CONTROL AND ESTIMATION FOR

LARGE-SCALE SYSTEMS HAVING SPATIAL SYMMETRY

by

JOSEPH EDWARD WALL, JR.

B.S.E., Tulane University
(1973)

S.M., Massachusetts Institute of Technology
(1976)

E.E., Massachusetts Institute of Technology
(1976)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August, 1978

Signature of Author .....
Department of Electrical Engineering and Computer Science,
August 11, 1978

Certified by .....
Thesis Supervisor

Certified by .....
Thesis Supervisor

Accepted by .....
Chairman, Department Committee

CONTROL AND ESTIMATION FOR

LARGE-SCALE SYSTEMS HAVING SPATIAL SYMMETRY

by

JOSEPH EDWARD WALL, JR.

Submitted to the Department of Electrical Engineering and Computer Science on August 11, 1978, in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

ABSTRACT

Control and estimation problems for circulant and Toeplitz systems are studied using spatial transform techniques. In the finite-dimensional case, the discrete Fourier transform is used to provide a complete treatment of the system theoretic properties of circulant systems in terms of lower dimensional, transformed subsystems. The centralized and decentralized control and estimation problems for circulant systems are approached in the spatial frequency domain. Efficient off-line and on-line solutions to the optimal centralized problem are obtained. For the decentralized problem, suboptimal design procedures are proposed by analogy with the design of finite impulse response digital filters.

In the infinite-dimensional case, the z-transform is employed to solve the Toeplitz estimation problem. Motivated by recent work in the image processing field dealing with recursive estimation based on two-parameter models, the update step of a discrete-time Kalman filter for a Toeplitz system is shown to be equivalent to a smoothing problem. An investigation of the smoothing problem yields new insight into the two-filter smoother and yields formulas for sensitivity analysis and reduced order smoother analysis. This new two-filter smoother is then used to perform the filter update step for Toeplitz systems. Some implementation issues of such a filter are then discussed. Finally, the control problem dual of the fixed-interval smoothing problem is posed and solved.

Spatial transformations play a crucial, fundamental role throughout this dissertation. The spatial frequency domain is found to be quite appealing when addressing control and estimation problems for spatially symmetric large-scale systems.

THESIS SUPERVISORS:

Professor Alan S. Willsky
Title: Associate Professor of Electrical Engineering

Professor Nils R. Sandell, Jr.
Title: Associate Professor of Electrical Engineering

ACKNOWLEDGEMENTS

I gratefully acknowledge the inestimable contribubtion of my two thesis supervisors, Professor Alan S. Willsky and Professor Nils R. Sandell, Jr. Professor Willsky has provided constant enthusiasm and a continuous stream of ideas for this work. The thoughtful and thought-provoking comments and questions of Professor Sandell have contributed greatly to this research. The friendship of these two men has also been important to me as a graduate student at M.I.T., and is thankfully acknowledged.

The interest and suggestions of Dr. Alan J. Laub and Dr. David A. Castañon, thesis readers, is also greatly appreciated.

The efforts of Ms. Barbara Peacock, Mrs. Evelyn Holmes and Ms. Fifa Monserrate in typing this manuscript have not been insignificant and are sincerely appreciated. Thanks are also due to Mr. Arthur Giordani for his excellent work and cooperation in preparing the figures in this document.

My parents. Joseph and Marjorie Wall, have steadily provided me with generous amounts of advice, support, and encouragement, and I wish to take this opportunity to express my deep thanks to them.

I also wish to thank my wife, Barbara, and my children, Joey and Susan, for their most special contribution to this research effort.

DEDICATED

TO

BARBARA, JOEY, SUSAN

TABLE OF CONTENTS

Table of Contents (continued)

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivation

There has been a great deal of recent activity in the area of large-scale systems [1]-[3]. These systems are found in such diverse fields as power systems [4]-[6], transportation systems [7]-[9], econometric systems [10], [11], and packet switched data networks [12]-[14]. For large-scale systems, the control and estimation problems are often of such great complexity that the standard modern techniques are computationally intractable. This intractability may be because of either the on-line or off-line computational requirements. One concludes that the adjective large as used in "large-scale systems" usually has the meaning of too large.

Various techniques have been proposed to reduce the computational burden for control and estimation problems by exploiting special structural properties frequently found in large-scale systems. For example, singular perturbation theory has been successfully employed to construct simplified controllers and estimators for systems having multiple time scales [15]-[18]. Also, systems composed of weakly coupled subsystems have been attacked by nonsingular perturbation theory to obtain decentralized controllers [19], [20] or off-line computational savings [21], [22]. Finally, there have been numerous approaches for determining the stability of a large-scale system on the basis of the properties of its individual subsystems and the nature of their interactions.

Decomposition and decentralization are two crucial elements in large-scale system theory. Decentralization is directed toward the reduction of on-line computational requirements and intersubsystem communication. The price of the advantages, however, is often increased off-line complexity. As pointed out in [23], other proposed benefits of decentralization, such as increased reliability or increased adaptability, may be more imagined than real. The decomposition issue concerns reducing the off-line computational burden associated with obtaining a desired controller or estimator. A discussion of decomposition vis-a-vis decentralization is found in [24].

It is often natural and useful to view a large-scale system as an interconnection of much simpler subsystems. In many cases these subsystems are actually distinct physical entities, while in other cases they are merely chosen for mathematical convenience. Decomposition procedures are frequently based at the subsystem level. The nonsingular perturbation decomposition and the stability tests for interconnected systems, for instance, are based on a tearing of the system into subsystems which are usually distinct entities. The singular perturbation methods, on the other hand, usually involve fast and slow subsystems which are chosen on the basis of physical insight for mathematical convenience. In the case of a decentralized controller or estimator, the partitioning of inputs and outputs is also frequently done at the subsystem level.

Concurrent with this work in large-scale systems has been the

substantial activity in the image processing field. The estimation of discrete-space images from observations corrupted by additive noise using techniques such as two-dimensional Wiener filtering result in enormous computational burdens. Thus one of the primary objectives of work in this area is the discovery of computationally tractable estimation formulas. One approach toward the recursive estimation of images has been motivated by the success of model based estimators such as the Kalman filter [25]. Of interest here is the use of two-dimensional models for the image process [26], [27].

Attasi [28] has considered least squares recursive estimation of an image $z(i,k)$ under noisy observations

$$y(i,k) = z(i,k) + v(i,k) \tag{1.1}$$

where the image is generated by the two-parameter model

$$\begin{aligned} x(i,k) &= F_1\, x(i-1,k) + F_2\, x(i,k-1) - F_1F_2\, x(i-1,k-1) + w(i-1,k-1) \\ z(i,k) &= H\, x(i,k) \end{aligned} \tag{1.2}$$

with the requirement $F_1F_2 = F_2F_1$. The problem is to determine the optimal estimate $\hat{x}(i,k)$ of $x(i,k)$ given the observations $y(m,n)$ for $m \leq i$ and all n. This estimate is shown to be obtained from a two step procedure. First a predicted value $\tilde{x}(i,k)$ is computed from the estimates $\hat{x}(i-1,n)$ for all n, as

$$\tilde{x}(i,k) = F_1\, \hat{x}(i-1,k). \tag{1.3}$$

The error $e(i,k)$ is now defined as $e(i,k) = x(i,k) - \tilde{x}(i,k)$. Then the estimation problem is solved by

$$\hat{x}(i,k) = x(i,k) + \hat{e}(i,k) \tag{1.4}$$

where $\hat{e}(i,k)$ is the solution of the one-parameter smoothing problem to estimate $e(i,n)$, for all n, given $y(i,n)$ for all n [28]. This smoothing problem is solved by two Kalman filters, one moving in the positive n direction and one in the negative n direction.

Two-parameter models such as Attasi's have had mixed success when used for recursive estimation of images. Attasi's model will now be considered in the context of large-scale systems.

Suppose that the vector quantity $x(i,k)$ in Attasi's model is interpreted as the state of subsystem k at time i. The term state is used loosely here since the dynamics (1.2) are not standard state space dynamics. Nevertheless, view Attasi's model as an infinite-dimensional system propagating in time. His estimation problem, then, is just the filtering problem for this infinite-dimensional system given observations up to the present. The update step of this filtering problem is solved by one Kalman filter moving up the line of subsystems and one Kalman filter moving down the line. The Kalman filter moving up the line can be implemented by having each subsystem (say subsystem k) perform a measurement update and then transmit this estimate to the next $(k+1)$ subsystem. Likewise, implementation of the Kalman filter moving down the line involves a measurement update at each subsystem

and then communication of this estimate to the previous (k-1) subsystem. This update procedure is not decentralized, but it is quite efficient and interesting.

As discussed by Willsky [29], the proof of Attasi's estimation algorithm employs a bilateral z-transform along the k direction and treats the i direction as the time variable. In terms of the interpretation of the model as an infinite-dimensional system, this corresponds to taking a z-transform with respect to subsystem index (essentially a spatial z-transform). In this manner, Attasi obtains independent subproblems indexed by the variable z. That is to say, this problem is very nicely decomposed by the z-transform.

The same approach of taking spatial z-transforms has also been used to address control problems for infinite-dimensional one-parameter systems of the form

$$x(i,j) = \sum_{k=-\infty}^{+\infty} A_{j-k}\, x(i-1,k) + B_{j-k}\, u(i-1,k) \qquad (1.5)$$

Melzer and Kuo [30] design optimal centralized regulators in this manner for spatially invariant quadratic cost functions. Optimal constrained decentralized regulators for the same class of cost functions are similarly designed by Chu [31]. The important result, at least in the centralized case, is that the optimal control problem decomposes into a set of optimal control problems of dimension equal to the dimension of the substates $x(i,j)$ and indexed by the variable z.

In both the estimation problem of Attasi and the control problems of Melzer and Kuo and of Chu, the key step is the use of the spatial z-transform to decompose the problem. It is possible to use the z-transform in these cases because the systems are spatially invariant, i.e. all the subsystems are identical and the influence of subsystem k on subsystem ℓ depends on only $k-\ell$. The objective of this thesis is to examine in depth large-scale systems possessing the structural property of spatial symmetry. Specifically, the control and estimation problems for infinite-dimensional Toeplitz systems [see (1.5)] and their finite-dimensional analog are studied. Such finite-dimensional systems are called circulant systems. In the case of circulant systems, the discrete Fourier transform will be found to be the analog of the z-transform used for Toeplitz systems.

The spatially symmetric systems studied in this thesis are obviously an extremely special type of large-scale system. The purpose in studying such a special class of systems is to determine just how far one can go in exploiting spatial symmetry to obtain efficient on-line implementations of controllers and estimators, or separating a large problem into more tractable subproblems. The issue of decentralization and decomposition, therefore, are crucial throughout the thesis. Also, for spatially invariant systems one can study such phenomena as the spatial propagation of disturbances which are obviously present here.

There has apparently been only very limited work done on circulant systems. Dickerson and Erickson [32] have obtained some weak stability results concerning circulant systems, and these are analyzed in the sequel. Some areas where circulant matrices have been used include the study of certain binary codes [33], [34], the generation of Markov chains used as digital signals [35], the generalization of Clarke components for polyphase networks [36], and the spherical model of a ferromagnet [37]. Circulant matrices have had extensive use in the field of digital image processing [38]-[40].

Let the image radiant energy at the point (x,y) be represented as g(x,y) and the object radiant energy as f(x,y). Then the image and object distributions are modelled as obeying an integral equation involving the point spread function h as follows:

$$g(x,y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x,y,u,v,f(u,v))\,dudv \tag{1.6}$$

By making the simplifying assumptions that

(i) h acts as a scalar multiplier, i.e.

$$h(x,y,u,v,f(u,v)) = h(x,y,u,v)f(u,v)$$

(ii) h is spatially invariant, i.e. $h(x,y,u,v) = h(x-u;y-v)$

one reduces (1.6) to the two-dimensional convolution

$$g(x,y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x-u,y-v)f(u,v)\,dudv \tag{1.7}$$

The image restoration problem is to estimate f from possibly noisy measurements of g. One digital approach to this problem is to sample g at points on a rectangular grid and then form a vector g_i by lexicographically ordering these samples. A vector f_i is similarly obtained from f and the relationship expressed by (1.7) may be approximated as

$$g_i = H_T f_i \tag{1.8}$$

where the matrix H_T is block Toeplitz with Toeplitz blocks [40]. By using a circulant approximation H_C to H_T, effective and computationally tractable algorithms for the inversion of (1.8) have been obtained [38], [39]. The key to these algorithms is the use of the fast Fourier Transform to diagonalize circulant matrices. This same diagonalization procedure is used heavily in the sequel to study circulant systems.

