CHAPTER 5

PARAMETRIC OPTIMIZATION

1. INTRODUCTION

The previous chapters have developed the tools which are required to analyze dynamical systems whose input signals are stochastic processes. This chapter will show how these tools can be exploited to design control systems. We will thus consider systems whose environment can be described by disturbances which are stochastic processes. It is also assumed that the control system is given apart from a number of parameters which can be chosen arbitrarily. We will then analyze how to choose the parameters in order to optimize the performance of the system. It is assumed that the system can be characterized by linear equations and the performance by the expected value of a loss function, which is a quadratic function of the state variables of the system.

The problem of parametric optimization can be divided into two parts:

- evaluation of performance
- optimization of performance with respect to parameters

The optimization can occasionally be done analytically but in most cases we must use numerical methods. There are many numerical methods available; some require evaluation of the loss functions only, while other require evaluation of gradients and possibly also higher order derivatives of the loss function. It turns out that the problem of evaluating derivatives of the loss function is the same type of problem as the evaluation of the loss function itself. We will therefore concentrate on the evaluation of the loss function.

The problem can be approached in two different ways: in the time domain or the frequency domain. Analysis in the frequency domain leads to the problem of evaluating integrals such as

$$\int_{-i\infty}^{i\infty} G(s)\,G(-s)\,ds$$

or

$$\oint H(z)\,H(z^{-1})\frac{dz}{z}$$

where G and H are rational functions of a complex variable. The details are given in Sections 2 and 3 for discrete time systems and continuous time systems respectively. It is interesting to see that the continuous and discrete time systems will require approximately the same amount of work and the same degree of complexity. Analysis in the time domain leads to equations of the type

$$P(t+1) = \Phi P(t)\Phi^T + R_1$$

in the discrete time case or

$$\frac{dP}{dt} = AP + PA^T + R_1$$

in the continuous time case.

The frequency domain approach to the problem is discussed in Section 2 for discrete time systems and in Section 3 for continuous time systems. The time domain aspect of the problem is discussed in connection with the problem of reconstructing the state of a noisy dynamical system from noisy observations. Using heuristic arguments, we arrive at a structure for the reconstructor which seems reasonable. The reconstructor has a number of undetermined parameters. These parameters are determined in such a way as to minimize the mean square reconstruction error. It will be shown later in Chapter 7 that the structures derived heuristically are in fact optimal. The reconstructors obtained are thus Kalman filters. The discrete time case is discussed in Section 4 and the continuous time case in Section 5. It is of interest to observe that the parameters can actually be time-varying.

2. EVALUATION OF LOSS FUNCTIONS FOR DISCRETE TIME SYSTEMS

Statement of the Problem

In Chapter 4 we developed the tools which were required to analyze linear systems subject to disturbances which can be described as stochastic processes. For linear time invariant systems whose disturbances are stationary stochastic processes with rational spectral densities, we found that the spectral density for any system variable can be expressed as

$$\phi(\omega) = H(z)\,H(z^{-1})$$

where $z = e^{i\omega}$ and H is a rational function. The variance of the system variable is then given by

$$\sigma^2 = \int_{-\pi}^{\pi} \phi(\omega)\,d\omega = \frac{1}{i}\int_{-\pi}^{\pi} H(e^{i\omega})H(e^{-i\omega})e^{-i\omega}d(e^{i\omega}) = \frac{1}{i}\oint H(z)H(z^{-1})\frac{dz}{z}$$

where $\oint$ denotes the integral along the unit circle in the complex plane. To compute the variance of a signal in such a case we are thus led to the problem of evaluating the integral

$$I = \frac{1}{2\pi i}\oint \frac{B(z)B(z^{-1})}{A(z)A(z^{-1})}\cdot\frac{dz}{z} \tag{2.1}$$

where A and B are polynomials with real coefficients

$$A(z) = a_0z^n + a_1 z^{n-1} + \cdots + a_n \tag{2.2}$$

$$B(z) = b_0z^n + b_1 z^{n-1} + \cdots + b_n \tag{2.3}$$

and $\oint$ denotes the integral along the unit circle in the positive direction. The factor $1/2\pi$ is introduced for convenience only. We also assume that a_0 is positive.

The integral (2.1) can of course be evaluated in a straightforward manner using residue calculus. It turns out, however, that the general formulas are not practical to handle for systems of high order. For this purpose we will present recursive formulas for the evaluation of the integral (2.1) which are convenient both for hand and machine calculations.

Notations and Preliminaries

We first observe that the integral (2.1) will always exist if the polynomial $A(z)$ is stable, which means that all its zeros are inside the unit circle. In such a case we can always find a stable dynamical system with the pulse transfer function $B(z)/A(z)$, and the integral (2.1) is then simply the variance of the output when the input is white noise.

If $A(z)$ has zeros on the unit circle, the integral diverges. If $A(z)$ has zeros both inside and outside the unit circle, but not on the unit circle, the integral (2.1) still exists. In such a case we can always find a polynomial $A'(z)$ with all its zeros inside the unit circle such that

$$A(z)\,A(z^{-1}) = A'(z)\,A'(z^{-1})$$

and the integral then represents the variance of the output of a stable dynamical system with the pulse transfer function $B(z)/A'(z)$.

In many practical cases, however, we obtain the integral as a result of an analysis of a dynamical system whose pulse transfer function is $B(z)/A(z)$. In such a case, it is naturally of great importance to test that the denominator $A(z)$ of the pulse transfer function has all its zeros inside the unit circle because, when this is not the case, the dynamical system will be unstable although the integral (2.1) exists.

In order to present the result in a simple form we will first introduce some notations. Let A^* denote the reciprocal polynomial defined by

$$A^*(z) = z^n A(z^{-1}) = a_0 + a_1 z + \cdots + a_n z^n \tag{2.4}$$

Further introduce the polynomials†

$$A_k(z) = a_0^k z^k + a_1^k z^{k-1} + \cdots + a_k^k \tag{2.5}$$

$$B_k(z) = b_0^k z^k + b_1^k z^{k-1} + \cdots + b_k^k \tag{2.6}$$

which are defined recursively by

$$A_{k-1}(z) = z^{-1}\{A_k(z) - \alpha_k A_k^*(z)\} \tag{2.7}$$

$$B_{k-1}(z) = z^{-1}\{B_k(z) - \beta_k A_k^*(z)\} \tag{2.8}$$

where

$$\alpha_k = a_k^k / a_0^k \tag{2.9}$$

$$\beta_k = b_k^k / a_0^k \tag{2.10}$$

and

$$A_n(z) = A(z) \tag{2.11}$$

$$B_n(z) = B(z) \tag{2.12}$$

The coefficients of the polynomials A_k and B_k are thus given by the recursive equations

$$a_i^{k-1} = a_i^k - \alpha_k a_{k-i}^k \qquad i = 0, 1, \ldots, k-1 \tag{2.13}$$

$$b_i^{k-1} = b_i^k - \beta_k a_{k-i}^k \qquad i = 0, 1, \ldots, k-1 \tag{2.14}$$

with the initial conditions

$$a_i^n = a_i \tag{2.15}$$

$$b_i^n = b_i \tag{2.16}$$

If the equations given above should have any meaning, we must naturally require that all a_0^k are different from zero. The coefficient a_0^n can of course always be chosen different from zero. The following theorem gives necessary and sufficient conditions.

THEOREM 2.1

Let $a_0^k > 0$, then the following conditions are equivalent:

1. The polynomial $A_k(z)$ is stable

† Notice that k in a_0^k is a superscript!

2. The polynomial $A_{k-1}(z)$ is stable and a_0^{k-1} is positive.

By repeated application of this theorem, we thus find that if the polynomial $A_n(z)$ is stable, then all coefficients $a_0{}^k$ are positive. In order to prove Theorem 2.1 we need the following result.

LEMMA 2.1

Let the polynomial $f(z)$, with real coefficients, have all its roots inside the unit circle, then

$$\begin{aligned} |f(z)| &< |f^*(z)| \quad &\text{for} \quad |z| < 1 \\ |f(z)| &= |f^*(z)| \quad &\text{for} \quad |z| = 1 \\ |f(z)| &> |f^*(z)| \quad &\text{for} \quad |z| > 1 \end{aligned}$$

Proof

Put

$$f(z) = \beta \prod_{i=1}^{n} (z - \alpha_i) \qquad |\alpha_i| < 1$$

Then

$$f^*(z) = \beta \prod_{i=1}^{n} (1 - \alpha_i z)$$

Introduce

$$w(z) = \frac{f(z)}{f^*(z)} = \prod_{i=1}^{n} \frac{z - \alpha_i}{1 - \alpha_i z} = \prod_{i=1}^{n} \frac{z - \alpha_i}{1 - \bar{\alpha}_i z}$$

where $\bar{\alpha}_i$ denotes the complex conjugate of α_i. The last equality follows from the fact that f has real coefficients. If α_i is a zero of f, then $\bar{\alpha}_i$ is also a zero. Now consider the transformation

$$w_i(z) = \frac{z - \alpha_i}{1 - \bar{\alpha}_i z}$$

This transformation transforms the interior of the unit circle on to itself. The unit circle is an invariant of the transformation. The transformation

$$w(z) = \prod_{i=1}^{n} w_i(z) = \prod_{i=1}^{n} \frac{z - \alpha_i}{1 - \bar{\alpha}_i z} = \frac{f(z)}{f^*(z)}$$

then also has the same properties and the lemma is proven.

Proof of Theorem 2.1

After these preliminaries we can now prove Theorem 2.1. We will first show that 1.$\Longrightarrow$2. If 1. holds, it follows from Lemma 2.1 that

$$|A_k(0)| < |A_k^*(0)|$$

But $A_k(0) = a_k^k$ and $A_k^*(0) = a_0^k$. Hence

$$| \alpha_k | = | a_k^k / a_0^k | < 1 \tag{2.17}$$

Equation (2.13) then gives

$$a_0^{k-1} = a_0^k - (a_k^k)^2/a_0^k = [(a_0^k)^2 - (a_k^k)^2]/a_0^k > 0$$

Notice that it was assumed that $a_0^k > 0$. As $A_k(z)$ is stable, it also follows from Lemma 2.1 that

$$| A_k(z) | \geqslant | A_k^*(z) | \quad \text{for} \quad | z | \geqslant 1$$

Combining this with (2.17) we get

$$| A_k(z) | > | \alpha_k | \cdot | A_k^*(z) | \quad \text{for} \quad | z | \geqslant 1$$

We now find from (2.7) that

$$\begin{aligned} | z | \cdot | A_{k-1}(z) | &= | A_k(z) - \alpha_k A_k^*(z) | \\ &\geqslant | A_k(z) | - | \alpha_k | \cdot | A_k^*(z) | > 0 \quad \text{for} \quad | z | \geqslant 1 \end{aligned}$$

This implies that $A_{k-1}(z)$ has no roots outside the unit circle. The first part of the theorem is thus proven.

Now assume that condition 2. holds. Then

$$a_0^{k-1} = a_0^k - (a_k^k)^2/a_0^k = [(a_0^k)^2 - (a_k^k)^2]/a_0^k > 0$$

As a_0^k was assumed positive we get

$$| \alpha_k | = | a_k^k / a_0^k | < 1$$

It follows from (2.7) that

$$A_k(z) - \alpha_k A_k^*(z) = z A_{k-1}(z) \tag{2.18}$$

Hence

$$z^k A_k(z^{-1}) - \alpha_k z^k A_k^*(z^{-1}) = z^{k-1} A_{k-1}(z^{-1})$$

or

$$A_k^*(z) - \alpha_k A_k(z) = A_{k-1}^*(z) \tag{2.19}$$

Elimination of $A_k^*(z)$ between (2.18) and (2.19) gives

$$A_k(z) = \frac{z}{1 - \alpha_k^2} A_{k-1}(z) + \frac{\alpha_k}{1 - \alpha_k^2} A_{k-1}^*(z)$$

As $| \alpha_k | < 1$, the elimination is always possible.

For $| z | \geqslant 1$ we now have (Lemma 2.1)

$$| A_{k-1}(z) | \geqslant | A_{k-1}^*(z) |$$

Furthermore $| \alpha_k | < 1$. Hence for $| z | \geqslant 1$ we have

$$| A_k(z) | \geqslant \left| \frac{z}{1 - \alpha_k^2} \right| | A_{k-1}(z) | - \left| \frac{\alpha_k}{1 - \alpha_k^2} \right| \cdot | A_{k-1}^*(z) | > 0$$

The polynomial $A_k(z)$ has no zeros outside the unit circle and the theorem is proven.

We have previously found that $a_0^k > 0$ for all k was a necessary condition for $A(z)$ to be stable. We will now show that the converse is also true. Hence assume that all a_0^k are positive. The trivial polynomial A_0 is stable as $a_0^0 > 0$. Theorem 2.1 then implies that A_1 is stable. By repeated application of Theorem 2.1, we thus find that the polynomial A_k is stable. Hence, if the polynomial $A(z)$ has all zeros inside the unit circle, then all coefficients a_0^k are positive. If any coefficient a_0^k is nonpositive, then the system with the pulse-transfer function $B(z)/A(z)$ is unstable. Summing up we get Theorem 2.2.

