice-Hall, 1996.
186. K. Zhou, G. Salomon and E. Wu, "Balanced realization and model reduction for unstable systems," *Int. J. Robust and Nonlinear Control*, vol. 9, no. 2, pp. 183–198, 1999.

Index