1.2 Summary

Chapter 2 introduces circulant systems and develops many of their properties. Circulant systems are linear systems defined in terms of (block) circulant matrices. Such matrices are discussed in Appendix A where it is shown that the eigenvectors of a circulant matrix are fixed by its dimension. This property is then used to develop the diagonalizing property of the discrete Fourier transform. Using this transformation, various system theoretic results are obtained for circulant systems, and circulant Riccati and Lyapunov are efficiently decomposed. Chapter 2 also shows how symmetric and antisymmetric block tridiagonal systems can

be imbedded in circulant systems roughly two or four times as large. Further, examples of circulant systems and tridiagonal systems that can be imbedded in circulant systems are given.

The control and estimation problems for circulant systems are the subject of Chapter 3. For the most part, Chapter 3 deals directly with the control problem; the estimation problem is treated by duality in Section 3.5. The centralized linear-quadratic control problem is shown to (i) decompose into low order control problems and (ii) have an efficient on-line implementation employing parallel processing. Both these results are obtained by using the spatial transformation introduced in Chapter 2. The fixed structure decentralized control problem is treated in Section 3.2. Necessary conditions for the optimal decentralized gains are obtained but are not found to decompose. Suboptimal decentralized controllers are then proposed by considering the analogous situation of the design of finite impulse response digital filters. Various digital filter design techniques are discussed for designing decentralized control gains. A computer example of circulant control for a rectangular membrane is also included. The possibility of using a circulant control law for a general large-scale system is examined in Section 3.4. Throughout Chapter 3 an attempt is made to use the spatial transform concepts to obtain centralized and decentralized controllers. That this goes much further than just using transforms to decompose the centralized problem can be clearly seen in Section 3.2.3. In this section, the spatial frequency

viewpoint is essential for understanding the suboptimal decentralized control laws that are presented.

It has been stated that the update step for Attasi's estimation problem is equivalent to a smoothing problem. This statement will be generalized in Chapter 5. The purpose of Chapter 4, however, is to carefully study the fixed-interval smoothing problem. In particular, the Mayne-Fraser two-filter smoother is studied here. Section 4.2 presents an historical review of the two-filter smoother, discussing the work of Mayne [41], Fraser [42], and Mehra [43]. By using reversed-time Markov models, a new solution to the fixed-interval smoothing problem is obtained which clearly demonstrates the use of (i) a priori data, (ii) past measurements, and (iii) future measurements in computing the smoothed estimate. Using this new solution, a sensitivity analysis and an analysis of reduced order smoothers are performed. Also, using the insight gained in this approach, a new change of initial conditions formula for the smoothed estimate is obtained.

Chapter 5 deals with the control and estimation problems for infinite-dimensional Toeplitz systems. The view here is to consider explicitly the filtering problem and then treat the optimal control problem by duality in Section 5.4. After defining Toeplitz systems, the work of Melzer and Kuo [30] and Chu [31] on optimal control of Toeplitz systems is reviewed. Attasi's [28] estimation problem is then considered, and the model (1.2) is shown to correspond to a Toeplitz system. Motivated by Attasi's work, the filtering problem

for general Toeplitz systems is then treated. It is shown that the update step in this case is equivalent to a smoothing problem along the subsystems. Finite-dimensional realizations of the update operation are presented and the implications of these realizations for filtering in large-scale systems are discussed. Also, some filter implementation issues are examined. Chapter 5 tries to give a cohesive treatment of Toeplitz systems and the associated control and estimation problems by employing a spatial z-transform. This treatment is much deeper than that of Melzer and Kuo or Chu in that the spatial transform is not merely used for decomposition purposes. Rather, the spatial frequency domain provides the necessary insight for the proposed filters and controllers in Chapter 5.

In conclusion, Chapter 6 presents the contributions of this thesis and suggestions for future research. In listing the contributions of this work, a summary of the thesis is also provided.

CHAPTER 2

CIRCULANT SYSTEMS

2.1 Introduction to Circulant Systems

2.1.1 Definition

A circulant matrix is a square N×N matrix in which each row is a circular right shift of the row directly above it, i.e. a matrix of the form

$$A = \begin{pmatrix} a_0 & a_{N-1} & a_{N-2} \cdots & a_1 \\ a_1 & a_0 & a_{N-1} \cdots & a_2 \\ a_2 & a_1 & a_0 \cdots & a_3 \\ \vdots & \vdots & \vdots & \vdots \\ a_{N-1} & a_{N-2} & a_{N-3} \cdots & a_0 \end{pmatrix} \qquad (2.1)$$

The right shift of each row is called a circular shift because the element that is shifted out on the right side re-enters the matrix on the left. Block circulant matrices are defined similarly, with the elements a_k being replaced by submatrices A_k. A matrix is called block circulant of order N if it can be partitioned into N^2 blocks A_k such that the resulting structure is the same as (2.1).

Deterministic continuous-time circulant systems are defined in terms of block circulant matrices as follows: the state of a circulant system evolves according to the differential equation

$$\frac{d}{dt} x(t) = A\, x(t) + B\, u(t) \tag{2.2}$$

and the output is given by

$$y(t) = C\, x(t) \tag{2.3}$$

where

$x(t) \in \mathbb{R}^{nN}$ is the state vector

$u(t) \in \mathbb{R}^{mN}$ is the input vector

$y(t) \in \mathbb{R}^{pN}$ is the output vector

$A \in \mathbb{R}^{nN \times nN}$ is a block circulant matrix of order N

$B \in \mathbb{R}^{nN \times mN}$ is a block circulant matrix of order N

$C \in \mathbb{R}^{pN \times nN}$ is a block circulant matrix of order N

This circulant system is just a finite-dimensional linear system for which the system, input, and output matrices are all block circulant. Discrete-time circulant systems are similarly defined.

The state x(t) of a circulant system may be viewed as consisting of N substates $x_k(t) \in \mathbb{R}^n$, k=0,1,...,N-1, according to the partition

$$x(t) = \begin{pmatrix} x_0(t) \\ x_1(t) \\ x_2(t) \\ \vdots \\ x_{N-1}(t) \end{pmatrix} \tag{2.4}$$

Likewise, the input u(t) and the output y(t) consist of subinputs $u_k(t) \in \mathbb{R}^m$ and suboutputs $y_k(t) \in \mathbb{R}^p$. The system equations (2.2) and (2.3) may be written in component form as

$$\frac{d}{dt} x_k(t) = \sum_{i=o}^{N-1} A_i \, x_{(k-i) \bmod N}(t) + B_i \, u_{(k-i) \bmod N}(t) \qquad (2.5)$$

$$y_k(t) = \sum_{i=o}^{N-1} C_i \, y_{(k-i) \bmod N}(t) \qquad (2.6)$$

where the notation (k-i)mod N is used to denote the unique integer j in the set {0,1,...,N-1} such that (k-i)+j is divisible by N. Figure 2.1 illustrates the unforced dynamics of a circulant system. All of the subsystems are identical in the sense that each substate $x_k(t)$ has

- the same self-dynamics A_0
- for any i, the same interaction with substate $x_{(k+i) \bmod N}(t)$.

Also, for any i, the subinput $u_{(k+i) \bmod N}(t)$ affects each $x_k(t)$ in the same way, and each $x_k(t)$ contributes equally to the suboutput $y_{(k+i) \bmod N}(t)$. Thus circulant systems may be called spatially symmetric where spatial refers to the subsystem index number. One way to view this is that someone located at subsystem k looking out at the rest of the subsystem, observes behavior that is independent of k.

2.1.2 A Useful Spatial Transformation

As discussed in Appendix A, a circulant matrix of dimension N has

83702AW073

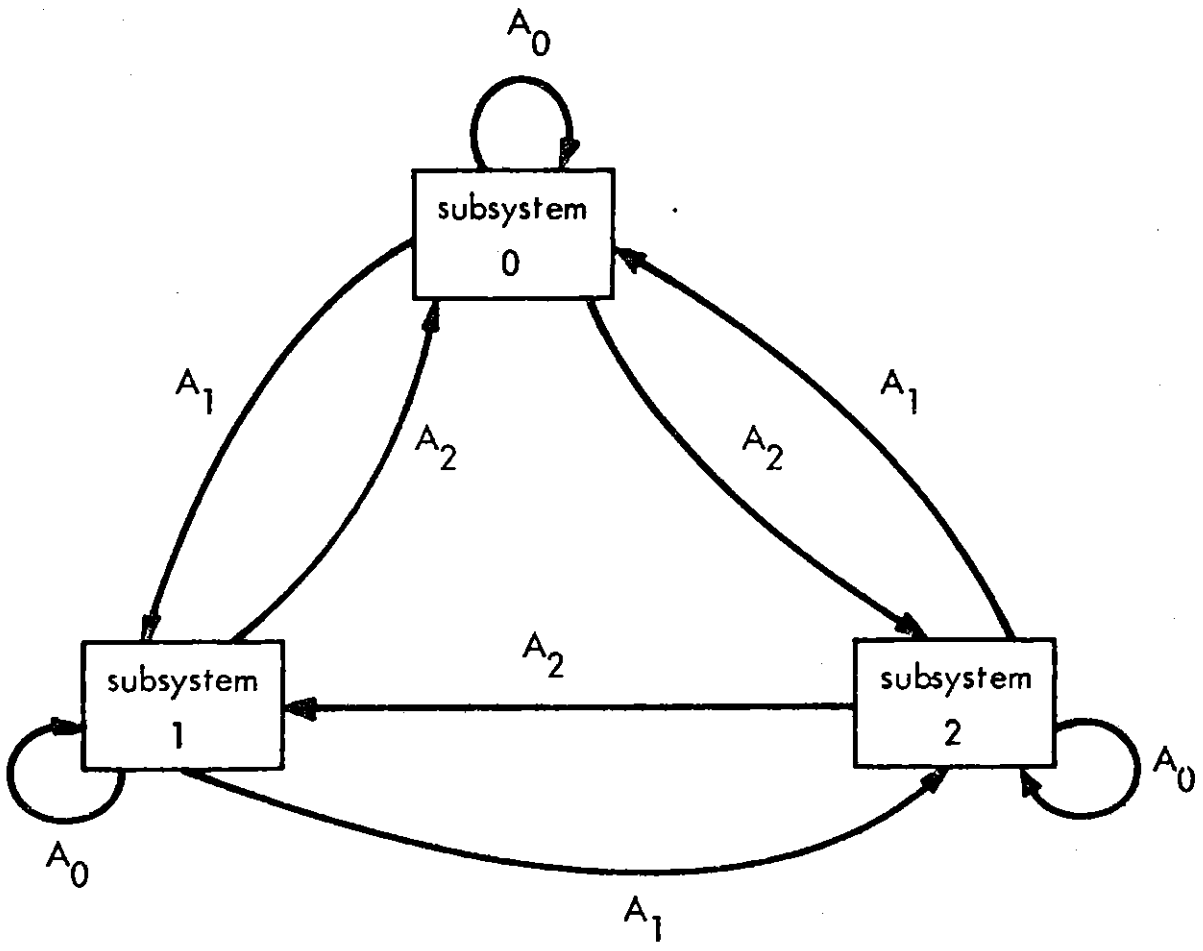


Figure 2.1 The unforced dynamics of a circulant system with three subsystems .

fixed eigenvectors given by

$$\phi_k = \begin{pmatrix} 1 \\ W_N^k \\ W_N^{2k} \\ \vdots \\ W_N^{(N-1)k} \end{pmatrix} \quad , \; k=0,1,\ldots,N-1 \tag{2.7}$$

where $W_N = \exp\left(\frac{j2\pi}{N}\right)$. Since the N values W_N^k, k=0,1,...,N-1, are all distinct, the eigenvectors ϕ_k form a linearly independent set, and so any circulant matrix can be diagonalized. If Φ is the matrix of eigenvectors,

$$\Phi = \left(\phi_0 \;\vdots\; \phi_1 \;\vdots\; \phi_2 \;\vdots\; \cdots \;\vdots\; \phi_{N-1} \right) \tag{2.8}$$

then $\Phi^{-1} A \Phi$ is diagonal, i.e.

$$\Lambda = \Phi^{-1} A \Phi \tag{2.9}$$

$$= \begin{pmatrix} \lambda_0 & & & & \\ & \lambda_1 & & & \\ & & \lambda_2 & & \\ & & & \ddots & \\ & & & & \lambda_{N-1} \end{pmatrix}$$

It is easily shown [44] that the eigenvalues λ_k of A may be computed from

$$\lambda_k = \sum_{i=o}^{N-1} a_i W_N^{-ki} \tag{2.10}$$

This equation means that the finite sequences $(\lambda_0,\lambda_1,\ldots,\lambda_{N-1})$ and $(a_0,a_1,\ldots,a_{N-1})$ are related by the discrete Fourier transform (DFT). Thus the eigenvalues of a circulant matrix can be obtained by applying the fast Fourier transform (FFT) algorithm to the top row of the matrix.