THEOREM 2.2

Let $a_0^n > 0$ then the following conditions are equivalent

1. $A_n(z)$ is stable
2. $a_0^k > 0$ for $k = 0, 1, \ldots, n-1$

The Main Result

We will now show that the integral (2.1) can be computed recursively. For this purpose, we introduce the integrals I_k defined by

$$I_k = \frac{1}{2\pi i} \oint \frac{B_k(z)\, B_k(z^{-1})}{A_k(z)\, A_k(z^{-1})} \cdot \frac{dz}{z} \tag{2.20}$$

It follows from (2.1) that $I = I_n$. We now have Theorem 2.3.

THEOREM 2.3

Let the polynomial $A(z)$ have all its zeros inside the unit circle. The integrals I_k defined by (2.20) then satisfy the following recursive equation

$$[1 - \alpha_k^2]\, I_{k-1} = I_k - \beta_k^2 \tag{2.21}$$

$$I_0 = \beta_0^2 \tag{2.22}$$

Proof

As $A(z)$ has all its zeros inside the unit circle, it follows from Theorem 2.2 that all a_0^k are different from zero. It thus follows from (2.7) and (2.8) that all polynomials A_k and B_k can be defined. Furthermore it follows from Theorem 2.2 that all polynomials A_k have all zeros inside the unit circle. All integrals I_k thus exist.

To prove the theorem, we will make use of the theory of analytic functions. We will first assume that the polynomial $A(z)$ has distinct roots, which are different from zero. The integral (2.20) equals the sum of residues at the poles of the function $B_k(z)\, B_k(z^{-1})/\{z A_k(z)\, A_k(z^{-1})\}$ inside the unit circle. As the integral is invariant under the change of variables

$z \to 1/z$, we also find that the integral equals the sum of residues of the poles outside the unit circle.

Now consider

$$I_{k-1} = \frac{1}{2\pi i} \oint \frac{B_{k-1}(z)\, B_{k-1}(z^{-1})}{A_{k-1}(z)\, A_{k-1}(z^{-1})} \cdot \frac{dz}{z}$$

The poles of the integrand inside the unit circle are $z = 0$ and the zeros z_i of the polynomial $A_{k-1}(z)$. Since $A_{k-1}(z_i) = 0$, it follows from (2.7) and (2.4) that

$$A_k(z_i) = \alpha_k A_k^*(z_i) = \alpha_k z_i^k A_k(z_i^{-1})$$

Combining this equation with (2.7) and (2.4) we find

$$\begin{aligned} A_{k-1}(z_i^{-1}) &= z_i[A_k(z_i^{-1}) - \alpha_k A_k^*(z_i^{-1})] \\ &= z_i[A_k(z_i^{-1}) - \alpha_k z_i^{-k} A_k(z_i)] = (1 - \alpha_k^2) z_i A_k(z_i^{-1}) \end{aligned}$$

Further it follows from (2.4) and (2.7) that

$$A_{k-1}^*(z) = A_k^*(z) - \alpha_k A_k(z)$$

Hence

$$A_{k-1}^*(0) = A_k^*(0) - \alpha_k A_k(0) = a_0^k - \alpha_k a_k^k = a_0^k(1 - \alpha_k^2)$$

The functions

$$\frac{B_{k-1}(z) B_{k-1}(z^{-1})}{A_{k-1}(z) A_{k-1}(z^{-1})} \cdot \frac{1}{z} = \frac{B_{k-1}(z) B_{k-1}^*(z)}{A_{k-1}(z) A_{k-1}^*(z)} \cdot \frac{1}{z}$$

and

$$\frac{B_{k-1}(z) B_{k-1}(z^{-1})}{A_{k-1}(z)[z(1 - \alpha_k^2) A_k(z^{-1})]} \cdot \frac{1}{z} = \frac{B_{k-1}(z) B_{k-1}^*(z)}{A_{k-1}(z)[(1 - \alpha_k^2) A_k^*(z)]} \cdot \frac{1}{z}$$

have the same poles inside the unit circle and the same residues at these poles. Hence

$$\begin{aligned} I_{k-1} &= \frac{1}{1 - \alpha_k^2} \cdot \frac{1}{2\pi i} \oint \frac{B_{k-1}(z) B_{k-1}(z^{-1})}{A_{k-1}(z) A_k(z^{-1})} \cdot \frac{dz}{z^2} \\ &= \frac{1}{1 - \alpha_k^2} \cdot \frac{1}{2\pi i} \oint \frac{B_{k-1}(z) B_{k-1}(z^{-1})}{A_k(z) A_{k-1}(z^{-1})}\, dz \end{aligned} \tag{2.23}$$

where the second equality is obtained by making the variable substitution $z \to z^{-1}$. The integrand has poles at the zeros of $A_k(z)$. It follows, however, from (2.7) that

$$A_{k-1}(z^{-1}) = z\{A_k(z^{-1}) - \alpha_k A_k^*(z^{-1})\} = z\{A_k(z^{-1}) - \alpha_k z^{-k} A_k(z)\}$$

Hence for z_i such that $A_k(z_i) = 0$, we get

$$A_{k-1}(z_i^{-1}) = z_i A_k(z_i^{-1})$$

The functions

$$\frac{B_{k-1}(z)B_{k-1}(z^{-1})}{A_k(z)A_{k-1}(z^{-1})}$$

and

$$\frac{B_{k-1}(z)B_{k-1}(z^{-1})}{A_k(z)A_k(z^{-1})}\cdot\frac{1}{z}=\frac{B_{k-1}(z)B_{k-1}^*(z)}{A_k(z)A_k^*(z)}$$

thus have the same poles inside the unit circle and the same residues at these poles. The integral of these functions around the unit circle are thus the same. Equation (2.23) now gives

$$I_{k-1}=\frac{1}{1-\alpha_k^2}\cdot\frac{1}{2\pi i}\oint\frac{B_{k-1}(z)B_{k-1}(z^{-1})}{A_k(z)A_k(z^{-1})}\cdot\frac{dz}{z}$$

Now introduce (2.8) and we find

$$\begin{aligned}(1-\alpha_k^2)I_{k-1}&=\frac{1}{2\pi i}\oint\frac{[B_k(z)-\beta_kA_k^*(z)][B_k(z^{-1})-\beta_kA_k^*(z^{-1})]}{A_k(z)A_k(z^{-1})}\cdot\frac{dz}{z}\\&=\frac{1}{2\pi i}\oint\frac{B_k(z)B_k(z^{-1})}{A_k(z)A_k(z^{-1})}\cdot\frac{dz}{z}-\frac{\beta_k}{2\pi i}\oint\frac{B_k(z)A_k^*(z^{-1})}{A_k(z)A_k(z^{-1})}\cdot\frac{dz}{z}\\&\quad-\frac{\beta_k}{2\pi i}\oint\frac{A_k^*(z)B_k(z^{-1})}{A_k(z)A_k(z^{-1})}\cdot\frac{dz}{z}+\frac{\beta_k^2}{2\pi i}\oint\frac{A_k^*(z)A_k^*(z^{-1})}{A_k(z)A_k(z^{-1})}\cdot\frac{dz}{z}\end{aligned}\tag{2.24}$$

The first integral equals I_k. The second integral can be reduced as follows

$$\begin{aligned}\frac{\beta_k}{2\pi i}\oint\frac{B_k(z)A_k^*(z^{-1})}{A_k(z)A_k(z^{-1})}\cdot\frac{dz}{z}&=\frac{\beta_k}{2\pi i}\oint\frac{B_k(z)A_k(z)}{A_k(z)A_k^*(z)}\cdot\frac{dz}{z}\\&=\frac{\beta_k}{2\pi i}\oint\frac{B_k(z)}{A_k^*(z)}\cdot\frac{dz}{z}=\beta_k\frac{B_k(0)}{A_k^*(0)}=\beta_k\frac{b_k^k}{a_0^k}=\beta_k^2\end{aligned}$$

where the first equality follows from (2.4), the third from the residue theorem, and the fifth from (2.10). Similarly we find that the third integral of the right member of (2.24) also equals β_k^2.

Using (2.4), the fourth term of the right member of (2.24) can be reduced as follows

$$\frac{\beta_k^2}{2\pi i}\oint\frac{A_k^*(z)A_k^*(z^{-1})}{A_k(z)A_k(z^{-1})}\cdot\frac{dz}{z}=\frac{\beta_k^2}{2\pi i}\oint\frac{dz}{z}=\beta_k^2$$

Summarizing we find (2.21). When $k=0$ we get from (2.20)

$$I_0=\frac{1}{2\pi i}\oint\left(\frac{b_0^0}{a_0^0}\right)^2\cdot\frac{dz}{z}=\beta_0^2$$

We have thus proven the formulas (2.21) and (2.22) when $A(z)$ has distinct roots. If A has multiple roots or roots equal to zero, we can always perturb its coefficients in order to obtain distinct and nonzero roots. Equations (2.21) and (2.22) then hold. As the numbers α_k and β_k are continuous functions of the parameters, we find that (2.20) and (2.21) hold even when A has multiple roots.

Notice that it follows from (2.13) that

$$a_0^{k-1} = a_0^k - \alpha_k a_k^k = a_0^k(1 - \alpha_k^2)$$

Equation (2.21) can then be written as

$$a_0^{k-1} I_{k-1} = a_0^k I_k - a_0^k \beta_k^2$$

or

$$a_0^k I_k = a_0^{k-1} I_{k-1} + \beta_k b_k^k = a_0^{k-1} I_{k-1} + (b_k^k)^2 / a_0^k$$

COROLLARY 2.1

The integral I_k is given by

$$I_k = \frac{1}{a_0^k} \sum_{i=0}^{k} \frac{(b_i^i)^2}{a_0^i} \tag{2.25}$$

Computational Aspects

Having obtained the recursive formula given by Theorem 2.3 we will now turn to the practical aspects of the computations. To obtain the integrals we must first compute the coefficients of the polynomials $A_k(z)$ and $B_k(z)$. This is easily done with the following tables

$$\begin{array}{lllll}
a_0 & a_1 & \cdots & a_{n-1} & a_n \\
a_n & a_{n-1} & \cdots & a_1 & a_0 \\
a_0^{n-1} & a_1^{n-1} & \cdots & a_{n-1}^{n-1} & \\
a_{n-1}^{n-1} & a_{n-2}^{n-1} & \cdots & a_0^{n-1} & \\
\vdots & & & & \\
a_0^1 & a_1^1 & & & \\
a_1^1 & a_0^1 & & & \\
a_0^0 & & & &
\end{array}
\qquad
\begin{array}{lllll}
b_0 & b_1 & \cdots & b_{n-1} & b_n \\
a_n & a_{n-1} & \cdots & a_1 & a_0 \\
b_0^{n-1} & b_1^{n-1} & \cdots & b_{n-1}^{n-1} & \\
a_{n-1}^{n-1} & a_{n-2}^{n-1} & \cdots & a_0^{n-1} & \\
\vdots & & & & \\
b_0^1 & b_1^1 & & & \\
a_1^1 & a_0^1 & & & \\
b_0^0 & & & &
\end{array}$$

Each even row in the table of coefficients of A (A-table) is obtained by writing the coefficients of the proceeding row in reverse order. The even rows of the A- and B-tables are the same. The coefficients of the odd rows of both tables are obtained from the two elements above using (2.13) and (2.14)

$$a_i^{k-1} = a_i^k - \alpha_k a_{k-i}^k, \qquad \alpha_k = a_k^k / a_0^k \tag{2.13}$$

$$b_i^{k-1} = b_i^k - \beta_k a_{k-i}^k, \qquad \beta_k = b_k^k / a_0^k \tag{2.14}$$

Using the stability criterion of Theorem 2.2, we find that the polynomial $A(z)$ has all zeros inside the unit circle if all coefficients a_0^k are positive. These elements are boldfaced in the table above. Having obtained the coefficients α_k and β_k we can now easily obtain the value of the integral from (2.25).

Notice that in order to investigate the stability of the polynomial $A(z)$ we have to form the A-table only. Hence the work required to calculate the integral I is roughly twice the work required to test the stability of the polynomial $A(z)$.

EXAMPLE

As an illustration, we will evaluate the integral for

$$A(z) = z^3 + 0.7z^2 + 0.5z - 0.3$$
$$B(z) = z^3 + 0.3z^2 + 0.2z + 0.1$$

We get the following tables

				α_k					β_k
1	0.7	0.5	−0.3		1	0.3	0.2	0.1	
−0.3	0.5	0.7	1.0	−0.3	−0.3	0.5	0.7	1.0	0.1
0.91	0.85	0.71			1.03	0.25	0.13		
0.71	0.85	0.91		0.780	0.71	0.85	0.91		0.143
0.356	0.187				0.929	0.129			
0.187	0.356			0.525	0.187	0.356			0.361
0.258				1	0.861				3.338

and we find $I = 2.9488$

The given formulas are well suited for machine calculations. See the FORTRAN program (p. 126) for the computation.

Exercises

1. Evaluate the integral (2.1) for

$$A(z) = z^2 + 0.4z + 0.1$$
$$B(z) = z^2 + 0.9z + 0.8$$

(Answer $I = 1.565079$)

2. A simple inventory control system can be described by the equations

$$I(t) = I(t-1) + P(t) - S(t)$$
$$P(t) = P(t-1) + u(t-k)$$

where I denotes inventory level, P production, and S sales. The deci-

```
      SUBROUTINE SALOSS (A, B, N, IERR, V, IN)
C
C     PROGRAM FOR EVALUATING THE INTEGRAL OF THE RATIONAL
C     FUNCTION
C        1/(2*PI*I)*B(Z)*B(1/Z)/(A(Z)*A(1/Z))*(1/Z)
C     AROUND THE UNIT CIRCLE
C
C     A—VECTOR WITH THE COEFFICIENTS OF THE POLYNOMIAL
C        A(1)*Z**N + A(2)*Z**(N - 1) + ··· + A(N + 1)
C     B—VECTOR WITH THE COEFFICIENTS OF THE POLYNOMIAL
C        B(1)*Z**N + B(2)*Z**(N - 1) + ··· + B(N + 1)
C
C        THE VECTORS A AND B ARE DESTROYED
C
C     N—ORDER OF THE POLYNOMIALS A AND B (MAX 10)
C     IERR—WHEN RETURNING IERR = 1 IF A HAS ALL ZEROS INSIDE UNIT
C        CIRCLE IERR = 0 IF THE POLYNOMIAL A HAS ANY ROOT OUTSIDE
C        OR ON THE UNIT CIRCLE OR IF A(1) IS NOT POSITIVE
C     V—THE RETURNED LOSS
C     IN—DIMENSION OF A AND B IN MAIN PROGRAM
C
C     SUBROUTINE REQUIRED
C        NONE
C
      DIMENSION A(IN), B(IN), AS(11)
C
      A0 = A(1)
      IERR = 1
      V = 0.0
      DO 10 K = 1, N
      L = N + 1 - K
      L1 = L + 1
      ALFA = A(L1)/A(1)
      BETA = B(L1)/A(1)
      V = V + BETA*B(L1)
      DO 20 I = 1, L
      M = L + 2 - I
      AS(I) = A(I) - ALFA*A(M)
20    B(I) = B(I) - BETA*A(M)
      IF (AS(1)) 50, 50, 30
30    DO 40 I = 1, L
40    A(I) = AS(I)
10    CONTINUE
      V = V + B(1)**2/A(1)
      V = V/A0
      RETURN
50    IERR = 0
      RETURN
      END
```

sion variable is denoted by u, and the production delay is k units. Assume that the following decision rule is used to refill the inventory

$$u(t) = \alpha[I_0 - I(t)]$$

Determine the variance of the fluctuations in production and inventory level when the fluctuations in sales can be described as a sequence of independent equally distributed random variables with zero mean values and standard deviation σ.

3. Show that the integral I defined by (2.1) can be computed as the first component x_1 of the solution of the following linear equation

$$\begin{bmatrix} 2a_0 & 2a_1 & 2a_2 & 2a_3 & \cdots 2a_n \\ a_1 & a_0 + a_2 & a_1 + a_3 & a_2 + a_4 \cdots & a_{n-1} \\ a_2 & a_3 & a_0 + a_4 & a_1 + a_5 \cdots & a_{n-2} \\ \vdots & & & & \\ a_{n-1} & a_n & 0 & 0 & \cdots a_1 \\ a_n & 0 & 0 & 0 & \cdots a_0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \\ x_{n+1} \end{bmatrix} = \begin{bmatrix} 2\sum_{i=0}^{n} b_i^2 \\ 2\sum_{i=0}^{n-1} b_i b_{i+1} \\ 2\sum_{i=0}^{n-2} b_i b_{i+2} \\ \vdots \\ 2\sum_{i=0}^{1} b_i b_{i+n-1} \\ 2\, b_0 b_n \end{bmatrix}$$

Compare the number of computations required when evaluating the integral as a solution to this linear equation with the computations required when using Theorem 2.1.

4. When the function A_{k-1} is determined from A_k using (2.7), the constant term of A_k is eliminated. Show that a result which is similar to Theorem 2.3 can be obtained by the following reductions

$$A_{k-1}(z) = A_k(z) - \frac{a_0{}^k}{a_k{}^k} A_k^*(z)$$

$$B_{k-1}(z) = B_k(z) - \frac{b_0{}^k}{a_k{}^k} A_k^*(z)$$

which eliminate the terms of highest order in the polynomials A_k and B_k.
Hint:

$$I_k = \left[\left(\frac{a_0{}^k}{a_k{}^k}\right)^2 - 1\right] I_{k-1} + 2\frac{b_0{}^k b_k{}^k}{a_0{}^k a_k{}^k} - \left(\frac{b_0{}^k}{a_k{}^k}\right)^2$$

5. Give recursive algorithms for evaluating the integrals

$$\frac{1}{2\pi i}\oint\frac{B(z)\,B(z^{-1})}{A(z)\,A(z^{-1})}z^k\,dz$$

$$\frac{1}{2\pi i}\oint\frac{z^k\,B(z^{-1})}{A(z)\,A(z^{-1})}dz$$

6. Verify that the given FORTRAN program gives the desired result.

3. EVALUATION OF LOSS FUNCTIONS FOR CONTINUOUS TIME SYSTEMS

Statement of the Problem

We will now analyze the continuous time version of the problem discussed in Section 2. Consider a continuous time, linear time invariant dynamical system subject to a disturbance which is a stationary stochastic process with a rational spectral density. The variance of a system variable can be expressed by an integral of the type

$$I = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\frac{B(s)\,B(-s)}{A(s)\,A(-s)}ds \tag{3.1}$$

where A and B are polynomials with real coefficients

$$A(s) = a_0 s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n \tag{3.2}$$

$$B(s) = \qquad b_1 s^{n-1} + \cdots + b_{n-1}s + b_n \tag{3.3}$$

The evaluation of the integral (3.1) will be discussed in this section. The integral (3.1) can also be interpreted as the variance of the signal obtained when white noise is fed through a stable filter with the transfer function $B(s)/A(s)$. The integral (3.1) will always exist if the polynomial $A(s)$ does not have any zeros on the imaginary axis. Notice that the polynomial B must be at least one degree less than the polynomial A. The physical interpretation is analogous to that of Section 2.

Notations and Preliminaries

To formulate the result we need some notations which will now be developed. A decomposition of the polynomial $A(s)$ into odd and even terms is first introduced. Hence

$$A(s) = \bar{A}(s) + \tilde{A}(s) \tag{3.4}$$

where

$$\bar{A}(s) = a_0 s^n + a_2 s^{n-2} + \cdots = \frac{1}{2}[A(s) + (-1)^n A(-s)] \tag{3.5}$$

$$\tilde{A}(s) = a_1 s^{n-1} + a_3 s^{n-3} + \cdots = \frac{1}{2}[A(s) - (-1)^n A(-s)] \tag{3.6}$$

We also introduce the polynomials $A_k(s)$ and $B_k(s)$ of lower order than n

$$A_k(s) = a_0^k s^k + a_1^k s^{k-1} + \cdots + a_k^k \tag{3.7}$$

$$B_k(s) = b_1^k s^{k-1} + b_2^k s^{k-2} + \cdots + b_k^k \tag{3.8}$$

which are defined recursively from the equations

$$A_{k-1}(s) = A_k(s) - \alpha_k s \tilde{A}_k(s) \tag{3.9}$$

$$B_{k-1}(s) = B_k(s) - \beta_k \tilde{A}_k(s) \tag{3.10}$$

where

$$\alpha_k = a_0^k / a_1^k \tag{3.11}$$

$$\beta_k = b_1^k / a_1^k \tag{3.12}$$

and

$$A_n(s) = A(s) \tag{3.13}$$

$$B_n(s) = B(s) \tag{3.14}$$

The polynomials A_{k-1} and B_{k-1} can apparently only be defined if $a_1^k \neq 0$. We will first establish necessary and sufficient conditions for this. It turns out that this problem is closely associated with the stability of the polynomials $A_k(s)$. We have the following result.

THEOREM 3.1

Let $a_0^k > 0$ then the following conditions are equivalent

1. The polynomial $A_k(s)$ has all zeros in the left half plane
2. The polynomial $A_{k-1}(s)$ has all zeros in the left half plane and a_1^k is positive.

To prove this theorem we will use the following lemma.

LEMMA 3.1

Let the real polynomial $f(s)$ with real coefficients have all zeros in the left half plane then

$$|f(s)| < |f(-s)| \qquad \text{Re } s < 0$$

$$|f(s)| = |f(-s)| \qquad \text{Re } s = 0$$

$$|f(s)| > |f(-s)| \qquad \text{Re } s > 0$$

Proof

As f is a polynomial with zeros in the left half plane we have

$$f(s) = \beta \prod_{i=1}^{n} (s - \alpha_i), \qquad \text{Re } \alpha_i < 0$$

Then

$$f(-s) = \beta \prod_{i=1}^{n} (-s - \alpha_i) = \beta \prod_{i=1}^{n} (-s - \bar{\alpha}_i)$$

Introduce

$$w(s) = \frac{f(s)}{f(-s)} = \prod_{i=1}^{n} \frac{s - \alpha_i}{-s - \bar{\alpha}_i}$$

Now consider the transformation

$$w_i(s) = \frac{\alpha_i - s}{\bar{\alpha}_i + s}$$

The transformation w_i maps the complex plane onto itself in such a way that the left half plane is mapped on the interior of the unit circle, the imaginary axis on the unit circle, and the right half plane on the exterior of the unit circle.

The transformation

$$w(s) = \prod_{i=1}^{n} w_i(s) = \prod_{i=1}^{n} \frac{\alpha_i - s}{\bar{\alpha}_i + s} = \frac{f(s)}{f(-s)}$$

has the same properties and the statements of the lemma are thus proven.

Proof of Theorem 3.1

We will first show that 1.$\Longrightarrow$2. Let $a_0^k > 0$ and let $A_k(s)$ have all its zeros in the left half plane. The proof that a_1^k is positive is by contradiction. Hence assume a_1^k nonpositive. Then take s real, positive, and sufficiently large. We then find $|A_k(s)| < |A_k(-s)|$ which contradicts Lemma 3.1. We can prove by a similar argument that a_1^k cannot be zero.

To prove that $A_{k-1}(s)$ has all its zeros in the left half plane, we observe that (3.6) and (3.9) give

$$A_{k-1}(s) = \left(1 - \frac{\alpha_k s}{2}\right) A_k(s) + (-1)^k \frac{\alpha_k s}{2} A_k(-s) \tag{3.15}$$

The polynomial $A_{k-1}(s)$ is of order $k - 1$. If we can show that the reciprocal polynomial

$$A_{k-1}^*(s) = s^{k-1} A_{k-1}(s^{-1}) = s^{-2}\left[\left(s - \frac{\alpha_k}{2}\right) A_k^*(s) + \frac{\alpha_k}{2}(-1)^k A_k^*(-s)\right] \tag{3.16}$$

has all zeros in the left half plane, it follows that $A_{k-1}(s)$ also has all zeros in the left half plane. Instead of analyzing (3.16) we will use imbedding, and consider the function

$$F(s, \alpha) = s^{-1}\left[\left(s - \frac{\alpha}{2}\right) A_k^*(s) + \frac{\alpha}{2}(-1)^k A_k^*(-s)\right] \tag{3.17}$$

for arbitrary real α in the interval $0 \leqslant \alpha \leqslant \alpha_k$. Notice that

$$F(s, \alpha_k) = sA^*_{k-1}(s) \tag{3.18}$$

As A^*_k has no zeros in the right half plane it follows from Lemma 3.1 that

$$|A^*_k(s)| \geqslant |A^*_k(-s)| \qquad \text{Re } s \geqslant 0$$

Now take s such that Re $s \geqslant 0$ and

$$\left| s - \frac{\alpha}{2} \right| > \left| \frac{\alpha}{2} \right|$$

It then follows from the triangle-inequality that

$$|F(s, \alpha)| = |s^{-1}| \left| \left(s - \frac{\alpha}{2} \right) A^*_k(s) + \frac{\alpha}{2} A^*_k(-s) \right|$$

$$\geqslant |s^{-1}| \left[\left| s - \frac{\alpha}{2} \right| |A^*_k(s)| - \left| \frac{\alpha}{2} \right| \cdot |A^*_k(-s)| \right] > 0$$

We thus find that the function F has no zeros in the set

$$S = \left\{ s; \ \text{Re } s \geqslant 0 \quad \text{and} \quad \left| s - \frac{\alpha}{2} \right| > \left| \frac{\alpha}{2} \right| \right\} \tag{3.19}$$

Compare with Fig. 5.1.

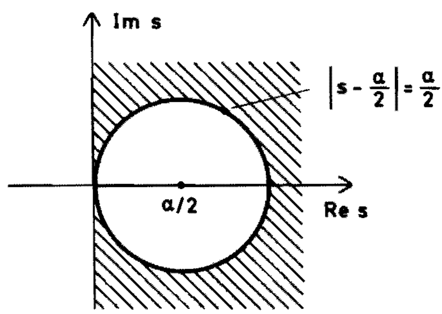

Fig. 5.1. Graph of the set $S = \{s;$ Re $s \geqslant 0$ and $|s - \alpha/2| > |\alpha/2|\}$.