Consider now a block circulant matrix A where the blocks A_k have dimensions r×s. In analogy with the circulant case, the partitioned matrix Φ_r is defined as

$$\Phi_r = \begin{pmatrix} I_r & I_r & I_r & \cdots & I_r \\ I_r & W_N I_r & W_N^2 I_r & \cdots & W_N^{N-1} I_r \\ I_r & W_N^2 I_r & W_N^4 I_r & \cdots & W_N^{2(N-1)} I_r \\ \vdots & \vdots & \vdots & & \vdots \\ I_r & W_N^{N-1} I_r & W_N^{2(N-1)} I_r & \cdots & W_N^{(N-1)(N-1)} I_r \end{pmatrix} \tag{2.11}$$

where I_r is the r×r identity matrix. In Appendix A it is shown that

$$\bar{A} = \Phi_r^{-1} A \Phi_s \tag{2.12}$$

$$= \begin{pmatrix} \bar{A}_0 & & & \\ & \bar{A}_1 & & \\ & & \bar{A}_2 & \\ & & & \ddots \\ & & & & \bar{A}_{N-1} \end{pmatrix}$$

where the blocks $\bar{A}_k$ on the diagonal are r×s matrices satisfying

$$\bar{A}_k = \sum_{i=o}^{N-1} A_i \, W_N^{-ki} \tag{2.13}$$

That is, in the block circulant case, the elements of the block diagonal form are the DFT of the top block row of the block circulant matrix.

A very useful change of basis can now be defined for circulant systems. Let

$$\bar{x}(t) = \Phi_n^{-1} \, x(t) \tag{2.14}$$

$$\bar{u}(t) = \Phi_m^{-1} \, u(t) \tag{2.15}$$

$$\bar{y}(t) = \Phi_p^{-1} \, y(t) \tag{2.16}$$

It is to be noted that $\Phi_n \bar{x}(t) = x(t)$ or, in component form,

$$x_k(t) = \sum_{i=o}^{N-1} \bar{x}_i(t) \, W_N^{ki} \tag{2.17}$$

i.e. the substates $\{x_k(t)\}$ are the inverse DFT of the $\{\bar{x}_k(t)\}$. Thus, the substates $\{\bar{x}_k(t)\}$, subinputs $\{\bar{u}_k(t)\}$, and suboutputs $\{\bar{y}_k(t)\}$ are simply the respective DFT's of $\{x_k(t)\}$, $\{u_k(t)\}$, and $\{y_k(t)\}$ where the transform is taken with respect to the index k. The transformed system is described by

$$\begin{aligned} \frac{d}{dt} \bar{x}(t) &= \left(\Phi_n^{-1} \, A \, \Phi_n\right)\bar{x}(t) + \left(\Phi_n^{-1} B \, \Phi_m\right) \bar{u}(t) \\ &= \bar{A} \, \bar{x}(t) + \bar{B} \, \bar{u}(t) \end{aligned} \tag{2.18}$$

$$\bar{y}(t) = \left(\Phi_p^{-1} C \Phi_n\right) \bar{x}(t)$$

$$= \bar{C}\,\bar{x}(t) \qquad (2.19)$$

where $\bar{A}$, $\bar{B}$, $\bar{C}$ are block diagonal matrices whose elements are given by (2.13).

Under this change of basis, the system is composed of N completely independent complex-valued subsystems

$$\frac{d}{dt}\bar{x}_k(t) = \bar{A}_k\,\bar{x}_k(k) + \bar{B}_k\,\bar{u}_k(t) \qquad (2.20)$$

$$\bar{y}_k(t) = \bar{C}_k\,\bar{x}_k(t) \quad , \; k=0,1,\ldots,N-1 \qquad (2.21)$$

This independence is in the "frequency domain", however, and is not directly applicable to decentralized control or estimation problems since each transformed subsystem $\bar{x}_k(t)$ depends on all of the subsystems $x_i(t)$, $i=0,1,\ldots,N-1$. Computation of the transition matrix, on the other hand, has been decomposed into the computation of the N lower dimensional transition matrices of the $\bar{A}_k$. Since the change of basis in (2.14)-(2.16) and the determination of $\bar{A}_k$, $\bar{B}_k$, and $\bar{C}_k$ are easily done using the FFT, the dynamic behavior of a circulant system is much easier to determine than that of a general nN-dimensional system.

In order to gain further insight as to why the transformed dynamics are uncoupled, consider the dynamics of the k^{th} substate,

$$\frac{d}{dt}x_k(t) = \sum_{i=o}^{N-1} A_i\, x_{(k-i) \bmod N}(t) + B_i\, u_{(k-i) \bmod N}(t) \qquad (2.5)$$

Notice that each of the two terms on the RHS of (2.5) is a circular convolution sum. Because of the relationship between convolution and Fourier transformation, one expects (2.5) to be decoupled in the frequency domain. This result is obtained by multiplying (2.19) by $W_N^{-k\ell}$ and summing over k,

$$\sum_{k=o}^{N-1} \frac{d}{dt} x_k(t) W_N^{-k\ell} = \sum_{k=o}^{N-1} \sum_{i=o}^{N-1} \left\{ A_i x_{(k-i) \bmod N}(t) W_N^{-k\ell} + B_i u_{(k-i) \bmod N}(t) W_N^{-k\ell} \right\}$$

$$= \sum_{i=o}^{N-1} \left\{ A_i W_N^{-i\ell} \right\} \sum_{k=o}^{N-1} \left\{ x_{(k-i) \bmod N}(t) W_N^{(i-k)\ell} \right\} +$$

$$+ \sum_{i=o}^{N-1} \left\{ B_i W_N^{-i\ell} \right\} \sum_{k=o}^{N-1} \left\{ u_{(k-i) \bmod N}(t) W_N^{(i-k)\ell} \right\} \tag{2.22}$$

But (2.22) is just an expanded version of

$$\frac{d}{dt} \bar{x}_\ell(t) = \bar{A}_\ell \bar{x}_\ell(t) + \bar{B}_\ell \bar{u}_\ell(t) \tag{2.18}$$

The key element here is the fact that a circulant matrix times a vector equals the circular convolution of the top row of the matrix with the vector. The Fourier transform is then applied to convert convolution into "multiplication in the spatial frequency domain".

Before concluding this discussion of the spatial transformation, several identities will be presented here for easy reference throughout the chapter. Consider first a real block circulant matrix A and the relationship between $\bar{A}_k$ and $\bar{A}_{N-k}$. From (2.13),

$$\mathrm{Re}\left[\bar{A}_k\right] = \mathrm{Re}\left[\sum_{i=0}^{N-1} A_i W_N^{-ki}\right]$$

$$= \sum_{i=0}^{N-1} A_i \,\mathrm{Re}\left[W_N^{-ki}\right]$$

$$= \sum_{i=0}^{N-1} A_i \,\mathrm{Re}\left[W_N^{(N-k)i}\right]$$

$$= \mathrm{Re}\left[\bar{A}_{N-k}\right] \qquad (2.23)$$

Similarly,

$$\mathrm{Im}\left[\bar{A}_k\right] = \mathrm{Im}\left[\sum_{i=0}^{N-1} A_i W_N^{ki}\right]$$

$$= -\mathrm{Im}\left[\sum_{i=0}^{N-1} A_i W_N^{(N-k)i}\right]$$

$$= -\mathrm{Im}\left[\bar{A}_{N-k}\right] \qquad (2.24)$$

Equations (2.23) and (2.24) just state the well-known property of the DFT that real-valued sequences have transforms with real parts that are even and imaginary parts that are odd. The sequence $\{\mathrm{Re}[\bar{A}_k]\}$ is called even because $\mathrm{Re}[A_k] = \mathrm{Re}[A_{-k}]$ where the indices are modulo N. The sequence $\{\mathrm{Im}[\bar{A}_k]\}$ is odd since $\mathrm{Im}[\bar{A}_k] = -\mathrm{Im}[\bar{A}_{-k}]$. Next, identities for a symmetric block circulant matrix Q will be obtained. The symmetry of Q is equivalent to the following condition on the blocks Q_k:

$$Q_k = Q'_{(-k) \bmod N} \qquad (2.25)$$

Therefore,

$$\mathrm{Re}\left[\bar{Q}_k\right] = \sum_{i=o}^{N-1} Q_i \ \mathrm{Re}\left[W_N^{-ki}\right]$$

$$= \sum_{i=o}^{N-1} Q'_{(-i)\bmod N} \ \mathrm{Re}\left[W_N^{ki}\right]$$

$$= \sum_{\ell=o}^{N-1} Q'_\ell \ \mathrm{Re}\left[W_N^{-k\ell}\right] \quad , \ \ell = (-i)\bmod N$$

$$= \mathrm{Re}\left[\bar{Q}'_k\right] \tag{2.26}$$

Similarly,

$$\mathrm{Im}\left[\bar{Q}_k\right] = \sum_{i=o}^{N-1} Q'_{(-i)\bmod N} \ \mathrm{Im}\left[W_N^{-ki}\right]$$

$$= -\sum_{\ell=o}^{N-1} Q'_\ell \ \mathrm{Im}\left[W_N^{-k\ell}\right]$$

$$= -\mathrm{Im}\left[\bar{Q}'_k\right] \tag{2.27}$$

Combining (2.23) and (2.26) for the real part of $\bar{Q}_k$ and (2.24) and (2.27) for the imaginary part of $\bar{Q}_k$ yields

$$\mathrm{Re}\left[\bar{Q}_k\right] = \mathrm{Re}\left[\bar{Q}'_{N-k}\right] \tag{2.28}$$

$$\mathrm{Im}\left[\bar{Q}_k\right] = \mathrm{Im}\left[\bar{Q}'_{N-k}\right] \tag{2.29}$$

The final identity relates the transforms of A and A'. If A is a block circulant matrix and B=A', then B is also block circulant. Partitioning

B according to (2.1), block B_k equals A'_{N-k}. Thus

$$\begin{aligned} \mathrm{Re}\left[\bar{B}_k\right] &= \sum_{i=o}^{N-1} B_i \ \mathrm{Re}\left[W_N^{-ki}\right] \\ &= \sum_{i=o}^{N-1} A'_{N-k} \ \mathrm{Re}\left[W_N^{ki}\right] \\ &= \sum_{\ell=o}^{N-1} A'_{\ell} \ \mathrm{Re}\left[W_N^{-k\ell}\right] \\ &= \mathrm{Re}\left[\bar{A}'_k\right] \end{aligned} \tag{2.30}$$

Also,

$$\begin{aligned} \mathrm{Im}\left[\bar{B}_k\right] &= \sum_{i=o}^{N-1} A'_{N-k} \ \mathrm{Im}\left[W_N^{-ki}\right] \\ &= -\sum_{\ell=o}^{N-1} A'_{\ell} \ \mathrm{Im}\left[W_N^{-k\ell}\right] \\ &= -\mathrm{Im}\left[\bar{A}'_k\right] \end{aligned} \tag{2.31}$$

Combining (2.30) and (2.31) yields

$$\overline{(A')} = \bar{A}^* \tag{2.32}$$

where * denotes Hermitian, i.e. the transposed complex conjugate. These identities will be heavily used in Sections 2.4 and 2.5.

2.2 Imbedding Tridiagonal Systems in Circulant Systems

The usefulness of circulant models is increased by showing how symmetric and antisymmetric tridiagonal systems can be imbedded in circulant

systems. Symmetric tridiagonal systems are linear systems for which the system matrix is block tridiagonal and the blocks on the subdiagonal and superdiagonal are equal. For antisymmetric tridiagonal systems, the blocks on the subdiagonal are the negative of the blocks on the superdiagonal. Brockett and Willems [45] have shown how a symmetric tridiagonal system could be imbedded in a circulant system roughly twice as large. The imbedding of an antisymmetric tridiagonal system is new, but strongly motivated by Brockett and Willems.

Before presenting the imbedding methods, it is interesting to consider why one would desire an imbedding of this type. One reason might be the computational advantages associated with determining the eigenvalues or solving a Riccati equation (see Section 2.5) for a circulant system. The work of Jain and Angel [26], however, suggests that similar savings can arise from a direct analysis of these tridiagonal systems. It is in the area of decentralized control and estimation that the motivation for this imbedding is found. Consider a decentralized control structure where feedback is allowed not just from the nearest neighbors but from the first and second nearest neighbor on each side. The resulting closed loop system matrix is no longer tridiagonal -- it is still circulant if the system has been imbedded in a circulant system. This is because, as is shown in Appendix A, the product or sum of circulant matrices is still circulant. Therefore, since feedback can destroy the tridiagonal property but not the circulant property, it is useful to imbed a tridiagonal system within a circulant system.