To show that this implies that F does not have any zeros in the right half plane, we will use a continuity argument. As the function F is continuous in α, its zeros will also be continuous in α. We have

$$F(s, 0) = A^*_k(s)$$

For $\alpha = 0$, we thus find that F has all zeros in the left half plane. As F does not have any zeros in the set (3.19), it follows that as α increases, no zeros will cross the imaginary axis for $s \neq 0$. The only possibility to obtain zeros in the right half plane is to have a zero enter the right half plane at the origin as α increases. But

$$F(0, \alpha) = a_0^k - \alpha a_1^k > 0, \qquad 0 \leqslant \alpha < \alpha_k$$

Hence as α increases $F(s, \alpha)$ will have a zero at the origin for $\alpha = a_0/a_1 = \alpha_k$. This is a single zero because $F'(0, \alpha_k) = a_1^k > 0$. The function $F(s, \alpha_k)$ has one zero at the origin and all other zeros in the left half plane. Equation (3.18) then implies that $A_{k-1}^*(s)$ and thus also $A_{k-1}(s)$ has all zeros in the left half plane.

To prove that 2. $\Longrightarrow$ 1. we assume that $A_{k-1}(s)$ has all zeros in the left half plane and that a_1^k and a_0^k are positive. Equations (3.6) and (3.9) give

$$A_{k-1}(s) = \left(1 - \frac{\alpha_k s}{2}\right) A_k(s) + (-1)^k \frac{\alpha_k s}{2} A_k(-s)$$

$$A_{k-1}(-s) = \left(1 + \frac{\alpha_k s}{2}\right) A_k(-s) - (-1)^k \frac{\alpha_k s}{2} A_k(s)$$

Elimination of $A_k(-s)$ between these equations gives

$$A_k(s) = \left(1 + \frac{\alpha_k s}{2}\right) A_{k-1}(s) - (-1)^k \frac{\alpha_k s}{2} A_{k-1}(-s)$$

As a_0^k and a_1^k are positive, α_k is positive. For Re $s \geqslant 0$ we have

$$\left| 1 + \frac{\alpha_k s}{2} \right| > \left| \frac{\alpha_k s}{2} \right|, \qquad \text{Re } s \geqslant 0$$

As $A_{k-1}(s)$ has all zeros in the left half plane, we can apply Lemma 3.1. Hence

$$| A_{k-1}(s) | \geqslant | A_{k-1}(-s) | \qquad \text{Re } s \geqslant 0$$

Combining the two inequalities given above, we find

$$\left| 1 + \frac{\alpha_k s}{2} \right| | A_{k-1}(s) | > \left| \frac{\alpha_k s}{2} \right| | A_{k-1}(-s) |, \qquad \text{Re } s \geqslant 0$$

Hence

$$| A_k(s) | = \left| \left(1 + \frac{\alpha_k s}{2}\right) A_{k-1}(s) - (-1)^k \frac{\alpha_k s}{2} A_{k-1}(-s) \right|$$

$$\geqslant \left| 1 + \frac{\alpha_k s}{2} \right| | A_{k-1}(s) | - \left| \frac{\alpha_k s}{2} \right| \cdot | A_{k-1}(-s) | > 0, \text{ Re } s \geqslant 0$$

The polynomial $A_k(s)$ thus does not have any zeros in the right half plane, and the proof of Theorem 3.1 is now completed.

By a repeated application of Theorem 3.1, we find that if the polynomial $A(s)$ has all roots in the left half plane, then all the polynomials $A_k(s)$, $k = n - 1, n - 2, \ldots, 0$ have also roots in the left half plane, and all the coefficients a_1^k are positive. Conversely, if all coefficients a_1^k are positive, we find that the polynomial $A(s)$ has all its roots in the left half plane. Hence we have Theorem 3.2.

THEOREM 3.2 (Routh)

Let $a_0^n > 0$, then the following conditions are equivalent

1. The polynomial $A(s)$ has all zeros in the left half plane.
2. All coefficients a_1^k are positive.

Main Result

We will now show how the integral (3.1) can be computed recursively. To do so we introduce

$$I_k = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{B_k(s) B_k(-s)}{A_k(s) A_k(-s)} \, ds$$

where the polynomials A_k and B_k are defined by (3.9) and (3.10). We observe that $I_n = I$. The main result is Theorem 3.3.

THEOREM 3.3

Assume that the polynomial A has all its roots in the left half plane. Then

$$I_k = I_{k-1} + \frac{\beta_k^2}{2\alpha_k} \qquad k = 1, 2, \cdots, n$$

$$I_0 = 0$$

Proof

The proof is based on elementary properties of analytic functions. As the coefficients α_k and β_k are continuous functions of the coefficients of the polynomial, it is sufficient to prove the theorem for the special case when all roots of $A_k(s)$ and $\tilde{A}_k(s)$ are distinct.

As the polynomial $A(s)$ has all zeros in the left half plane, it follows from Theorem 3.1 that all coefficients a_1^k are positive. The polynomials A_k and B_k can then be defined by (3.9) and (3.10). It also follows from Theorem 3.1 that all polynomials $A_k(s)$ have their zeros in the left half plane. We also observe that the polynomial $\tilde{A}_k(s)$ has all its zeros on the imaginary axis. It follows from (3.6) that

$$\tilde{A}_k(-s) = (-1)^{k-1} \tilde{A}_k(s)$$

and Lemma 3.1 gives

$$| \tilde{A}_k(s) | = \frac{1}{2} | A_k(s) - (-1)^k A_k(-s) | > \frac{1}{2} (| A_k(s) | - | A_k(s) |) = 0, \qquad \text{Re } s > 0$$

It follows from (3.4), (3.5), and (3.6) that

$$A_k(-s) = \bar{A}_k(-s) + \tilde{A}_k(-s) = (-1)^k\bar{A}_k(s) + (-1)^{k-1}\tilde{A}_k(s)$$
$$= (-1)^k[\bar{A}_k(s) - \tilde{A}_k(s)] = (-1)^k A_k(s) + 2\tilde{A}_k(-s) \qquad (3.20)$$

Now consider the functions

$$\frac{B_{k-1}(s)B_{k-1}(-s)}{A_{k-1}(s)A_{k-1}(-s)} \qquad (3.21)$$

and

$$\frac{B_{k-1}(s)B_{k-1}(-s)}{A_{k-1}(s)2\tilde{A}_k(-s)} \qquad (3.22)$$

These functions have the same poles s_i in the left half plane which are given by $A_{k-1}(s_i) = 0$. The function (3.21) also has poles in the right half plane and the function (3.22) has poles on the imaginary axis.

At the poles which are strictly in the left half plane we have

$$A_{k-1}(s_i) = A_k(s_i) - \alpha_k s_i \tilde{A}_k(s_i) = 0 \qquad (3.23)$$
$$A_{k-1}(-s_i) = A_k(-s_i) + \alpha_k s_i \tilde{A}_k(-s_i)$$
$$= A_k(-s_i) + \alpha_k s_i (-1)^{k-1}\tilde{A}_k(s_i)$$
$$= A_k(-s_i) + (-1)^{k-1}A_k(s_i) = 2\tilde{A}_k(-s_i)$$

where the first equality follows from (3.9), the second from (3.6), the third from (3.23), and the fourth from (3.6).

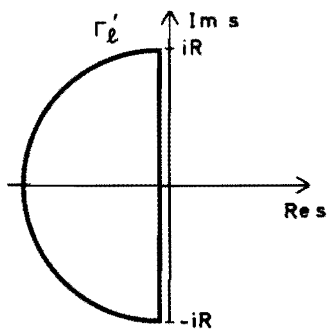

Fig. 5.2. The contour Γ_l is the limit of Γ_l' as $R \to \infty$.

As $A_{k-1}(s)$ has simple poles, the functions (3.21) and (3.22) have the same residues at the poles s_i. Integrating the functions (3.21) and (3.22) around a contour Γ_l (Fig. 5.2), which consists of a straight line slightly to the left of the imaginary axis and a semicircle with this line as a diameter, we find

$$I_{k-1} = \frac{1}{2\pi i}\int_{-i\infty-\varepsilon}^{i\infty-\varepsilon} \frac{B_{k-1}(s)B_{k-1}(-s)}{A_{k-1}(s)A_{k-1}(-s)}ds$$

$$= \frac{1}{2\pi i} \int_{-i\infty-\varepsilon}^{i\infty-\varepsilon} \frac{B_{k-1}(s)B_{k-1}(-s)}{A_{k-1}(s)2\tilde{A}_k(-s)} ds, \qquad \varepsilon > 0 \tag{3.24}$$

because the integrands tend to zero as $|s|^{-2}$ for large s and the integrals along the semicircle thus vanish.

Now consider the functions

$$\frac{B_{k-1}(s)B_{k-1}(-s)}{A_{k-1}(s)2\tilde{A}_k(-s)} \tag{3.25}$$

and

$$\frac{B_{k-1}(s)B_{k-1}(-s)}{A_k(s)2\tilde{A}_k(-s)} \tag{3.26}$$

These functions have the same poles on the imaginary axis, namely at the zeros of $\tilde{A}_k$. They have no poles in the right half plane since the polynomials A_k and A_{k-1} have no zeros in the right half plane. Since $\tilde{A}_k(s_i) = 0$ we get

$$A_{k-1}(s_i) = A_k(s_i) - \alpha_k s_i \tilde{A}_k(s_i) = A_k(s_i)$$

If the poles s_i are distinct, we thus find that the functions (3.25) and (3.26) have the same residues at s_i. Integrating the functions (3.25) and (3.26) around a contour Γ_r which consists of a straight line slight to the left of the imaginary axis and a semicirle to the right, with the line as diameter (Fig. 5.3), we get

$$I_{k-1} = \frac{1}{2\pi i} \int_{-i\infty-\varepsilon}^{i\infty-\varepsilon} \frac{B_{k-1}(s)B_{k-1}(-s)}{A_{k-1}(s)2\tilde{A}_k(-s)} ds$$

$$= \frac{1}{2\pi i} \int_{-i\infty-\varepsilon}^{i\infty-\varepsilon} \frac{B_{k-1}(s)B_{k-1}(-s)}{A_k(s)2\tilde{A}_k(-s)} ds, \qquad \varepsilon > 0 \tag{3.27}$$

because the integrands tend to zero as $|s|^{-2}$ for large $|s|$ and the integrals along the semicircles thus vanish.

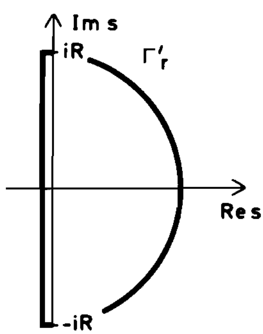

Fig. 5.3. The contour Γ_r is the limit of Γ_r' as $R \to \infty$.

Now consider the functions

$$\frac{B_{k-1}(s)B_{k-1}(-s)}{A_k(s)A_k(-s)} \tag{3.28}$$

and

$$\frac{B_{k-1}(s)B_{k-1}(-s)}{A_k(s)2\tilde{A}_k(-s)} \tag{3.29}$$

They have the same poles in the left half plane at the zeros of A_k. The function (3.28) also has poles in the right half plane and the function (3.29) has poles on the imaginary axis. At the poles in the left half plane we have

$$A_k(s_i) = 0 \tag{3.30}$$

It then follows from (3.20) that

$$A_k(-s_i) = (-1)^k A_k(s_i) + 2\tilde{A}_k(-s_i) = 2\tilde{A}_k(-s_i) \tag{3.31}$$

Since the zeros of A_k were assumed distinct, the functions (3.28) and (3.29) then have the same residues at their left half plane poles. Integrating (3.28) and (3.29) around the contour Γ_l (Fig. 5.2) we now find

$$\begin{aligned} I_{k-1} &= \frac{1}{2\pi i}\int_{-i\infty-\varepsilon}^{i\infty-\varepsilon} \frac{B_{k-1}(s)B_{k-1}(-s)}{A_k(s)2\tilde{A}_k(-s)}\,ds \\ &= \frac{1}{2\pi i}\int_{-i\infty-\varepsilon}^{i\infty-\varepsilon} \frac{B_{k-1}(s)B_{k-1}(-s)}{A_k(s)A_k(-s)}\,ds, \qquad \varepsilon > 0 \end{aligned} \tag{3.32}$$

because the integrals along the semicircle with infinite radius will vanish because the integrand tends to zero as $|s|^{-2}$ for large $|s|$.