2.2.1 Symmetric Tridiagonal Systems

The unforced symmetric tridiagonal system under consideration is given by

$$\frac{d}{dt} z(t) = F\, z(t)$$

$$= \begin{pmatrix} F_0 & F_1 & & & \\ F_1 & F_0 & F_1 & & \\ & F_1 & \ddots & \ddots & \\ & & \ddots & & F_1 \\ & & & F_1 & F_0 \end{pmatrix} \begin{pmatrix} z_1(t) \\ z_2(t) \\ \vdots \\ z_N(t) \end{pmatrix} \tag{2.33}$$

where $z(t) \in \mathbb{R}^{nN}$. This system is close to being circulant; all that is needed is for the upper right hand corner block and the lower left hand corner block of F to be F_1 instead of zero. What this means physically is that the symmetric tridiagonal system fails to be circulant because the two end subsystems do not directly interact with each other.

The system (2.33) is imbedded in the following circulant system of dimension $2n(N+1)$:

$$\frac{d}{dt} x(t) = A\, x(t)$$

$$= \begin{pmatrix} F_0 & F_1 & & & F_1 \\ F_1 & F_0 & F_1 & & \\ & F_1 & \ddots & \ddots & \\ & & \ddots & & F_1 \\ F_1 & & & F_1 & F_0 \end{pmatrix} \begin{pmatrix} x_0(t) \\ x_1(t) \\ \vdots \\ x_{2N+1}(t) \end{pmatrix} \tag{2.34}$$

The idea here is for $x_1(t),\ldots,x_N(t)$ to equal $z_1(t),\ldots,z_N(t)$, respectively, for all time t. In order for $x_1(t)$ to equal $z_1(t)$, it is necessary that

$$\frac{d}{dt}x_1(t) = \frac{d}{dt}z_1(t) \tag{2.35}$$

$$F_1x_0(t) + F_0x_1(t) + F_1x_2(t) = F_0z_1(t) + F_1z_2(t)$$

Equation (2.35) implies that $x_0(t)$ must be identically zero. Demanding that $x_N(t)$ equal $z_N(t)$ implies, by the same argument, that $x_{N+1}(t)$ must also be identically zero. Now the question becomes how can $x_0(t)$ and $x_{N+1}(t)$ be made to remain at zero? For $x_{N+1}(t)$ to be identically zero, it derivative must also be zero for all t,

$$\frac{d}{dt}x_{N+1}(t) = F_1\,x_N(t) + F_0x_{N+1}(t) + F_1\,x_{N+2}(t) \tag{2.36}$$

$$= F_1\left[x_N(t) + x_{N+2}(t)\right] \quad (\text{since } x_{N+1}(t)=0)$$

$$= 0$$

Clearly a sufficient condition is $x_{N+2}(t) = -x_N(t) = -z_N(t)$ for all t. That is, if x_N and x_{N+2} exert equal but opposite "forces" on x_{N+1}, then x_{N+1} will remain at zero. Similarly, for x_0 to remain at zero it suffices to have $x_{2N+1}(t) = -x_1(t) = -z_1(t)$ for all t. Continuing this line of reasoning for x_{N+2}, x_{N+3},... and x_{2N+1}, x_{2N},... yields the conclusion that if the initial state z(o) is extended in an odd way, i.e.

$$x_0(o) = 0 \tag{2.37}$$

$$x_k(o) = z_k(o)$$

$$x_{N+1}(o) = 0$$

$$x_{2N+2-k}(o) = -z_k(o) \quad , \; k=1,2,\ldots,N$$

then for all time t, the substates $x_0(t)$ and $x_{N+1}(t)$ remain fixed at zero and $z_k(t) = x_k(t) = -x_{2N+2-k}(t)$, $k=1,2,\ldots,N$. This extension is illustrated in Figure 2.2 for an example with scalar subsystems. In this manner, a symmetric tridiagonal system can be imbedded in a circulant system.

This imbedding procedure is analogous to a method used to determine the transverse displacement of a finite string [46]. The displacement of a string having fixed ends at O and L is given by the one-dimensional wave equation. In the case of an infinite string, the solution of the wave equation is the well known d'Alembert solution. The displacement of the finite string can be obtained from the infinite string analysis by means of the following procedure. The initial displacement of the finite string is extended to an odd periodic function of period 2L. The initial velocity is also extended in the same way. The resulting displacement between O and L of an infinite string having these extended initial displacement and velocity is identical to the displacement of the finite string. Thus the behavior of a finite string can be obtained from d'Alembert's solution to the infinite string problem. This odd

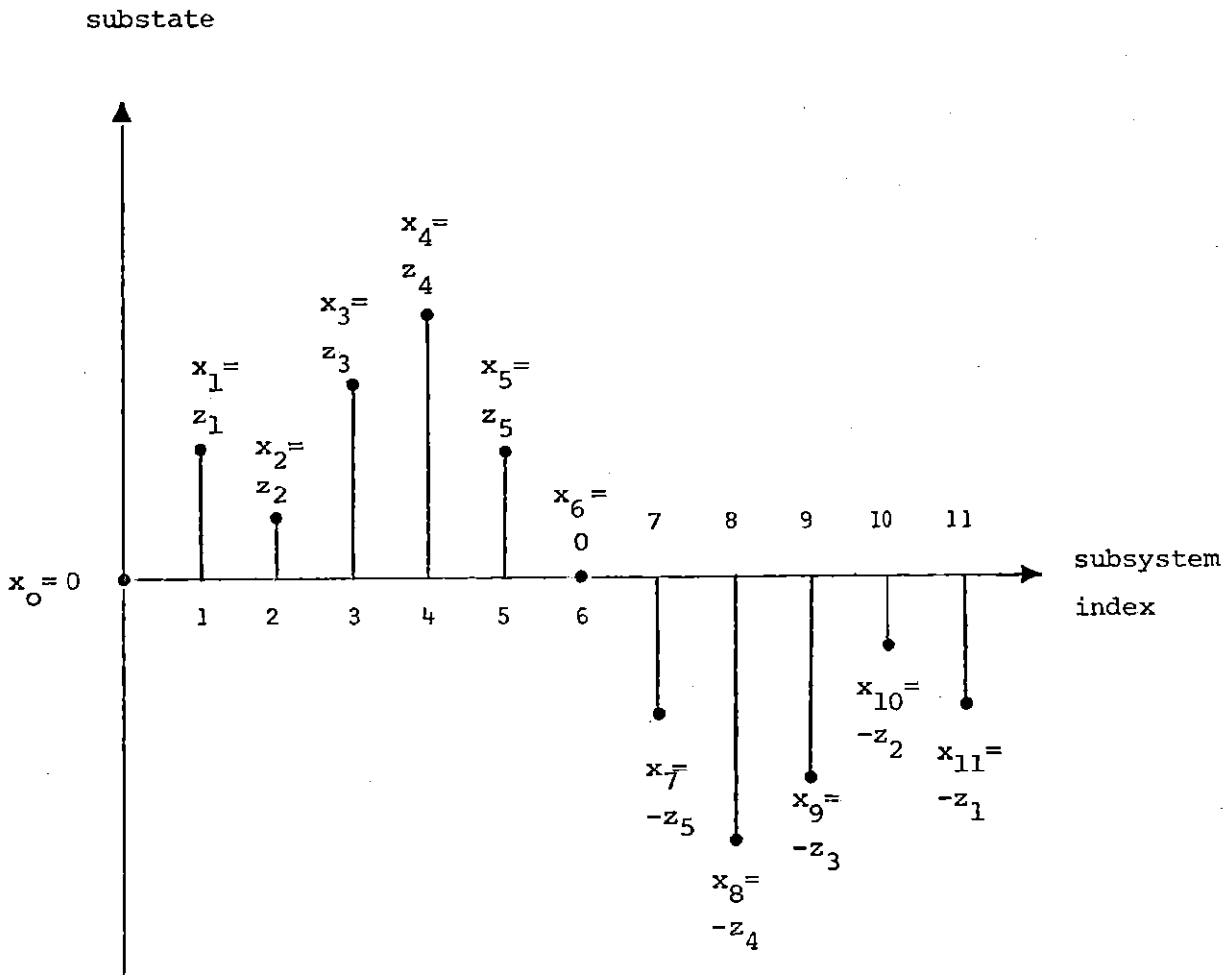


Figure 2.2 When a symmetric tridiagonal system is imbedded in a circulant system, the initial state must be extended in an odd manner.

periodic extension of the initial displacement and velocity is essentially the same procedure as the imbedding formula (2.37).

The idea of this imbedding procedure is also somewhat similar to the idea behind the method of images used in electrostatics. The method of images can be used, for example, to solve the problem of a point charge q placed near a perfect conducting plane of infinite extent. The method replaces the conducting plane with a point charge -q located at the mirror image of charge q. The potential of these two charges is zero where the conducting plane was located. In the case of imbedding a symmetric tridiagonal system in a circulant system, the substates $x_o(t)$ and $x_{N+1}(t)$ are identically zero. The mirror image of the state z(t) (see (2.37)) is used to ensure that these two subsystems remain at zero.

The case of the tridiagonal system with scalar subsystems, i.e. n=1, will be studied further. First, a nonsymmetric tridiagonal matrix

$$F = \begin{pmatrix} f_0 & f_{N-1} & & & & \\ f_1 & f_0 & f_{N-1} & & & \\ & f_1 & f_0 & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & & f_{N-1} & \\ & & & f_1 & f_0 & \end{pmatrix} \qquad (2.38)$$

can be transformed into a symmetric tridiagonal matrix by a similarity transformation if f_1 and f_{N-1} are both nonzero.

If

$$P = \begin{pmatrix} 1 & & & & \\ & (f_{N-1}/f_1)^{\frac{1}{2}} & & & \\ & & (f_{N-1}/f_1)^{\frac{2}{2}} & & \\ & & & \ddots & \\ & & & & (f_{N-1}/f_1)^{\frac{N-1}{2}} \end{pmatrix} \tag{2.39}$$

then PFP^{-1} is symmetric,

$$PFP^{-1} = \begin{pmatrix} f_0 & \sqrt{f_1 f_{N-1}} & & & \\ \sqrt{f_1 f_{N-1}} & f_0 & \sqrt{f_1 f_{N-1}} & & \\ & \sqrt{f_1 f_{N-1}} & f_0 & \ddots & \\ & & \ddots & \ddots & \sqrt{f_1 f_{N-1}} \\ & & & \sqrt{f_1 f_{N-2}} & f_0 \end{pmatrix} \tag{2.40}$$

Hence the assumption that $F = F'$ can be made without loss of generality.

Next the eigenvalues and eigenvectors of the symmetric tridiagonal matrix F will be found from the eigenvalues and eigenvectors of a circulant. Let A be the $2(N+1) \times 2(N+1)$ circulant matrix obtained from F according to (2.34). The linear transformation $S : \mathbb{R}^N \to \mathbb{R}^{2(N+1)}$ is defined componentwise as follows:

$$(Sw)_k = \begin{cases} 0 & , \quad k=0 \text{ or } N+1 \\ w_k & , \quad k=1,2,\ldots,N \\ -w_{2N+2-k} & , \quad k=N+2,\ldots,2N+1 \end{cases} \tag{2.41}$$

where $w \in \mathbb{R}^N$. Clearly S is just the linear transformation used to imbed a tridiagonal system within a circulant system, i.e. $x(o) = S\, z(o)$. For any vector $w \in \mathbb{R}^N$, it is trivially verified that

$$S\, F\, w = A\, S\, w \tag{2.42}$$

The equation SF = AS is the aggregation equation that arises in large-scale system analysis when an aggregated model is being constructed. In the context of aggregation, the state z is the state of a large-scale system, and F is the corresponding system matrix. The transformation S maps z into an aggregated state x = Sz where the system matrix for x is A and must satisfy the aggregation equation. It is well-known that the eigenvalues of the aggregated system matrix are a subset of the eigenvalues of the large-scale system matrix. In the context of imbedding a tridiagonal system within a circulant system, the equation SF=FA can be thought of a a "disaggregation" equation since the resulting state is of higher order than the state of the original model.