Equation (3.10) now gives

$$\begin{aligned} I_{k-1} = {} & \frac{1}{2\pi i}\int_{-i\infty-\varepsilon}^{i\infty-\varepsilon} \frac{B_k(s)B_k(-s)}{A_k(s)A_k(-s)}\,ds - \frac{\beta_k}{2\pi i}\int_{-i\infty-\varepsilon}^{i\infty-\varepsilon} \frac{B_k(s)\tilde{A}_k(-s)}{A_k(s)A_k(-s)}\,ds \\ & - \frac{\beta_k}{2\pi i}\int_{-i\infty-\varepsilon}^{i\infty-\varepsilon} \frac{\tilde{A}_k(s)B_k(-s)}{A_k(s)A_k(-s)}\,ds + \frac{\beta_k^2}{2\pi i}\int_{-i\infty-\varepsilon}^{i\infty-\varepsilon} \frac{\tilde{A}_k(s)\tilde{A}_k(-s)}{A_k(s)A_k(-s)}\,ds \end{aligned} \tag{3.33}$$

The functions

$$\frac{B_k(s)\tilde{A}_k(-s)}{A_k(s)A_k(-s)} \tag{3.34}$$

and

$$\frac{B_k(s)\tilde{A}_k(-s)}{A_k(s)2\tilde{A}_k(-s)} \tag{3.35}$$

have the same poles in the left half plane. The poles s_i are equal to the zeros of A_k. Since these zeros were assumed distinct, it follows from (3.30) and (3.31) that the functions (3.34) and (3.35) have the same residues at these poles. Integrating (3.34) and (3.35) around the contour Γ_l (Fig. 5.2) we now find

$$\begin{aligned}\frac{1}{2\pi i}\int_{-i\infty-\varepsilon}^{i\infty-\varepsilon}\frac{B_k(s)\tilde{A}_k(-s)}{A_k(s)A_k(-s)}\,ds &= \frac{1}{2\pi i}\int_{\Gamma_l}\frac{B_k(s)\tilde{A}_k(-s)}{A_k(s)A_k(-s)}\,ds \\ &= \frac{1}{2\pi i}\int_{\Gamma_l}\frac{B_k(s)\tilde{A}_k(-s)}{A_k(s)2\tilde{A}_k(-s)}\,ds = \frac{1}{2\pi i}\int_{\Gamma_l}\frac{B_k(s)}{2A_k(s)}\,ds = \frac{1}{2}\frac{b_1^k}{a_0^k}\end{aligned} \tag{3.36}$$

where the first equality follows from the fact that the integral along the semicircle vanishes because the integrand tends to zero as $|s|^{-2}$ for large $|s|$. The second equality follows from the fact that both integrands have the same poles inside Γ_l and the same residues at these poles. The third equality is just an identity. Since A_k has all its zeros in the left half plane, the contour Γ_l can be changed to a circle around the origin without changing the value of the last integral. Observing that the integrand $B_k(s)/(2A_k(s))$ has a pole at infinity with the residue $b_1^k/(2a_0^k)$ we finally get the last equality.

We find similarly

$$\begin{aligned}\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\frac{\tilde{A}_k(s)\tilde{A}_k(-s)}{A_k(s)A_k(-s)}\,ds &= \frac{1}{2\pi i}\int_{\Gamma_l}\frac{\tilde{A}_k(s)\tilde{A}_k(-s)}{A_k(s)2\tilde{A}_k(-s)}\,ds \\ &= \frac{1}{2\pi i}\int_{\Gamma_l}\frac{\tilde{A}_k(s)}{2A_k(s)}\,ds = \frac{a_1^k}{2a_0^k}\end{aligned} \tag{3.37}$$

Equations (3.33), (3.36), and (3.37) now give

$$I_{k-1} = I_k - \frac{\beta_k}{2}\frac{b_1^k}{a_0^k} - \frac{\beta_k}{2}\frac{b_1^k}{a_0^k} + \frac{\beta_k^2}{2}\frac{a_1^k}{a_0^k} = I_k - \frac{\beta_k^2}{2\alpha_k}$$

For $k = 1$ we have

$$I_1 = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\frac{b_1^1}{a_0^1 s + a_1^1}\cdot\frac{b_1^1}{-a_0^1 s + a_1^1}\,ds = \frac{(b_1^1)^2}{2a_0^1a_1^1} = \frac{\beta_1^2}{2\alpha_1}$$

The proof of the theorem is now completed.

Computational Aspects

Having established the recursive formula given by Theorem 3.3 we will now turn to the practical aspects of the computations. To obtain the value of the integral, we must first compute the coefficients of the polynomials $A_k(s)$ and $B_k(s)$. This is conveniently done using the following tables.

$$
\begin{array}{cccccccccc}
a_0^n & a_1^n & a_2^n & a_3^n & a_4^n & \cdots & & & & \\
\mathbf{a_1^n} & 0 & a_3^n & 0 & a_5^n & \cdots & & & & \\
 & a_0^{n-1} & a_1^{n-1} & a_2^{n-1} & a_3^{n-1} & \cdots & & & & \\
 & \mathbf{a_1^{n-1}} & 0 & a_3^{n-1} & 0 & \cdots & & & & \\
 & & \vdots & & & & & & & \\
 & & a_0^2 & a_1^2 & a_2^2 & & & & & \\
 & & \mathbf{a_1^2} & 0 & & & & & & \\
 & & & a_0^1 & a_1^1 & & & & & \\
 & & & \mathbf{a_1^1} & 0 & & & & & \\
 & & & & a_0^0 & & & & &
\end{array}
\qquad
\begin{array}{cccccc}
b_1^n & b_2^n & b_3^n & b_4^n & b_5^n & \cdots \\
a_1^n & 0 & a_3^n & 0 & a_5^n & \cdots \\
 & b_1^{n-1} & b_2^{n-1} & b_3^{n-1} & b_4^{n-1} & \cdots \\
 & a_1^{n-1} & 0 & a_3^{n-1} & 0 & \cdots \\
 & & & \vdots & & \\
 & & & b_1^2 & b_2^2 & \\
 & & & a_1^2 & 0 & \\
 & & & & b_1^1 &
\end{array}
$$

Each even row in the table of a_i^k coefficients is formed by shifting the elements of the preceding row one step to the left and putting zero in the every other position. The even rows of the table on the right are identical to those of the table on the left. The elements of the odd rows of both tables are formed from the two elements immediately above, using the formulas

$$
a_i^{k-1} = \begin{cases} a_{i+1}^k & i \text{ even} \\ a_{i+1}^k - \alpha_k a_{i+2}^k & i \text{ odd}, \quad \alpha_k = a_0^k / a_1^k \end{cases} \qquad i = 0, \cdots, k-1
$$

$$
b_i^{k-1} = \begin{cases} b_{i+1}^k & i \text{ even} \\ b_{i+1}^k - \beta_k a_{i+1}^k & i \text{ odd}, \quad \beta_k = b_1^k / a_1^k \end{cases} \qquad i = 1, \cdots, k-1
$$

These are obtained by identifying coefficients of powers of s in (3.9) and (3.10).

It follows from Routh's stability test (Theorem 3.2) that the polynomial A has all its zeros in the left-half plane if all the coefficients a_1^k are positive. The coefficients a_1^k are boldfaced in the table above.

Having obtained the values α_k and β_k, the value of the integral is then given by Theorem 3.3.

$$
I = \sum_{k=1}^{n} \beta_k^2/(2\alpha_k) = \sum_{k=1}^{n} (b_1^k)^2/(2a_0^k a_1^k)
$$

As the computations are defined recursively, it is now an easy matter to obtain a computer algorithm. See the FORTRAN program (p. 139) for the computation.

Exercises

1. Evaluate the integral (3.1) for

```
      SUBROUTINE COLOSS (A, B, N, IERR, V, IN)
C
C     PROGRAM FOR EVALUATING THE INTEGRAL OF THE RATIONAL
C     FUNCTION
C        1/(2*PI*I)*B(S)*B(-S)/(A(S)*A(-S))
C     ALONG THE IMAGINARY AXIS
C
C     A-VECTOR WITH THE COEFFICIENTS OF THE POLYNOMIAL
C        A(1)*S**N + A(2)*S**(N-1) + ··· + A(N + 1)
C     IT IS ASSUMED THAT A(1) IS POSITIVE
C     B-VECTOR WITH THE COEFFICIENTS OF THE POLYNOMIAL
C        B(1)*S**(N - 1) + B(2)*S**(N - 2) + ··· + B(N)
C
C        THE VECTORS A AND B ARE DESTROYED
C
C     N-ORDER OF THE POLYNOMIALS A AND B
C     IERR-WHEN RETURNING IERR = 1 IF ALL ZEROS OF A ARE IN LEFT
C        HALF PLANE IERR = 0 IF THE POLYNOMIAL A DOES NOT HAVE
C        ALL ZEROS IN LEFT HALF PLANE OR IF A(1) IS NOT POSITIVE
C     V-THE RETURNED LOSS
C     IN-DIMENSION OF A AND B IN MAIN PROGRAM
C
C     SUBROUTINE REQUIRED
C        NONE
C
      DIMENSION A(IN), B(IN)
C
      IERR = 1
      V = 0.
      IF (A(1)) 70, 70, 10
10    DO 20 K = 1, N
      IF (A(K + 1)) 70, 70, 30
30    ALFA = A(K)/A(K + 1)
      BETA = B(K)/A(K + 1)
      V = V + BETA**2/ALFA
      K1 = K + 2
      IF(K1 - N) 50, 50, 20
50    DO 60 I = K1, N, 2
      A(I) = A(I) - ALFA*A(I + 1)
60    B(I) = B(I) - BETA*A(I + 1)
20    CONTINUE
      V = V/2.
      RETURN
70    IERR = 0
      RETURN
      END
```

$$A(s) = s^6 + 3s^5 + 5s^4 + 12s^3 + 6s^2 + 9s + 1$$
$$B(s) = 3s^5 + s^4 + 12s^3 + 3s^2 + 9s + 1$$

2. Consider the feedback system whose block diagram is shown in Fig. 5.4, where the input signal u is a Wiener process with unit variance parameter. Determine the variance of the tracking error e as a function of K, and calculate the K-value which minimizes the variance of the tracking error.

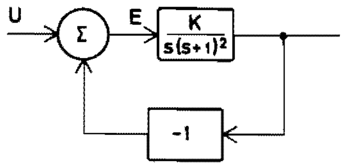

Fig. 5.4. Block diagram of the system of Exercise 2.

3. Consider the feedback system whose block diagram is shown in Fig. 5.5. The input u is a stationary process with the spectral density

$$\phi_u(\omega) = \frac{1}{\omega^2 + a^2}$$

and the measurement noise is white with the spectral density

$$\phi_n(\omega) = b^2$$

Determine the mean square deviation between the input and the output and the value of the gain parameter for which the mean square error is as small as possible.

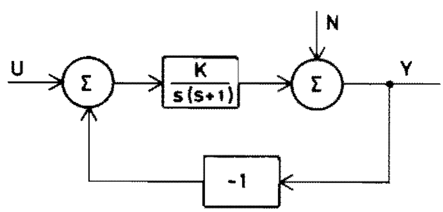

Fig. 5.5. Block diagram of the system of Exercise 3.

4. Show that for $n = 4$, the integral I defined by (3.1) can be computed as the first component x_1 of the following linear system

$$\begin{bmatrix} a_1 & a_0 & 0 & 0 \\ a_3 & a_2 & a_1 & a_0 \\ 0 & a_4 & a_3 & a_2 \\ 0 & 0 & 0 & a_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \frac{(-1)^{n+1}}{2a_0} \begin{bmatrix} b_1^2 + b_2^2 + b_3^2 + b_4^2 \\ b_1 b_2 + b_2 b_3 + b_3 b_4 \\ b_1 b_3 + b_2 b_4 \\ b_1 b_4 \end{bmatrix}$$

Notice that the matrix on the left is equal to the Hurwitz matrix for the polynomial $A(s)$.

5. Generalize the formula in Exercise 4 to arbitrary n.

6. When the function A_{k-1} is determined from A_k using (3.9), the highest power of A_k is eliminated. Show that it is possible to obtain results which are analogous to Theorem 3.3 through the following reductions

$$A_{k-1}(s) = \frac{1}{s}\left[A_k(s) - \frac{a_k{}^k}{a_{k-1}^k s}\tilde{A}_k(s)\right]$$

$$B_{k-1}(s) = \frac{1}{s}\left[B_k(s) - \frac{b_k{}^k}{a_{k-1}^k s}\tilde{A}_k(s)\right]$$

$$\tilde{A}_k(s) = \frac{1}{2}[A_k(s) - A_k(-s)]$$

Hint:

$$I_k = I_{k-1} + \frac{(b_k{}^k)^2}{2a_k{}^k a_{k-1}^k}$$

7. Derive recursive formulas for evaluating the integral

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \frac{s^k B(-s)}{A(s)A(-s)}\, ds$$

where A and B are polynomials with real coefficients

$$A(s) = a_0 s^n + a_1 s^{n-1} + \cdots + a_n$$
$$B(s) = b_0 s^m + b_1 s^{m-1} + \cdots + b_m$$

and $k + m \leqslant 2(n-1)$. The polynomial A has all its roots in the left half plane.

8. Derive recursive formulas for evaluating the integral

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \frac{B(s)}{A(s)C(-s)}\, ds$$

where

$$A(s) = a_0 s^n + a_1 s^{n-1} + \cdots + a_n$$
$$B(s) = b_0 s^m + b_1 s^{m-1} + \cdots + b_m$$
$$C(s) = c_0 s^k + c_1 s^{k-1} + \cdots + c_k$$

and $m < n + k - 2$. The polynomials A and C have all their zeros in the left half plane.

9. Show that

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\frac{1}{A(s)A(-s)}\,ds = \frac{1}{2a_1{}^1a_0{}^1}$$

10. Show that the polynomial

$$\tilde{A}(s) = \frac{1}{2}[A(s) - (-1)^n A(-s)]$$

has all its zeros on the imaginary axis.

11. Consider two stationary stochastic processes with the spectral densities

$$\begin{aligned}\phi_x(\omega) &= G_1(i\omega)G_1(-i\omega)\\ \phi_y(\omega) &= G_2(i\omega)G_2(-i\omega)\\ \phi_{xy}(\omega) &= G_1(i\omega)G_2(-i\omega)\end{aligned}$$

where

$$G_1(s) = \frac{\omega^2}{s^2 + 2\zeta\omega s + \omega^2}$$

$$G_2(s) = \frac{\omega s}{s^2 + 2\zeta\omega s + \omega^2}$$

Determine Ex^2, Ey^2, and Exy using Theorem 3.3. Also solve the same problem by first applying the representation theorem of Chapter 4 (Theorem 5.2) and then using Theorem 6.1 of Chapter 3. Compare the computational efforts in the two cases.

12. Verify that the given FORTRAN program gives the desired result.

4. RECONSTRUCTION OF STATE VARIABLES FOR DISCRETE TIME SYSTEMS

Introduction

There are in practice many situations when only a few state variables of a dynamical system can be measured directly. Consider for example the following discrete time dynamical system

$$x(t+1) = \Phi x(t) + \Gamma u(t) \tag{4.1}$$

$$y(t) = \theta x(t) \tag{4.2}$$

where x is an n-dimensional state vector, u an r-dimensional vector of inputs, and y a p-dimensional vector of output signals. The matrix Φ is $n \times n$, Γ is $n \times r$ and θ is $p \times n$. The elements of Φ, Γ, and θ may depend on t.

If the system (4.1), (4.2) is completely observable in Kalman's sense, the state vector can be reconstructed from at most n measurements of the output signal. The state variables can, however, also be reconstructed from a mathematical model of the system. Consider for example the following model

$$\hat{x}(t+1) = \Phi\hat{x}(t) + \Gamma u(t) \tag{4.3}$$

which has the same input as the original system (4.1).

If (4.3) is a perfect model, i.e., if the model parameters are identical to the system parameters, and if the initial conditions of (4.1) and (4.3) are identical, then the state $\hat{x}$ of the model will be identical to the true state variable x. If the initial conditions of (4.1) and (4.3) differ, the reconstruction $\hat{x}$ will converge to the true state variable x only if the system (4.1) is asymptotically stable. Notice, however, that the reconstruction (4.3) does not make use of the measurements of the state variables. By comparing y with $\theta\hat{x}$, it is possible to get an indication of how well the reconstruction (4.3) works. The difference $y - \theta\hat{x}$ can be interpreted physically as the difference between the actual measurements and the predictions of the measurements based on the reconstructed state variables. By exploiting the difference $y - \theta\hat{x}$, we can adjust the estimates $\hat{x}$ given by (4.3), for example, by using the reconstruction

$$\hat{x}(t+1) = \Phi\hat{x}(t) + \Gamma u(t) + K[y - \theta\hat{x}] \tag{4.4}$$

where K is a suitably chosen matrix. If the reconstructed state vector $\hat{x}$ is identical to the true state vector, the reconstructions (4.3) and (4.4) are identical and both will give the correct result. In a practical case, we might also expect (4.4) to give better results than (4.3) because in (4.4) we use the measurements as well as the input signals for the reconstruction. To get some insight into the proper choice of K, we will consider the reconstruction error $\tilde{x} = x - \hat{x}$. By subtracting (4.4) from (4.1), and using (4.2), we get

$$\tilde{x}(t+1) = \Phi\tilde{x}(t) - K[y(t) - \theta\hat{x}(t)] = [\Phi - K\theta]\tilde{x}(t) \tag{4.5}$$

If K is chosen in such a way that the system (4.5) is asymptotically stable, the reconstruction error $\tilde{x}$ will always converge to zero. Hence by introducing a feedback in the model it is possible to reconstruct state variables also in the case when the system itself is unstable. By a proper choice of K, the reconstruction error will always converge to zero for arbitrary states of (4.4).

A Parametric Optimization Problem

We have thus found that the state variables of a dynamical system can be reconstructed through the use of a mathematical model. The reconstruction contains a matrix K which can be chosen arbitrarily, subject to the constraint that the matrix Φ-$K\theta$ has all its zeros inside the unit circle. The problem now arises if there is an optimal choice of K. To pose such a problem we must introduce more structure into the problem. To do so we assume that the system is actually governed by a stochastic difference equation

$$x(t+1) = \Phi x(t) + \Gamma u(t) + v(t) \tag{4.6}$$

where $\{v(t), t \in T\}$ is a sequence of independent random n-vectors. The vector $v(t)$ has zero mean and covariance R_1. We also assume that the initial value $x(t_0)$ is normal with mean m and covariance R_0 and that there are measurement errors

$$y(t) = \theta x(t) + e(t) \tag{4.7}$$

where $\{e(t), t \in T\}$ is a sequence of independent random p-vectors. The vector $e(t)$ has zero mean and covariance R_2. The measurement errors e are assumed to be independent of v. The parameters Φ, Γ, θ, R_1, and R_2 may depend on time. Notice that even if the disturbances acting on the system are not white noise, they can often be described by a model of type (4.6) by enlarging the state space as was described in Chapter 4. A block diagram representation of the system described by the (4.6) and (4.7) is given in Fig. 5.6.

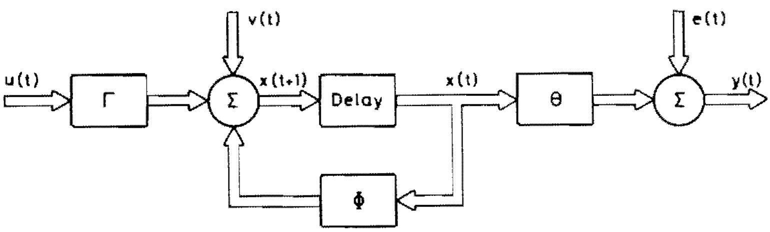

Fig. 5.6. Block diagram representation of the system described by (4.1) and (4.2).

In order to reconstruct the state variables we use the mathematical model

$$\hat{x}(t+1) = \Phi\hat{x}(t) + \Gamma u(t) + K[y(t) - \theta\hat{x}(t)] \tag{4.8}$$

A block diagram of the system (4.6), (4.7), and the reconstructor (4.8) is given in Fig. 5.7. We can now formulate a parametric optimization problem.

PROBLEM 4.1

Given an arbitrary constant vector a. Find a sequence of matrices $K(t)$ such that the error of the reconstruction of the scalar product $a^T x$ is as small as possible in the sense of mean square.

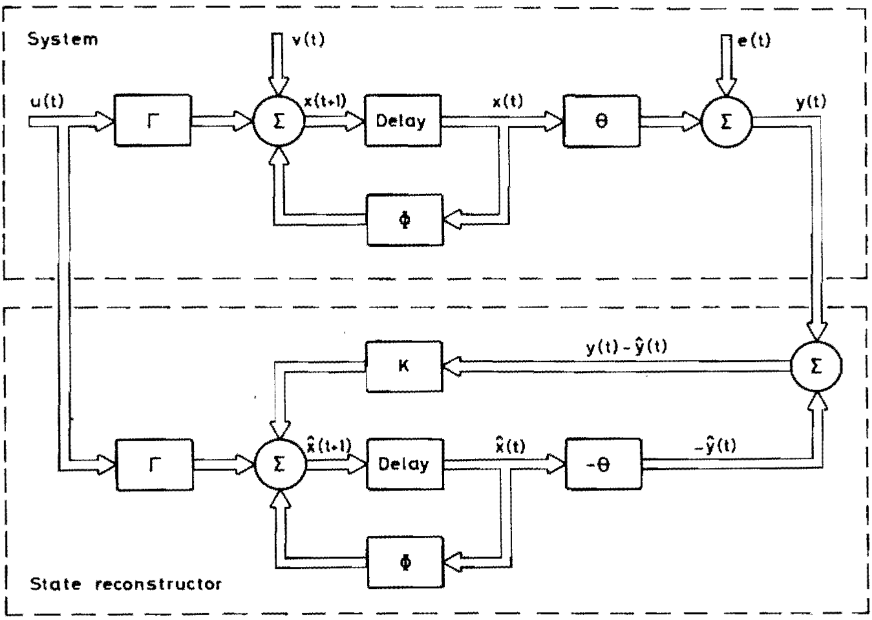

Fig. 5.7. Block diagram of system described by (4.6) and (4.7), of and state reconstructor given by (4.8).

Solution

The solution of the problem is straightforward. We first evaluate the mean and the variance of the reconstruction error, and we will then carry out the minimization. We will first derive an equation for the reconstruction error. Subtracting (4.8) from (4.6) we get

$$\tilde{x}(t+1) = x(t+1) - \hat{x}(t+1) = \Phi\tilde{x}(t) + v(t) - K[y(t) - \theta\hat{x}(t)]$$

Using (4.7) we find

$$\tilde{x}(t+1) = (\Phi - K\theta)\tilde{x}(t) + v(t) - Ke(t) \tag{4.9}$$

The reconstruction error is thus governed by a linear stochastic difference equation. The analysis of such equations was discussed in Chapter 3 (Theorem 3.1). The mean value is given by

$$E\tilde{x}(t+1) = [\Phi - K\theta]E\tilde{x}(t) \tag{4.10}$$

Hence if we choose the initial condition such that $\hat{x}(t_0) = m$, we find that $E\tilde{x}(t_0) = E(\hat{x}(t_0) - m) = 0$, and the reconstruction error has thus zero mean, irrespective of how K is chosen. The variance of the reconstruction error

$$P(t) = E[\tilde{x}(t) - E\tilde{x}(t)][\tilde{x}(t) - E\tilde{x}(t)]^T \tag{4.11}$$

is thus given by

$$P(t+1) = [\Phi - K\theta]P(t)[\Phi - K\theta]^T + R_1 + KR_2K^T$$
$$P(t_0) = R_0 \tag{4.12}$$

This equation follows from Theorem 3.1 of Chapter 3. It can also be derived directly by multiplying (4.9) with its transpose and taking mathematical expectation.

Having obtained an equation for the variance of the reconstruction error, we will now determine the gain-matrix K in such a way that the variance of the scalar product $a^T\tilde{x}$ is as small as possible. We have

$$E(a^T\tilde{x})^2 = Ea^T\tilde{x}\tilde{x}^Ta = a^T(E\tilde{x}\tilde{x}^T)a = a^TP(t)a$$

Using (4.12) we get

$$a^TP(t+1)a = a^T\{\Phi P(t)\Phi^T + R_1 - K\theta P(t)\Phi^T - \Phi P(t)\theta^TK^T + K[R_2 + \theta P(t)\theta^T]K^T\}a \tag{4.13}$$

We can now determine the time-varying gain recursively. Starting with $t = t_0$ we find that the right member of (4.13) is a quadratic function of K. By the proper choice of K we can then achieve that $P(t_0 + 1)$ is as small as possible. Having done this, we then put $t = t_0 + 1$ and we can then determine $K = K(t_0 + 1)$ in such a way that $P(t_0 + 2)$ is as small as possible. To carry out the details we rewrite (4.13) by completing the squares. We get

$$\begin{aligned} P(t+1) = {} & \Phi P(t)\Phi^T + R_1 - \Phi P(t)\theta^T[R_2 + \theta P(t)\theta^T]^{-1}\theta P(t)\Phi^T \\ & + \{K - \Phi P(t)\theta^T[R_2 + \theta P(t)\theta^T]^{-1}\}[R_2 + \theta P(t)\theta^T] \\ & \times \{K - \Phi P(t)\theta^T[R_2 + \theta P(t)\theta^T]^{-1}\}^T \end{aligned} \tag{4.14}$$

Now consider the scalar

$$\begin{aligned} a^TP(t+1)a = {} & a^T\{\Phi P(t)\Phi^T + R_1 - \Phi P(t)\theta^T[R_2 + \theta P(t)\theta^T]^{-1}\theta P(t)\Phi^T\}a \\ & + a^T\{K - \Phi P(t)\theta^T[R_2 + \theta P(t)\theta^T]^{-1}\}[R_2 + \theta P(t)\theta^T] \\ & \times \{K - \Phi P(t)\theta^T[R_2 + \theta P(t)\theta^T]^{-1}\}^Ta \end{aligned} \tag{4.15}$$

The right member is thus a function of two terms; the first term is independent of K and the second term is nonnegative because the matrix $R_2 + \theta P(t)\theta^T$ is nonnegative. The smallest value of the left member is thus obtained by choosing K in such a way that the second term of the right member of (4.14) vanishes. Doing so we find

$$K = K(t) = \Phi P(t)\theta^T[R_2 + \theta P(t)\theta^T]^{-1} \tag{4.16}$$

$$P(t+1) = \Phi P(t)\Phi^T + R_1 - \Phi P(t)\theta^T[R_2 + \theta P(t)\theta^T]^{-1}\theta P(t)\Phi^T \tag{4.17}$$

Notice that the result does not depend on a. Hence if we choose K in order to minimize the mean square reconstruction error of one linear combination of the state variables, we will at the same time minimize the mean square reconstruction error for all linear combinations. Also notice that (4.17) gives the variance of the reconstruction error for the optimal reconstruction.