If w is an eigenvector of F corresponding to eigenvalue λ, then

$$A(S\, w) = S\, F\, w = \lambda(s\, w) \tag{2.43}$$

Hence λ is also an eigenvalue of A with eignevector S w. What are the eigenvalues of A? From (2.10), the k^{th} eigenvalue is

$$\lambda_k = f_0 + f_1 W_{2N+2}^{-k} + f_1 W_{2N+2}^{+k}$$

$$= f_0 + 2f_1 \cos\left(\frac{k\Pi}{N+1}\right) \quad , k=0,1,\ldots,2N+1 \tag{2.44}$$

Thus there are only N+2 distinct eigenvalues -- eigenvalues $\lambda_0 = f_0 + 2f_1$ and $\lambda_{N+1} = f_0 - 2f_1$ are not repeated; the other N eigenvalues occur twice. Can λ_0 also be an eigenvalue of F? Since the corresponding eigenvector is

$$\phi_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{2.45}$$

and ϕ_0 is obviously not in the range space of S, λ_0 cannot be an eigenvalue of F. Similarly, eigenvalue λ_{N+1} of A has eigenvector

$$\phi_{N+1} = \begin{pmatrix} +1 \\ -1 \\ \vdots \\ +1 \\ -1 \end{pmatrix} \tag{2.46}$$

and, since ϕ_{N+1} cannot be in the range space of S, λ_{N+1} is not an eigenvalue of F. The only remaining candidates for eigenvalues of F are the

repeated eigenvalues of A. Consider the pair of eigenvalues λ_k and λ_{2N+2-k}. Any eigenvector χ_k of A corresponding to this eigenvalue can be written as a linear combination of the eigenvectors ϕ_k and ϕ_{2N+2-k},

$$\chi_k = \alpha \begin{pmatrix} 1 \\ \cos\left(\frac{k\pi}{N+1}\right) - j \sin\left(\frac{k\pi}{N+1}\right) \\ \cos\left(\frac{k2\pi}{N+1}\right) - j \sin\left(\frac{k2\pi}{N+1}\right) \\ \vdots \\ \cos\left(\frac{k(N+1)\pi}{N+1}\right) - j \sin\left(\frac{k(N+1)\pi}{N+1}\right) \\ \vdots \\ \cos\left(\frac{k(2N+1)\pi}{N+1}\right) - j \sin\left(\frac{k(2N+1)\pi}{N+1}\right) \end{pmatrix} + \beta \begin{pmatrix} 1 \\ \cos\left(\frac{k\pi}{N+1}\right) + j \sin\left(\frac{k\pi}{N+1}\right) \\ \cos\left(\frac{k2\pi}{N+1}\right) + j \sin\left(\frac{k2\pi}{N+1}\right) \\ \vdots \\ \cos\left(\frac{k(N+1)\pi}{N+1}\right) + j \sin\left(\frac{k(N+1)\pi}{N+1}\right) \\ \vdots \\ \cos\left(\frac{k(2N+1)\pi}{N+1}\right) + j \sin\left(\frac{k(2N+1)\pi}{N+1}\right) \end{pmatrix} \tag{2.47}$$

$$= \begin{pmatrix} (\alpha+\beta) \\ (\alpha+\beta)\cos\left(\frac{k\pi}{N+1}\right) + j\,(\beta-\alpha)\sin\left(\frac{k\pi}{N+1}\right) \\ (\alpha+\beta)\cos\left(\frac{2k\pi}{N+1}\right) + j\,(\beta-\alpha)\sin\left(\frac{2k\pi}{N+1}\right) \\ \vdots \\ (\alpha+\beta)\cos\left(k\pi\right) + j\,(\beta-\alpha)\sin\left(k\pi\right) \\ \vdots \\ (\alpha+\beta)\cos\left(\frac{k(2N+1)\pi}{N+1}\right) + j\,(\beta-\alpha)\sin\left(\frac{k(2N+1)\pi}{N+1}\right) \end{pmatrix}$$

For χ_k to be in the range of S requires, from (2.41), that the 0^{th} and the $(N+1)^{th}$ components of χ_k be zero, i.e. $(\alpha+\beta) = 0$ and $(\alpha+\beta)\cos(k\pi) = 0$. Therefore, choosing $\alpha = -\beta$ yields an eigenvector ψ_k of F, i.e.

$$\chi_k = S\psi_k \tag{2.48}$$

A purely real eigenvector is obtained by letting $\alpha = -\beta = j/2$,

$$\psi_k = \begin{pmatrix} \sin\left(\frac{k\pi}{N+1}\right) \\ \sin\left(\frac{2k\pi}{N+1}\right) \\ \vdots \\ \sin\left(\frac{2kN\pi}{N+1}\right) \end{pmatrix} \tag{2.49}$$

components 1 through N of χ_k. The corresponding eigenvalue of ψ_k is $\lambda_k = f_o + 2f_1 \cos\left(\frac{k\pi}{N+1}\right)$.

From the above analysis, it is clear that the eigenvalues of a symmetric tridiagonal (finite) Toeplitz matrix can be found using the FFT algorithm. This is proved more directly in [26] by Jain and Angel. Using the DFT to efficiently diagonalize symmetric tridiagonal matrices, Jain and Angel are able to decompose the vector filtering equations arising in the restoration of images into a set of uncoupled scalar filtering equations. This is completely analogous to the decomposition of Lyapunov and Riccati equations in Section 2.5. In fact, their results on efficient algorithms for image processing were one of the motivating factors for undertaking the work presented here.

Finally, it is noted that the idea of imbedding a symmetric tridiagonal system within a circulant system can be extended to higher dimensions. Consider a tridiagonal system

$$\frac{d}{dt}\begin{pmatrix} z_1(t) \\ z_2(t) \\ \vdots \\ z_N(t) \end{pmatrix} = \begin{pmatrix} F_0 & F_1 & & \bigcirc \\ F_1 & F_0 & \ddots & \\ & \ddots & \ddots & F_1 \\ \bigcirc & & F_1 & F_0 \end{pmatrix} \begin{pmatrix} z_1(t) \\ z_2(t) \\ \vdots \\ z_N(t) \end{pmatrix} \tag{2.50}$$

where each substate $z_k(t)$ is itself composed of q sub-substates,

$$z_k(t) = \begin{pmatrix} z_{k1}(t) \\ z_{k2}(t) \\ \vdots \\ z_{kq}(t) \end{pmatrix} \quad , k=1,2,\ldots,N \tag{2.51}$$

and the matrices F_0 and F_1 have the special forms

$$F_0 = \begin{pmatrix} F_{00} & F_{01} & & \bigcirc \\ F_{01} & F_{00} & \ddots & \\ & \ddots & \ddots & F_{01} \\ \bigcirc & & F_{01} & F_{00} \end{pmatrix} \tag{2.52}$$

$$F_1 = \begin{pmatrix} F_{10} & & & \\ & F_{10} & & \\ & & \ddots & \\ & & & F_{10} \end{pmatrix} \qquad (2.53)$$

Since the individual subsystems are symmetric tridiagonal systems, the imbedding procedure can be applied at the subsystem level to yield

$$\frac{d}{dt}\begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{pmatrix} = \begin{pmatrix} A_0 & A_1 & & \\ A_1 & A_0 & \ddots & \\ & \ddots & \ddots & A_1 \\ & & A_1 & A_0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{pmatrix} \qquad (2.54)$$

where the substates $x_k(t)$ are composed of (2q+2) sub-substates, $x_k(t) = Sz_k(t)$ and

$$A_0 = \begin{pmatrix} F_{00} & F_{01} & & & F_{01} \\ F_{01} & F_{00} & F_{01} & & \\ & F_{01} & F_{00} & F_{01} & \\ & & \ddots & \ddots & F_{01} \\ F_{01} & & & F_{01} & F_{00} \end{pmatrix} \qquad (2.55)$$

$$A_1 = \begin{pmatrix} F_{10} & & & \\ & F_{10} & & \\ & & \ddots & \\ & & & F_{10} \end{pmatrix} \qquad (2.56)$$

For this imbedding to be valid, it is necessary that $x_{k,0}(t)$ and $x_{k,q+1}(t)$ be identically zero. Writing out (2.54) for $x_{k,0}(t)$,

$$\frac{d}{dt} x_{k,0}(t) = F_{10}x_{k-1,0}(t) + F_{01}x_{k,2q+2}(t) + F_{00}x_{k,0}(t) + F_{01}x_{k,1}(t) + F_{10}x_{k+1,0} \qquad (2.57)$$

A consistent solution is obtained if $x_{k-1,0}(t)$, $x_{k,0}(t)$, $x_{k+1,0}(t)$ are all zero and $x_{k,1}(t) = -x_{k,2q+2}(t)$. Similarly, one sees that all the $x_{k,q+1}(t)$ are identically zero, and so this imbedding is valid. The resulting system (2.54), moreover, is itself a symmetric tridiagonal system. Applying the imbedding procedure to it yields

$$\frac{d}{dt} \begin{pmatrix} x_0(t) \\ x_1(t) \\ \vdots \\ x_{2N+2}(t) \end{pmatrix} = \begin{pmatrix} A_0 & A_1 & & A_1 \\ A_1 & A_0 & A_1 & \\ & \ddots & \ddots & A_1 \\ A_1 & & A_1 & A_0 \end{pmatrix} \begin{pmatrix} x_0(t) \\ x_1(t) \\ \vdots \\ x_{2N+2}(t) \end{pmatrix} \qquad (2.58)$$

a circulant system where the blocks of the system matrix are themselves block circulant matrices. The original system of order nqN is imbedded

in a circulant system of order $4n(q+1)(N+1)$. In Section 2.3, an example of such a symmetric tridiagonal system will be described.

2.2.2 Antisymmetric Tridiagonal Systems

An unforced antisymmetric tridiagonal system has the form

$$\frac{d}{dt} z(t) = F\, z(t)$$

$$= \begin{pmatrix} F_0 & -F_1 & & & \\ F_1 & F_0 & -F_1 & & \\ & F_1 & \ddots & \ddots & \\ & & \ddots & & -F_1 \\ & & & F_1 & F_0 \end{pmatrix} \begin{pmatrix} z_1(t) \\ z_2(t) \\ \vdots \\ z_N(t) \end{pmatrix} \qquad (2.59)$$

where $z(t) \in \mathbb{R}^{nN}$. It will be shown that there are two different imbedding procedures for antisymmetric tridiagonal systems depending on whether N, the number of subsystems, is even or odd. For odd N, the proper circulant system has order $2n(N+1)$, just as in the symmetric tridiagonal case. For even N, however, a circulant system of order $4n(N+1)$ is necessary.

The situation when N is odd will be considered first. The candidate circulant system is

$$\frac{d}{dt}\, x(t) = A\, x(t)$$

$$= \begin{pmatrix} F_0 & -F_1 & & & F_1 \\ F_1 & F_0 & -F_1 & & \\ & F_1 & \ddots & \ddots & \\ & & \ddots & & -F_1 \\ -F_1 & & & F_1 & F_0 \end{pmatrix} \begin{pmatrix} x_0(t) \\ x_1(t) \\ \vdots \\ x_{2N+1}(t) \end{pmatrix} \qquad (2.60)$$

This circulant system is very similar to the circulant system (2.34) used for imbedding symmetric tridiagonal systems. Just as in the symmetric case, the substates $x_1(t),\ldots,x_N(t)$ equal $z_1(t),\ldots,z_N(t)$, respectively, for all time. This immediately implies that $x_0(t)$ and $x_{N+1}(t)$ must be identically zero. Since $x_{N+1}(t)$ is zero,

$$\frac{d}{dt}\, x_{N+1}(t) = F_1\, x_N(t) + F_0\, x_{N+1}(t) - F_1\, x_{N+2}(t) \qquad (2.61)$$

$$= F_1\left[x_N(t) - x_{N+2}(t)\right] \quad (\text{since } x_{N+1}(t) = 0)$$

$$= 0$$

A sufficient condition here is that $x_{N+2}(t) = x_N(t) = z_N(t)$, in contrast to the symmetric case. Once again, x_N and x_{N+2} exert equal but opposite forces on x_{N+1}, but in the antisymmetric case this requires $x_{N+2}(t)=x_N(t)$. Examination of the derivative of $x_0(t)$ quickly leads to the conclusion that $x_{2N+1}(t) = x_1(t) = z_1(t)$. In order to clearly illustrate the differ-

ence between the symmetric and antisymmetric cases, the argument will be continued for $x_{N+2}(t)$ and $x_{2N+1}(t)$. The requirement is that $x_{N+2}(t)=x_N(t)$; hence

$$\frac{d}{dt} x_{N+2}(t) = \frac{d}{dt} x_N(t) \tag{2.62}$$

$$F_1 x_{N+1}(t) + F_0 x_{N+2}(t) - F_1 x_{N+3}(t) = F_1 x_{N-1}(t) + F_0 x_N(t) - F_1 x_{N+1}(t)$$

$$F_0 x_{N+2}(t) - F_1 x_{N+3}(t) = F_1 x_{N-1}(t) + F_0 x_N(t)$$

$$\text{(since } x_{N+1}(t) = 0\text{)}$$

Requiring $x_{N+3}(t) = -x_{N-1}(t) = -z_{N-1}(t)$ obviously suffices. The condition $x_{2N}(t) = -x_2(t) = -z_2(t)$ is obtained similarly by considering $x_{2N+1}(t)$. Continuing this argument leads to the conclusion that if

$$\begin{aligned} x_0(o) &= 0 \\ x_k(o) &= z_k(o) \\ x_{N+1}(o) &= 0 \\ x_{2N+2-k}(o) &= (-1)^{k+1} z_k(o) \quad , \quad k=1,2,\ldots,N \end{aligned} \tag{2.63}$$

then $x_1(t),\ldots,x_N(t)$ will equal $z_1(t),\ldots,z_N(t)$, respectively, for all t. Figure 2.3 illustrates this extension in the antisymmetric case for the same system shown in Figure 2.2. It is to be noted that this procedure only works for odd N. From (2.63),

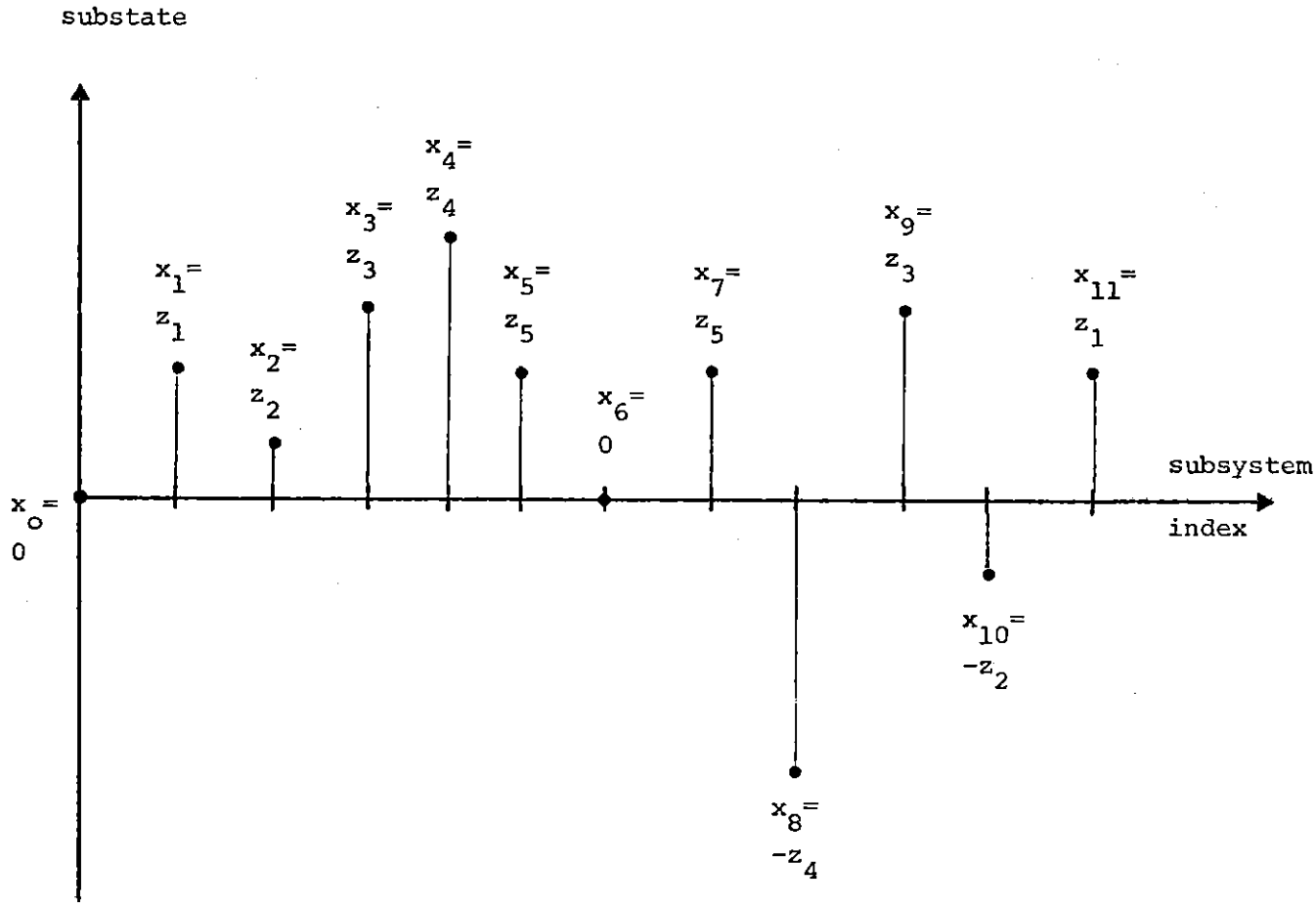


Figure 2.3 Imbedding an antisymmetric tridiagonal system with an odd number of subsystems in a circulant system.

$$x_{N+2}(o) = (-1)^{N+1} z_N(o) \tag{2.64}$$

and so the requirement that $x_{N+2}(t)$ equals $z_N(t)$ is only met for odd N.

Imbedding an antisymmetric tridiagonal system in a circulant system when N is even will now be discussed. Suppose the system matrix of the circulant system in which this antisymmetric system is imbedded has the same form as the system matrix A in (2.60), only now the order of the circulant system is as yet unspecified. It is further supposed that substates $x_1(t),\ldots,x_N(t)$ of the circulant system are to equal $z_1(t),\ldots,z_N(t)$, respectively, for all time. Of course, the immediate implication is that $x_0(t)$ and $x_{N+1}(t)$ are both identically zero. By examining the derivatives of $x_{N+1}(t),\ldots,x_{2N}(t)$, the following relations are obtained:

$$\begin{aligned} x_{N+2}(t) &= z_N(t) \\ x_{N+3}(t) &= -z_{N-1}(t) \\ x_{N+4}(t) &= z_{N-2}(t) \\ &\vdots \qquad\quad \vdots \\ x_{2N}(t) &= z_2(t) \\ x_{2N+1}(t) &= -z_1(t) \end{aligned} \tag{2.65}$$

For odd N, the identity is $x_{2N+1}(t) = z_1(t)$, and so subsystems 0 and (2N+1) could be "tied together" to complete the circulant system. In the case of even N being considered here, tying together subsystems 0 and (2N+1) fails since this would imply

$$\frac{d}{dt} x_0(t) = F_0\, x_0(t) - F_1\, x_1(t) + F_1\, x_{2N+1}(t) \qquad (2.66)$$

$$= -F_1 \left[x_1(t) - x_{2N+1}(t) \right] \quad \text{(since } x_0(t) = 0\text{)}$$

$$= -F_1 \left[z_1(t) + z_1(t) \right] \quad \text{(since } x_1(t) = z_1(t) = -x_{2N+1}(t)\text{)}$$

$$\neq 0$$

a contradiction. Since the subsystems 0 and (2N+1) cannot be joined, additional subsystems are added to the circulant model. Using the device of examining derivatives for $x_{2N+1}(t),\ldots,x_{3N+1}(t)$ yields the following:

$$x_{2N+2}(t) = 0 \qquad (2.67a)$$

$$x_{2N+3}(t) = -z_1(t)$$

$$x_{2N+4}(t) = -z_2(t)$$

$$\vdots \qquad\qquad \vdots$$

$$x_{3N+1}(t) = -z_{N-1}(t)$$

$$x_{3N+2}(t) = -z_N(t)$$

Unfortunately, subsystem (3N+2) cannot be joined to subsystem 0 and have $x_0(t)$ remain equal to zero since $x_{3N+2}(t) = z_N(t)$ and $x_1(t) = z_1(t)$. Therefore, another N+1 subsystems are added to the circulant model,

$$x_{3N+3}(t) = 0 \qquad (2.67b)$$

$$x_{3N+4}(t) = -z_N(t)$$

$$x_{3N+5}(t) = z_{N-1}(t)$$
$$\vdots \qquad\qquad \vdots$$
$$x_{4N+2}(t) = -z_2(t)$$
$$x_{4N+3}(t) = z_1(t)$$

Now subsystem (4N+3) can be joined to subsystem 0 since this yields

$$\frac{d}{dt} x_0(t) = F_0 x_0(t) - F_1 x_1(t) + F_1 x_{4N+3}(t) \tag{2.68}$$
$$= -F_1\left[x_1(t) - x_{4N+3}(t)\right]$$
$$= 0$$

Summarizing, an antisymmetric tridiagonal system with an even number of subsystems can be imbedded in the following circulant system of order 4n(N+1):

$$\frac{d}{dt}\begin{pmatrix} x_0(t) \\ x_1(t) \\ \vdots \\ x_{4N+3}(t) \end{pmatrix} = \begin{pmatrix} F_0 & -F_1 & & & F_1 \\ F_1 & F_0 & -F_1 & & \\ & F_1 & \ddots & \ddots & \\ & & \ddots & & -F_1 \\ -F_1 & & & F_1 & F_0 \end{pmatrix} \begin{pmatrix} x_0(t) \\ x_1(t) \\ \vdots \\ x_{4N+3}(t) \end{pmatrix} \tag{2.69}$$

where

$$x_0(o) = z_k(o)$$
$$x_k(o) = z_k(o)$$

$$x_{N+1}(o) = 0$$

$$x_{2N+2-k}(o) = (-1)^k z_k(o)$$

$$x_{2N+2}(o) = 0$$

$$x_{2N+2+k}(o) = -z_k(o)$$

$$x_{3N+3}(o) = 0$$

$$x_{4N+4-k}(o) = (-1)^{k+1} z_k(o) \quad , \quad k=1,2,\ldots,N$$

This extension is illustrated in Figure 2.4 for a system having only two subsystems.

2.3 Examples

Circulant models have rarely been used in the literature on dynamical linear systems. Dickerson and Erickson [32], however, are authors who have employed circulant systems to investigate the stability of a closed loop of vehicles. Their model will be presented first, and then their stability results will be discussed in light of Section 2.1. The underlying physical situation is a string of a large number of vehicles moving along a single lane, e.g. a personal rapid transit system. The string is modelled as a closed loop in order to avoid difficulties associated with the ends of the string and in order to obtain controllers which are identical for each vehicle.

Figure 2.5 illustrates the vehicle traffic loop and is taken from [32]. The loop consists of N identical cars moving around a circular

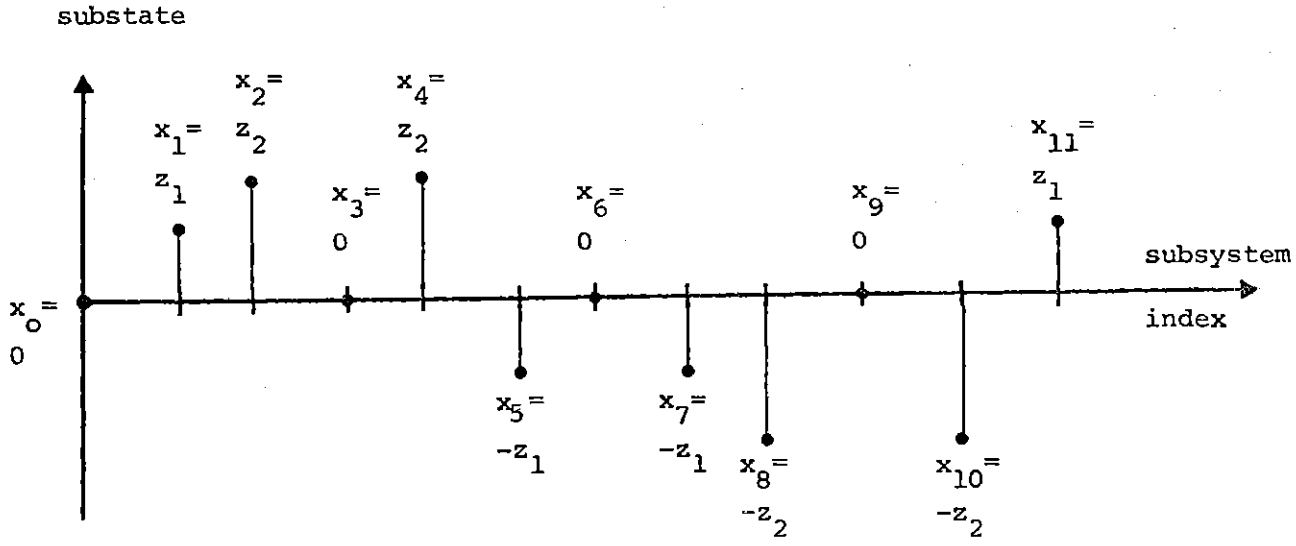


Figure 2.4 Imbedding an antisymmetric tridiagonal system with an even number of subsystems requires a circulant system roughly four times as large.