The first term of the right member $\Phi P(t)\Phi^T$ shows how the reconstruction error at stage t will propagate to stage $t + 1$ through the system dynamics. The term R_1 represents the increase of the variance of the reconstruction error due to the disturbance v which acts on the system, and the third term of (4.17) shows how the reconstruction error decreases due to the information obtained from the measurements.

It follows from (4.16) and (4.17) that

$$P(t + 1) = \Phi P(t)\Phi^T + R_1 - K(t)\theta P(t)\Phi^T = [\Phi - K(t)\theta]P(t)\Phi^T + R_1$$
$$K(t)R_2 + K(t)\theta P(t)\theta^T = \Phi P(t)\theta^T$$

Postmultiplying the last equation by $K^T(t)$ and subtracting we find

$$\begin{aligned} P(t + 1) &= \Phi P(t)\Phi^T + R_1 - K(t)\theta P(t)\Phi^T - \Phi P(t)\theta^T K^T(t) \\ &\quad + K(t)R_2K^T(t) + K(t)\theta P(t)\theta^T K^T(t) \\ &= [\Phi - K(t)\theta]P(t)[\Phi - K(t)\theta]^T + R_1 + K(t)R_2K^T(t) \end{aligned}$$

From this equation we can immediately deduce that in a pure algebraical way, if $P(t)$ is nonnegative definite, then $P(t + 1)$ is also nonnegative definite.

Summarizing the results we find that the solution to the reconstruction problem is given by Theorem 4.1.

THEOREM 4.1

Consider the dynamical system (4.6) with the output signal (4.7). A reconstruction of the state variables of the system using the mathematical model (4.8) is optimal in the sense of mean square if the gain parameter K is chosen as

$$K(t) = \Phi P(t)\theta^T[R_2 + \theta P(t)\theta^T]^{-1} \tag{4.16}$$

where $P(t)$ is the variance of the optimal reconstruction which is given by

$$\begin{aligned} P(t + 1) &= \Phi P(t)\Phi^T + R_1 - \Phi P(t)\theta^T[R_2 + \theta P(t)\theta^T]^{-1}\theta P(t)\Phi^T \\ &= [\Phi - K(t)\theta]P(t)\Phi^T + R_1 \\ &= [\Phi - K(t)\theta]P(t)[\Phi - K(t)\theta]^T + R_1 + K(t)R_2K^T(t) \end{aligned} \tag{4.18}$$

with

$$P(t_0) = R_0$$

Remark 1

Notice that we have solved a parametric optimization problem, i. e., we have determined the reconstruction with the structure (4.8) which gives the smallest mean square error. Hence there might be reconstructions with a different structure which give still smaller errors. Chapter 7 will show that the chosen structure is in fact optimal.

Remark 2

It follows from the derivation that Theorem 4.1 holds when the matrices Φ, Γ, θ, R_1 and R_2 are time dependent. Writing the time dependence explicitly the model given by (4.6) and (4.7) becomes

$$x(t+1) = \Phi(t+1;\, t)x(t) + \Gamma(t)u(t) + v(t)$$
$$y(t) = \theta(t)x(t) + e(t)$$

where the covariances of $v(t)$ and $e(t)$ are $R_1(t)$ and $R_2(t)$. The optimal reconstructor is then given by

$$\hat{x}(t+1) = \Phi(t+1;\, t)\hat{x}(t) + \Gamma(t)u(t) + K(t)[y(t) - \theta(t)\hat{x}(t)]$$

where $K(t)$ is given by (4.16) and (4.18) with $\Phi = \Phi(t+1;\, t)$,

$$\Gamma = \Gamma(t), \qquad \theta = \theta(t), \qquad R_1 = R_1(t) \qquad \text{and} \qquad R_2 = R_2(t).$$

Exercises

1. Consider a system described by

$$x(t+1) = \begin{bmatrix} 1 & h \\ 0 & 1 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} e(t)$$
$$y(t) = [1 \quad 0]\, x(t)$$

where $\{e(t),\, t \in T\}$ is a sequence of independent normal (0, 1) random variables. Assume that the initial state $x(0)$ is normal with mean value

$$Ex(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and covariance

$$\operatorname{cov}\,[x(0),\, x(0)] = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

Determine the gain vector in a minimal reconstructor of the form (4.8). Also determine the covariance matrix of the reconstruction error.

2. Consider the dynamical system

$$x(t+1) = \Phi x(t) + \Gamma u(t) + v(t)$$

whose output is given by

$$y(t) = \theta x(t) + e(t)$$

where $\{e(t)\}$ and $\{v(t)\}$ are discrete time white noise with zero mean values and the covariances

$$Ev(t)v^T(s) = \delta_{s,t}R_1$$
$$Ev(t)e^T(s) = \delta_{s,t}R_{12}$$
$$Ee(t)e^T(s) = \delta_{s,t}R_2$$

Show that the state variables can be reconstructed with the mathematical model

$$\hat{x}(t+1) = \Phi\hat{x}(t) + \Gamma u(t) + K[y(t) - \theta\hat{x}(t)]$$

where the best value of K is given by

$$K = K(t) = [\Phi P(t)\theta^T + R_{12}][\theta P(t)\theta^T + R_2]^{-1}$$

where

$$P(t+1) = \Phi P(t)\Phi^T + R_1 - K(t)[R_2 + \theta P(t)\theta^T]K^T(t)$$

3. The reconstructor given by (4.8) has the property that the value of the state vector at time t is reconstructed from observations $y(t-1)$, $y(t-2)$, ... It is possible to find reconstructors which also make use of the correct observation $y(t)$ to reconstruct $x(t)$. This can be achieved by the following equation

$$\begin{aligned}\hat{x}(t+1) &= \Phi\hat{x}(t) + \Gamma u(t)\\ &\quad + K(t+1)\{y(t+1) - \theta[\Phi\hat{x}(t) + \Gamma u(t)]\}\end{aligned} \tag{*}$$

Show that if the system is governed by Eqs. (4.6) and (4.7) the optimal choice of K is given by

$$K(t) = P(t)\theta^T[R_2 + \theta P(t)\theta^T]^{-1}$$
$$P(t) = \Phi S(t-1)\Phi^T + R_1$$
$$S(t) = P(t) - K(t)\,\theta P(t)$$
$$S(t_0) = R_0$$

Also give a physical interpretation of the matrices P and S.

4. Consider the system of Exercise 1. Find the gain matrix of an optimal reconstructor of type (*). Determine the covariance matrix of the reconstruction error. Compare with the results of Exercise 1.

5. Equation (4.18) of Theorem 4.1 suggests three different ways of computing the matrix $P(t)$ recursively. Discuss the computational aspects of the different schemes.

5. RECONSTRUCTION OF STATE VARIABLES FOR CONTINUOUS TIME SYSTEMS

Introduction

We will now discuss the continuous time version of the problem of Section 4. Consider a system described by the stochastic differential equation

$$dx = Ax\,dt + Bu\,dt + dv \tag{5.1}$$

where x is an n-dimensional state vector, u an r-dimensional input vector, and v a Wiener process with incremental covariance $R_1 dt$. Let the output of the system be observed through a noisy channel described by

$$dy = Cx\,dt + de \tag{5.2}$$

where the output y is a p-vector and e is a Wiener process with incremental covariance $R_2\,dt$. The matrices A, B, C, R_1 and R_2 may be time varying. The elements are assumed to be continuous functions of time. R_2 is positive definite and R_1 positive semidefinite.

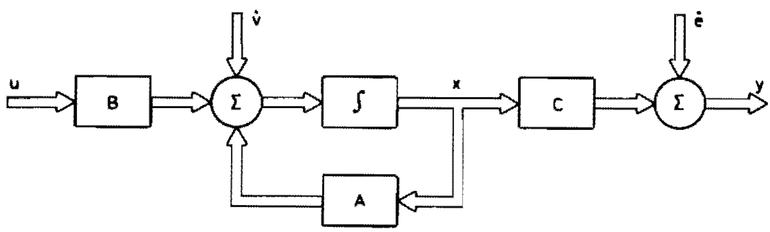

Fig. 5.8. Block diagram of a physical system which can be modeled by (5.1) and (5.2). The stochastic processes $\dot{v}$ and $\dot{e}$ have finite variances and spectral densities which are essentially constant over the frequency range $(-\omega_0, \omega_0)$ where ω_0 is considerably larger than the largest eigenvalue of A.

In Fig. 5.8 we give a block diagram of a physical system which can be modeled approximatively by (5.1) and (5.2). Using heuristic arguments analogous to those presented in Section 4, we find that it looks reasonable to try a reconstructor of the form

$$d\hat{x} = A\hat{x}\,dt + Bu\,dt + K[dy - C\hat{x}\,dt] \tag{5.3}$$

where K is an $n \times p$ matrix with time varying elements. To investigate if the model (5.3) can give a reasonable reconstruction we introduce the reconstruction error defined by

$$\tilde{x} = x - \hat{x}$$

Using (5.1) and (5.2) we find that the reconstruction error is thus a Gauss-

Markov stochastic process governed by the following stochastic differential equation

$$d\tilde{x} = (A - KC)\tilde{x}\, dt + dv - K\, de \tag{5.4}$$

To investigate the properties of the reconstructor (5.3) we thus have to analyze the stochastic differential equation (5.4). Using Theorem 6.1 of Chapter 3, we find that the mean value of the reconstruction error is given by

$$\frac{d}{dt}(E\tilde{x}) = (A - KC)(E\tilde{x}) \tag{5.5}$$

$$E\tilde{x}(t_0) = Ex(t_0) - \hat{x}(t_0) \tag{5.6}$$

and that the covariance of the reconstruction error

$$P(t) = E\{\tilde{x}(t) - E\tilde{x}(t)\}\{\tilde{x}(t) - E\tilde{x}(t)\}^T \tag{5.7}$$

is given by

$$\frac{dP}{dt} = (A - KC)P + P(A - KC)^T + R_1 + KR_2K^T \tag{5.8}$$

$$P(t_0) = E\{x(t_0) - \hat{x}(t_0)\}\{x(t_0) - \hat{x}(t_0)\}^T = R_0 \tag{5.9}$$

Hence if K is chosen in such a way that (5.5) is stable, the mean value of the reconstruction error will converge to zero. It also follows from (5.5) that if the reconstruction error has zero mean value at any time t_0, it will have zero mean value for all t.

It is instructive to give the physical interpretation of the different terms of (5.8). For this purpose we rewrite it as

$$\frac{dP}{dt} = AP + PA^T + R_1 - (KCP + PC^TK^T - KR_2K^T) \tag{5.10}$$

The first two terms represent the propagation of the covariance of the reconstruction error through the system dynamics. The term R_1 represents the increase in the covariance of the reconstruction error due to the disturbance v which acts on the system, and the last term represents the reduction of the covariance of the reconstruction error due to the measurements. The last term naturally depends on the choice of the gain matrix.

A Parametric Optimization Problem

Having found the equations which characterize the reconstruction error we will now investigate if there is a choice of K which is optimal. It is assumed that $E\tilde{x}(t_0) = 0$ which means that the reconstruction error always has a zero mean value. The mean square error of the reconstruction of the linear combination of state variables a^Tx is chosen as the criterion. Hence

$$E(a^T\tilde{x})^2 = E(a^T\tilde{x})(\tilde{x}^Ta) = a^TE\tilde{x}\tilde{x}^Ta = a^TP(t)a \tag{5.11}$$

where the last equality holds because the mean value of the reconstruction error is zero. Using the differential equation (5.8) we find

$$\frac{d}{dt} a^T P(t) a = a^T(AP + PA^T + R_1)a + a^T(KR_2K^T - KCP - PC^TK^T)a \tag{5.12}$$

To proceed we need the following result.