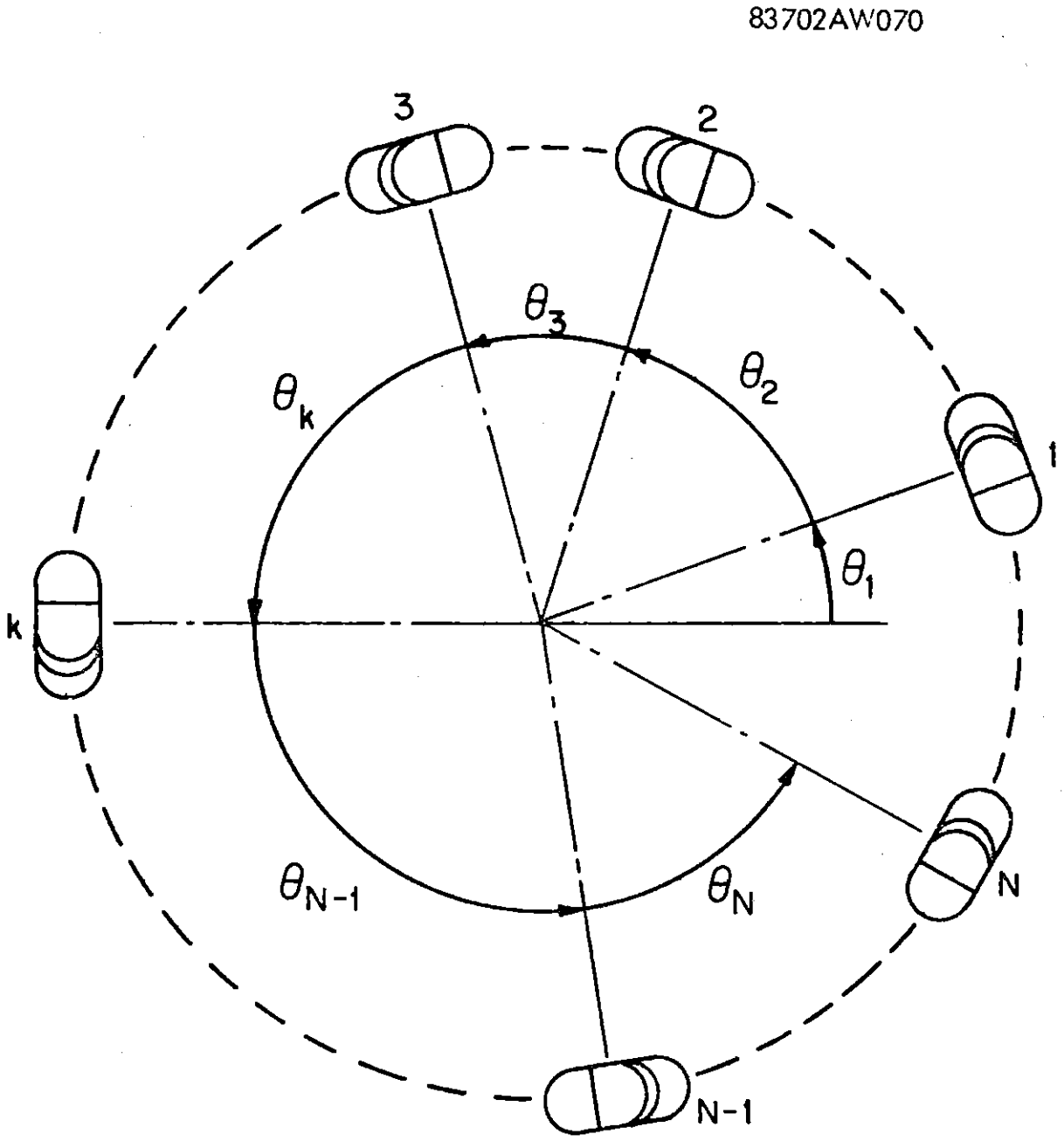


Figure 2.5 (from [32]). Illustration of the classical vehicle traffic loop.

track. It is desired to regulate the system so that the cars traverse the track with a nominal angular velocity V and a nominal separation $\frac{2\pi}{N}$. Assuming linearly proportional feedback from the position and velocity of the nearest fore and aft neighbors and its own position and velocity, a simplified description of the motion of the k^{th} car is given by

$$\frac{d^2}{dt^2}\theta_k(t) = K_f\left[\theta_{k+1}(t) - \theta_k(t) - \frac{2\pi}{N}\right] + G_f\left[\frac{d}{dt}\theta_{k+1}(t) - \frac{d}{dt}\theta_k(t)\right]$$

$$+ K_b\left[\theta_{k-1}(t) - \theta_k(t) + \frac{2\pi}{N}\right] + G_b\left[\frac{d}{dt}\theta_{k-1}(t) - \frac{d}{dt}\theta_k(t)\right]$$

$$+ K\left[Vt - \theta_k(t) + \frac{k2\pi}{N}\right] + G\left[V - \frac{d}{dt}\theta_k(t)\right] \qquad (2.70)$$

where all indices are modulo N. The K's are gains on positions, and the G's are gains on velocities. The resulting model is clearly a circulant system. Moreover, the fact that the feedback was restricted to be from the nearest neighbors is not crucial for obtaining a circulant model. The important characteristic is that the loop is closed in that cars 1 and N interact in the same way as (say) cars 1 and 2.

The infinite-dimensional version of this problem has been studied by several authors [30], [31], [47]. The problem is to regulate an infinite string of identical systems moving in a straight line. Of course the model is a Toeplitz system (the subject of Chapter 5) and not a circulant system. However, these authors claim that "the infinite object theory accurately describes the properties of the typical vehicle controller in a long finite string" [30]. Typical vehicles are those

having many vehicles in front and behind them, i.e. ones not near the ends of the string. But typical vehicles could alternatively be described by a circulant model obtained by connecting the two ends of the finite string. This procedure yields the approximate circulant model of Dickerson and Erikson which does not model the behavior of objects near the ends. But if the behavior of objects near the ends is of no particular interest, or if this behavior can be determined in some *ad hoc* fashion, or if it is desired to equip all the vehicles with the same controller (for simplicity of design, for easy replacement of any vehicle, etc.), then the circulant model could be quite useful.

The stability results of Dickerson and Erickson [32] for circulant systems will now be discussed keeping in mind the decoupling property of the DFT. That is, since the DFT converts a block circulant matrix into a block diagonal matrix, the stability of the overall system can be determined from the stability of the individual spatial subsystems. If all the subsystems are stable, then the system is stable; if one subsystem is unstable, then the system is unstable. This elementary observation together with the formula (2.13) for the diagonal blocks will explain two of the three stability results found in [32]. The third is just a direct application of Lyapunov's direct method and will not be considered further.

Dickerson and Erickson's first stability condition is given in terms of the measure μ of a matrix. For any matrix norm such that $\|I\| = 1$, the measure of a matrix A is defined by

$$\mu(A) = \lim_{h \downarrow 0} \frac{1}{h} \left[\| I + hA \| - 1 \right] \tag{2.71}$$

It is easily shown that for any eigenvalue λ of A,

$$\mathrm{Re}[\lambda] \leq \mu(A) \tag{2.72}$$

If ϕ is an eigenvector of A corresponding to λ with norm 1, then

$$\lim_{h \downarrow 0} \frac{1}{h} \left[\| (I + hA)\phi \| - \| \phi \| \right] = \lim_{h \downarrow 0} \frac{1}{h} \left[|1 + h\lambda| - 1 \right] = \mathrm{Re}[\lambda] \tag{2.73}$$

But,

$$\lim_{h \downarrow 0} \frac{1}{h} \left[\| (I + hA)\phi \| - \| \phi \| \right] \leq \lim_{h \downarrow 0} \frac{1}{h} \left[\| I + hA \| - 1 \right] = \mu(A) \tag{2.74}$$

Thus (2.72) is valid. Using this result, Dickerson and Erickson give the following stability criterion: the circulant system (2.2) is stable if

$$\mu(A_0) + \sum_{k=1}^{N-1} \| A_k \| < 0 \tag{2.75}$$

The condition (2.75) will now be obtained by considering the stability of each diagonal block. Recall that the k^{th} diagonal block is

$$\bar{A}_k = \sum_{i=0}^{N-1} A_1 \, W_N^{-ik} \tag{2.13}$$

Then

$$\mu(\bar{A}_k) = \mu\left(A_0 + \sum_{i=1}^{N-1} A_1 W_N^{-ik}\right) \qquad (2.76)$$

$$\leq \mu(A_0) + \left\| \sum_{i=1}^{N-1} A_i W_N^{-ik} \right\| \qquad \text{(triangle inequality)}$$

$$\leq \mu(A_0) + \sum_{i=1}^{N-1} \left\| A_i W_N^{-ik} \right\| \qquad \text{(triangle inequality)}$$

$$\leq \mu(A_0) + \sum_{i=1}^{N-1} \| A_i \|$$

$$< 0 \qquad \text{(by (2.75))}$$

That is, if (2.75) holds, then each subsystem is stable. This immediately implies that the overall circulant system is stable. It is obvious that the sufficient condition (2.75) for stability is simply derived by using the formula (2.13) for the diagonal blocks of a circulant matrix.

A sufficient condition for instability is also presented by Dickerson and Erickson. Consider the first diagonal block,

$$\bar{A}_0 = \sum_{i=o}^{N-1} A_i \qquad (2.77)$$

Clearly, if $\bar{A}_0$ is not a stability matrix, then the system must be unstable. $\bar{A}_0$ is not a stability matrix if there exists an eigenvalue λ of A_0 having real part greater than or equal to zero. Hence if there exists a λ with $\mathrm{Re}(\lambda) \geq 0$ such that $\lambda I - \sum_{i=o}^{N-1} A_i$ is not one-to-one, then the circulant

system (2.2) is unstable. But this last statement is just Theorem 3 in [32]. The diagonalization property of the DFT, therefore, provides an easy means of obtaining the stability conditions of Dickerson and Erickson.

Condition (2.75) has an interesting interpretation when the circulant system is viewed as an interconnection of subsystems. Since $\mu(A_0)$ is required to be negative, each isolated subsystem is stable. Moreover, the interactions A_k, k=1,...,N-1 with other subsystems are considered to be only destabilizing forces in (2.75). Given this interpretation, the sufficient condition (2.75) is reminiscent of stability formulas for diagonally dominant large-scale systems, see for example [48], where each subsystem is required to be stable and the interactions with other subsystems are only considered to be destabilizing.

In some cases, only some of the matrices A_k, k=0,...,N-1 are nonzero. For example, $A_k = 0$ for $k \neq 0,1,N-1$ corresponds to only nearest neighbor interactions. In these cases, the stability condition (2.75) and the instability condition provide tests which are independent of the number N of subsystems. Thus if condition (2.75) is satisfied, the circulant system is stable no matter how many subsystems are included. This observation was first made in [32].

As shown in Section 2.2, circulant models can be used to study symmetric and antisymmetric tridiagonal systems. There are many examples of tridiagonal systems; one is the longitudinal power system depicted in Figure 2.6. This system is similar to one analyzed by Arcidiacono,

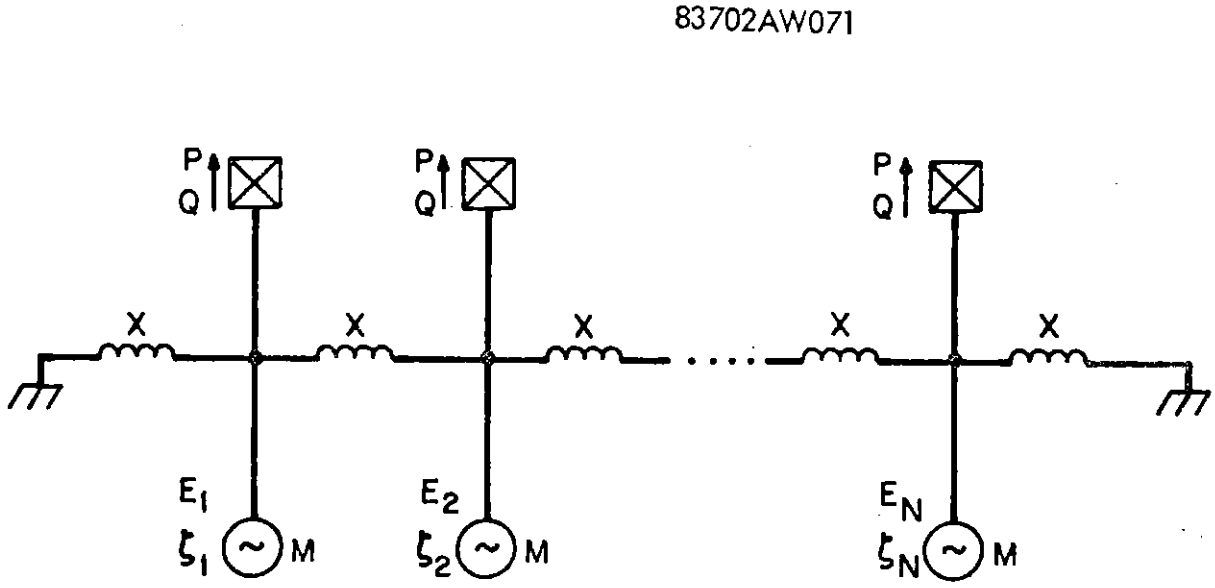


Figure 2.6 A uniform longitudinal power system consisting of N generators [49].

Ferrari, and Saccomanno [49] for low frequency electromechanical oscillations. A simplified linear model will now be obtained based on the modelling assumptions made in [49].