LEMMA 5.1

Let P and Q be solutions of the Riccati equations

$$\frac{dP}{dt} = AP + PA^T + R_1 + KR_2K^T - KCP - PC^TK^T \tag{5.13}$$

$$\frac{dQ}{dt} = AQ + QA^T + R_1 - QC^TR_2^{-1}CQ \tag{5.14}$$

where R_2 is assumed positive definite and the initial conditions are

$$P(t_0) = Q(t_0) = R_0 \tag{5.15}$$

where R_0 is symmetric. The matrix $P(t) - Q(t)$ is then positive semidefinite and

$$P(t) = Q(t) \tag{5.16}$$

for

$$K = PC^TR_2^{-1} \tag{5.17}$$

Proof

It follows from (5.13) and (5.14) that

$$\begin{aligned}\frac{d}{dt}(P - Q) &= A(P - Q) + (P - Q)A^T + KR_2K^T - KCP - PC^TK^T \\ &\quad + QC^TR_2^{-1}CQ = (A - KC)(P - Q) + (P - Q)(A-KC)^T \\ &\quad + (K - QC^TR_2^{-1})R_2(K - QC^TR_2^{-1})^T\end{aligned} \tag{5.18}$$

Let $\Psi(t; s)$ be the solution of the differential equation

$$\begin{aligned}\frac{d\Psi(t;s)}{dt} &= [A(t) - K(t)C(t)]\,\Psi(t; s) \\ \Psi(t; t) &= I\end{aligned} \tag{5.19}$$

The solution of (5.18) can then be written as

$$\begin{aligned}P(t) - Q(t) &= \int_{t_0}^{t} \Psi(t; s)[K(s) - Q(s)C^T(s)R_2^{-1}(s)]R_2(s) \\ &\quad \times [K(s) - Q(s)C^T(s)R_2^{-1}(s)]^T\,\Psi^T(t; s)\,ds\end{aligned} \tag{5.20}$$

But the matrix of the right member is always nonnegative definite for all K. Furthermore for

$$K = QC^T R_2^{-1}$$

we get $P(t) = Q(t)$ which then implies (5.16) and completes the proof of the lemma.

We thus find that there is a universal choice of K which makes $a^T P(t) a$ as small as possible for all a. The optimal value of the gain matrix is given by (5.17). The optimal value of K will thus give the smallest error for reconstructing any linear combination of the state variables. Summing up we get Theorem 5.1.

THEOREM 5.1

Let a dynamical system subject to disturbances and measurement errors be described by (5.1) and (5.2). A reconstructor of the form (5.3) is optimal in the sense of mean squares if the initial value is chosen as

$$\hat{x}(t_0) = Ex(t_0) \tag{5.21}$$

and the gain parameter K is chosen as

$$K(t) = P(t)C^T R_2^{-1} \tag{5.17}$$

where $P(t)$ is the covariance of the optimal reconstruction errors. The matrix $P(t)$ is given by the Riccati equation

$$\frac{dP(t)}{dt} = AP + PA^T + R_1 - PC^T R_2^{-1} C P \tag{5.22}$$

with the initial condition

$$P(t_0) = R_0 \tag{5.23}$$

Remark

Notice that we have solved a parametric optimization problem for a state reconstructor with the structure (5.3). Chapter 7 will show that the structure (5.3) is in fact optimal.

Exercises

1. Consider the motion of a particle on a straight line. Assume that the acceleration of the particle can be described as a white noise process with the spectral density $1/(2\pi)$ and that the position of the particle is observed with a white noise measurement error with the spectral density $r/(2\pi)$. Determine a minimal variance reconstructor of the form (5.3) for the position and velocity of the particle. Also find the covariance matrix of the reconstruction error.

2. Consider a dynamical system described by (5.1) and (5.2) where e and v are correlated Wiener processes with zero mean values and the incremental covariance

$$E\begin{bmatrix} dv \\ de \end{bmatrix}[dv^T \; de^T] = \begin{bmatrix} R_1 & R_{12} \\ R_{12}^T & R_2 \end{bmatrix} dt$$

Show that a minimum variance reconstructor with the structure (5.3) is obtained when the initial value is chosen as $\hat{x}(t_0) = Ex(t_0)$ and the gain matrix is given by

$$K(t) = [P(t)C^T + R_{12}]R_2^{-1}$$

where the covariance matrix of the reconstruction error satisfies the Riccati equation

$$\frac{dP}{dt} = [A - R_{12}R_2^{-1}C]P + P[A - R_{12}R_2^{-1}C]^T$$
$$+ R_1 - R_{12}R_2^{-1}R_{12}^T - PC^TR_2^{-1}CP$$
$$P(t_0) = R_0$$

3. A simple vertical alignment system which consists of a platform servo and an accelerometer is shown in Fig. 5.9. The accelerometer signal provides information about the vertical indication error. The signal is used to drive the platform normal to the vertical position. For small deviations the accelerometer signal is

$$y = \theta + n$$

where θ is the vertical indication error and n a disturbance due to horizontal accelerations. The platform can be described by the equation

$$\frac{d\theta}{dt} = u$$

where u is the control signal. The servo loop can be described by

$$u = -ky$$

where k is the gain. Assuming that the noise can be modeled by white noise, we find that the system can be described by the stochastic differential equation

$$d\theta = -k\theta\, dt + k\, dv$$

where $\{v(t)\}$ is a Wiener process with variance parameter r. Assume that the initial state is normal $(0, \sigma)$; find a time varying gain $k = k(t)$ such that the variance of the vertical alignment error is as small as possible. Compare the results with those obtained for a constant gain when the alignment period is kept constant T; see Fig. 5.10.

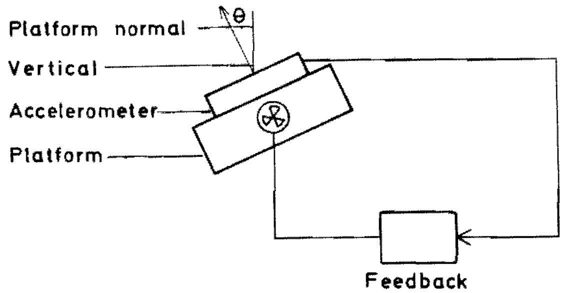

Fig. 5.9. Schematic diagram of a vertical alignment system.

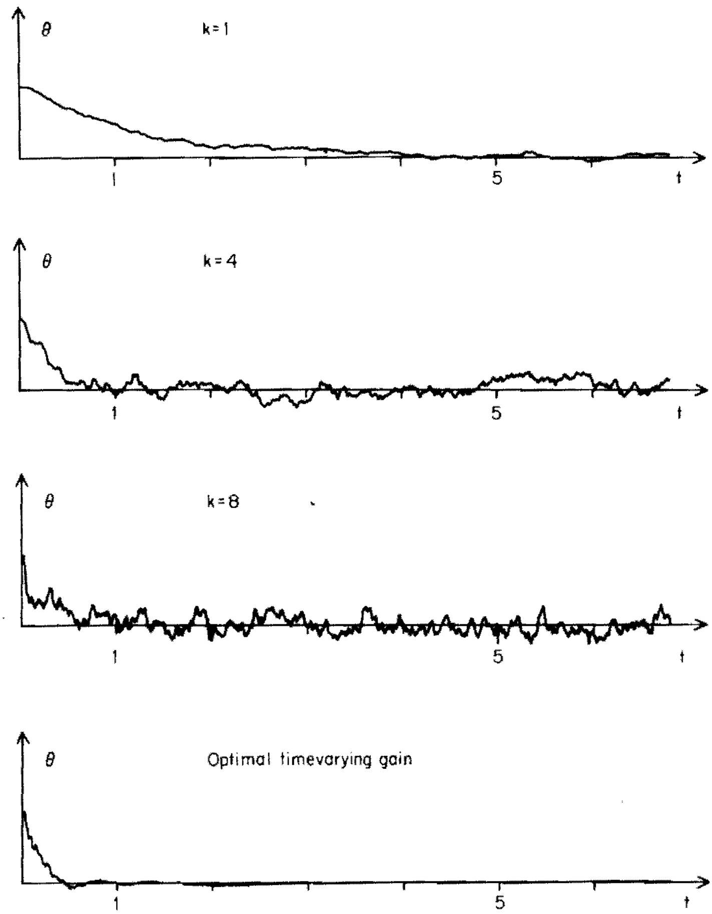

Fig. 5.10. Vertical alignment error θ versus time t for analog simulation of systems with different fixed gains and with optimal time varying gain.

4. Let P and Q be solutions to the Riccati equations

$$\begin{cases} \dfrac{dP}{dt} = AP + PA^T + R_1 - PC^TR_2^{-1}CP \\ P(t_0) = P_0 \end{cases}$$

$$\begin{cases} \dfrac{dQ}{dt} = AQ + QA^T + R_1 - QC^TR_2^{-1}CQ \\ Q(t_0) = Q_0 \end{cases}$$

Show that $P_0 > Q_0$ implies $P(t) > Q(t)$ for all t for which the Riccati equations have solutions.

5. Consider the Riccati equations

$$\begin{cases} \dfrac{dP}{dt} = AP + PA^T + R_1 - PC^TR_2^{-1}CP \\ P(t_0) = P_0 \end{cases}$$

$$\begin{cases} \dfrac{dQ}{dt} = AQ + QA^T + R_3 - QC^TR_2^{-1}CQ \\ Q(t_0) = P_0 \end{cases}$$

Show that $R_1 \geqslant R_3$ implies $P(t) \geqslant Q(t)$ for all t for which the Riccati equations have solutions.

6. Consider the Riccati equations

$$\begin{cases} \dfrac{dP}{dt} = AP + PA^T + R_1 - PC^TR_2^{-1}CP \\ P(t_0) = P_0 \end{cases}$$

$$\begin{cases} \dfrac{dQ}{dt} = AQ + QA^T + R_1 - QC^TR_3^{-1}CQ \\ Q(t_0) = P_0 \end{cases}$$

Show that $R_2 \geqslant R_3$ implies $P(t) \geqslant Q(t)$ for all t for which the Riccati equations have solutions.

6. BIBLIOGRAPHY AND COMMENTS

The idea of parametric optimization of stochastic systems is discussed in

James, H. M., Nichols, N. B., and Phillips, R. S., *Theory of Servomechanisms*, McGraw-Hill, New York, 1947

Newton, G. C., Jr., Gould, L. A., and Kaiser, J. F., *Analytical Design of Linear Feedback Controls*, Wiley, New York, 1957.

In these textbooks the integrals for the variance of the signals are evaluated using straightforward residue calculus. Tables for the integrals in the continuous time case are also given in these references.

Analogous results for discrete time systems are given in

Jury, E. I., *Theory and Application of the z-Transform Method*, Wiley, New York, 1964.

Theorem 2.2 is essentially the Schur-Cohn theorem for stability of a discrete time linear system. The proof presented is Section 2 is due to

Růžička, J., "Algebraická Kritéria Stability Impulsních Soustav," (Algebraic criteria for stability of sampled data systems, in Czech), *Strojnícky Časopis* XIII, č 5, 395–403 (1962).

The proof is also given in

Strejc, V., *Syntese von Regelungssystemen mit Prozessrechner*, Akademie-Verlag, Berlin, 1967.

Closely related theorems are also given in

Jury, E. I., "On the Roots of Real Polynomial Inside the Unit Circle and a Stability Criterion for Linear Discrete Systems," *Proc. of Sec. Congr. of IFAC*, Butterworths, London, 1964.

Toma, M., "Ein Einfaches Verfahren zur Stabilitätsprüfung von linearen Abtastsystemen," *Regelungstechnik* **10**, 302–306 (1962).

Theorem 3.2 is in essence the Routh-Hurwitz theorem. The proof given in Section 3 which is patterned after the discrete time case is believed to be new.

The idea of evaluating the integrals recursively is due to Nekolný. See, e.g.,

Nekolný, J., *Nova Jednoduchá Methodika Testu Jakosti Regulace*, Prague, 1957

Nekolný, J. and Beneš, J., "Simultaneous Control of Stability and Quality of Adjustment-Application of Statistical Dynamics," in Coales *et al.* "Automatic and Remote Control," *Proc. of First IFAC Congr. Moscow 1960*, Vol. 2, Butterworths, London, 1961.

The algorithm for the discrete time case Theorem 2.3 is given in

Åström, K. J., "Recursive Formulas for the Evaluation of Certain Complex Integrals," Rep. 6804, Lund Institute of Technology, Division of Automatic Control. September 1968.

The corollary of Theorem 2.3 and a computer algorithm is given in

Åström, K. J., Jury, E. I., and Agniel, R. G., "A Numerical Method for the Evaluation of Complex Integrals," *IEEE Trans.* AC (1970).

Algol programs for the continuous time algorithms are found in

Peterka, V. and Vidinčev, P., "Rational-Fraction Approximation of Transfer Functions," *Proc. IFAC Symp. on Identification in Autom. Control Sys.* Prague, 1967.

The idea of reconstructing the state of a dynamical system using a mathematical model discussed in Sections 4 and 5 is probably very old. It was, e.g., encountered in discussions with John Bertram in 1961. An early explicit reference to a model reconstructor without feedback from the measurements is given in

Kalman, R. E. and Bertram, J. E., "General Synthesis Procedure for Computer Control of Single and Multiloop Linear Systems," *Trans. AIEE* **77** (1958).

A detailed discussion with several refinements is given in

Luenberger, D. G., "Observing the State of a Linear System," *IEEE Trans. on Military Electron.* **8**, 74-80 (1964).
Luenberger, D. G., "Observers for Multivariable Systems," *IEEE Trans. AC* **11**, 190-191 (1966).

The recursive equations given by Theorem 4.1 and Theorem 5.1 are identical to the Kalman Bucy filtering algorithms. See

Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," *ASME J. Basic Eng.* **82**, 35-45 (1960)
Kalman, R. E. and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory," *ASME J. of Basic Eng.* **83**, 95-107, (1961).

Also compare with the results of Chapter 7.