Each generator is modelled by an e.m.f. E_i and phase angle δ_i, and it is assumed that the E_i are constant. For small variations, each generator has the second order model

$$\Delta P_{mi} - \Delta P_{ei} = m \Delta\ddot{\delta}_i \tag{2.78}$$

where m is an inertia coefficient and ΔP_{mi} and ΔP_{ei} are changes in the mechanical and electrical power, respectively, at generator i. Linearzing, the electrical power variation ΔP_{ei} is written

$$\begin{aligned} \Delta P_{ei} &= \sum_{j=1}^{N} \left(\frac{\partial P_{ei}}{\partial \Delta\delta_j} \right) \Delta\delta_j \\ &= \sum_{j=1}^{N} K_{ij} \Delta\delta_j \end{aligned} \tag{2.79}$$

where the partial derivatives are evaluated at the nominal values of their arguments. From the static load flow equations [50],

$$K_{ij} = \begin{cases} -\frac{1}{X} E_i E_j \cos(\delta_i - \delta_j) & , j = i \pm 1 \\ 0 & , \text{otherwise} \end{cases} \tag{2.80}$$

for the uniform longitudinal system in Figure 2.6. The mechanical power variation ΔP_{mi} is expressed as

$$\Delta P_{mi} = \Delta P_{mi}^{(o)} + \Delta P_{mi}^{(r)} \tag{2.81}$$

where the second term represents speed governor action. It is assumed that in the frequency range of interest,

$$\Delta P_{mi}^{(r)} = -d\,\Delta\dot{\delta}_i \tag{2.82}$$

Using (2.79)-(2.82), the model (2.78) may be written in vector form as

$$M\Delta\ddot{\delta} + D\Delta\dot{\delta} + K\Delta\delta = \Delta P_m^{(o)} \tag{2.83}$$

where

$$M = \begin{pmatrix} m & & & \\ & m & & \\ & & \ddots & \\ & & & m \end{pmatrix}, \quad D = \begin{pmatrix} d & & & \\ & d & & \\ & & \ddots & \\ & & & d \end{pmatrix}$$

$$K = \begin{pmatrix} 0 & k & & \\ -k & 0 & k & \\ & -k & \ddots & k \\ & & -k & 0 \end{pmatrix}$$

and $k = -\left(\frac{1}{X}\right) E_1 E_2 \cos(\delta_1 - \delta_2)$. Equation (2.83) can be written in state space form by defining a state vector x of substates $x_i = \begin{pmatrix} \Delta\delta_i \\ \Delta\dot{\delta}_i \end{pmatrix}$,

$$\dot{x}(t) = F\,x(t) + B\,\Delta P_m^{(o)}(t) \tag{2.84}$$

where the system matrix F is antisymmetric tridiagonal as in (2.59) with $F_0 = \begin{pmatrix} 0 & 1 \\ 0 & d/m \end{pmatrix}$ and $F_1 = \begin{pmatrix} 0 & 0 \\ 0 & -k/m \end{pmatrix}$. This uniform longitudinal system, therefore, does have a tridiagonal linear model and can, according to Section 2.2.2, be imbedded in a circulant system.

Tridiagonal systems can also arise when a finite difference method is applied to partial differential equations. Consider the one-dimensional wave equation

$$\frac{\partial^2 v}{\partial t^2} - c^2 \frac{\partial^2 v}{\partial y^2} = E(y,t) \tag{2.85}$$

for $0 < y < L$ and $t > 0$ that describes the small displacement motion of a vibrating string of length L subject to an external force $E(y,t)$. The ends of the string are fixed, hence the boundary conditions

$$v(o,t) = v(L,t) = 0 \tag{2.86}$$

The initial conditions

$$v(y,o) = f(y) \tag{2.87}$$

$$\dot{v}(y,o) = g(y) \tag{2.88}$$

mean that at time 0 the displacement and velocity of the string are given by $f(y)$ and $g(y)$, respectively.

The method of finite differences can be used to obtain an approximate solution to the vibrating string problem. Let

$$d_i(t) = v\left(\frac{iL}{N+1}, t\right) \tag{2.89}$$

for i=1,2,...,N. Replacing the partial derivative $\frac{\partial^2 v}{\partial y^2}$ in (2.85) with the approximating difference quotient [51],

$$\frac{d_{i+1}(t) - 2d_i(t) + d_i(t)}{\left(\frac{L}{N+1}\right)^2}$$

yields the following equation:

$$\ddot{d}_i(t) = \frac{c^2(N+1)^2}{L^2}\left[d_{i+1}(t) - 2d_i(t) + d_{i-1}(t)\right] + E\left(\frac{iL}{N+1}, t\right) \tag{2.90}$$

In state space form, these equations are

$$\dot{x}(t) = F\,x(t) + B\,E(t) \tag{2.91}$$

where x(t) consists of substates $x_i(t) = \begin{pmatrix} d_i(t) \\ \dot{d}_i(t) \end{pmatrix}$, F is of the form (2.33) with

$$F_0 = \begin{pmatrix} 0 & 1 \\ \frac{-2c^2(N+1)^2}{L^2} & 0 \end{pmatrix}, \quad F_1 = \begin{pmatrix} 0 & 0 \\ \frac{c^2(N+1)^2}{L^2} & 0 \end{pmatrix}$$

and the input E(t) is

$$E(t) = \begin{pmatrix} E\left(\frac{L}{N+1}, t\right) \\ E\left(\frac{2L}{N+1}, t\right) \\ \vdots \\ E\left(\frac{NL}{N+1}, t\right) \end{pmatrix}$$

This finite difference nethod, therefore, produces a symmetric tridiagonal system that approximates the solution of the one-dimensional wave equation.

The two-dimensional wave equation

$$\frac{\partial^2 v}{\partial t^2} - c^2 \frac{\partial^2 v}{\partial y^2} - b^2 \frac{\partial^2 v}{\partial z^2} = E(Y,z,t) \tag{2.92}$$

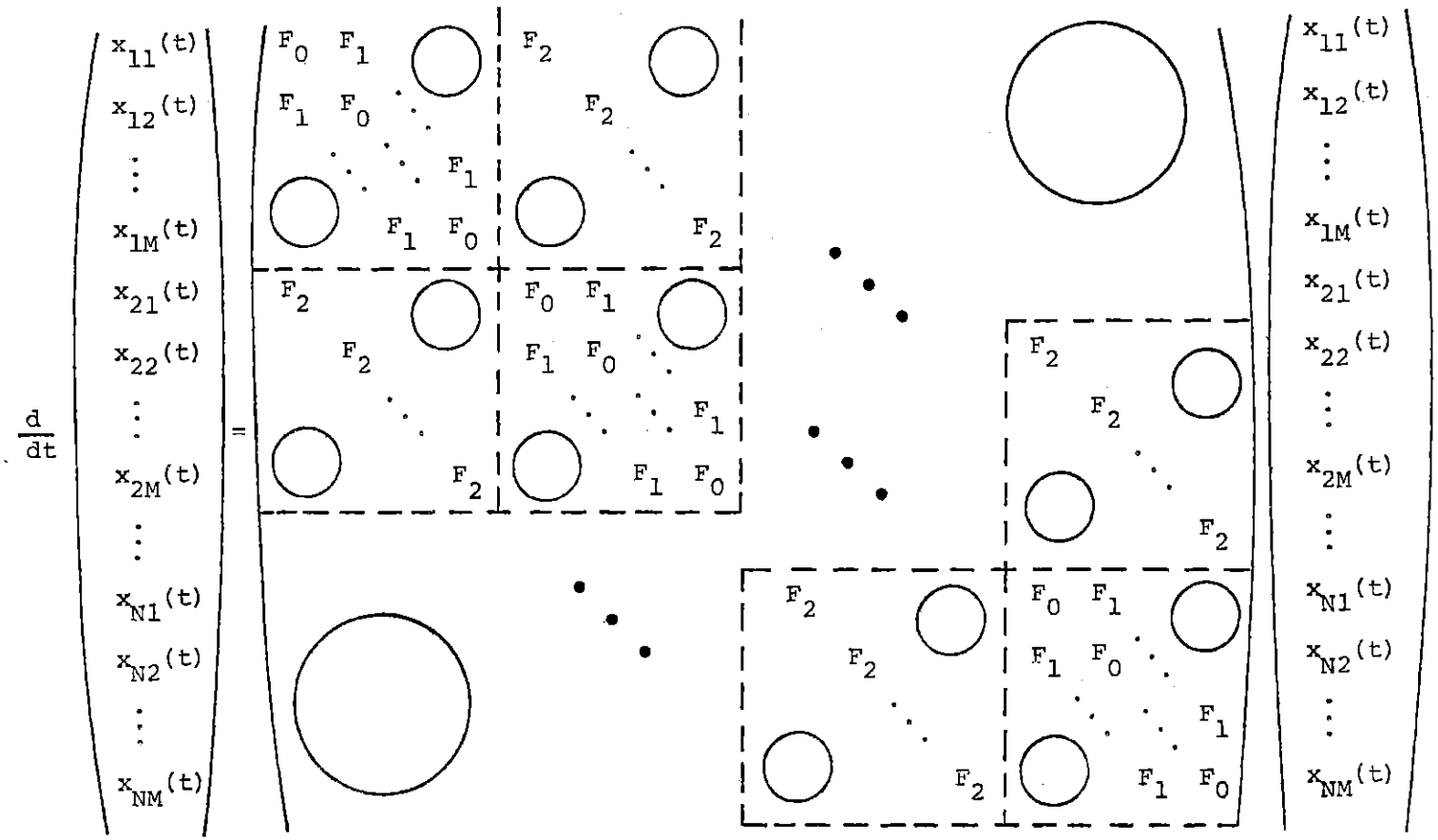
where $0 < y < 1$ and $0 < z < 1$, can be similarly approximated. This equation describes the vibrating motion of a square membrane having fixed edges. If b^2 does not equal c^2, then the membrane is nonhomogeneous in that its physical characteristics differ in the y and z directions. Let

$$d_{i,j}(t) = v\left(\frac{i}{N+1}, \frac{j}{M+1}, t\right) \tag{2.93}$$

for $i=1,2,\ldots,N$ and $j=1,2,\ldots,M$. Using difference quotients to approximate the two spatial partial derivatives in (2.92) yields

$$\begin{aligned} \ddot{d}_{i,j}(t) = {} & c^2(N+1)^2\left[d_{i+1,j}(t) - 2d_{i,j}(t) + d_{i-1,j}(t)\right] + \\ & + b^2(M+1)^2\left[d_{i,j+1}(t) - 2d_{i,j}(t) + d_{i,j-1}(t)\right] + \\ & + E\left(\frac{i}{N+1}, \frac{j}{M+1}, t\right) \end{aligned} \tag{2.94}$$

Letting $x_{i,j}(t) = \begin{pmatrix} d_{i,j}(t) \\ \dot{d}_{i,j}(t) \end{pmatrix}$ and lexicographically ordering the points yields the following lumped model:



(2.95)

where $F_0 = \begin{pmatrix} 0 & 1 \\ -2c^2(N+1)^2 - 2b^2(M+1)^2 & 0 \end{pmatrix}$, $F_1 = \begin{pmatrix} 0 & 0 \\ b^2(M+1)^2 & 0 \end{pmatrix}$, $F_2 = \begin{pmatrix} 0 & 0 \\ c^2(N+1)^2 & 0 \end{pmatrix}$

and the forcing function E has been taken to be zero. This system can be described as block tridiagonal where

- the blocks on the diagonal are themselves block tridiagonal and their super- and sub-diagonal blocks are equal
- the blocks on the super- and sub-diagonals are equal and block diagonal.

Consequently, this system has the same form as (2.50), and therefore the original 2NM dimensional system can be imbedded in a circulant system of order 8(N+1)(M+1) as described in Section 2.2.1.

2.4 System Theoretic Results

The question of controllability and observability and the problems of minimal realizations, pole placement, and state reconstruction will be examined for circulant systems. In each case, the spatial transformation of Section 2.1.2 will be used to decompose the usual tests or procedures into a series of lower order tests or procedures on complex-valued systems. Since the algebraic theory of linear systems may be performed over arbitrary fields (see [52]), considering complex systems instead of real systems presents no inherent difficulty.

2.4.1 Controllability and Observability

The block circulant pair under consideration is (A,B). Since controllability is invariant under a change of basis for either the state or the input, the pair (A,B) is controllable if and only if the pair

$(\bar{A},\bar{B})$ is controllable, i.e.

$$\text{rank}\left[B,\ AB,\ldots,A^{nN-1}B\right] = \text{rank}\left[\bar{B},\ \bar{A}\,\bar{B},\ldots,\bar{A}^{nN-1}\bar{B}\right] \qquad (2.96)$$

But the matrices $\bar{A}$ and $\bar{B}$ are block diagonal. Thus controllability of $(\bar{A},\bar{B})$ is equivalent to controllability of each pair $(\bar{A}_k,\bar{B}_k)$, $k=0,1,\ldots,N-1$. Therefore, the controllability of (A,B) can be determined from the following procedure:

(1.) use the FFT to obtain $\bar{A}$ and $\bar{B}$

(2.) test if each $(\bar{A}_k,\bar{B}_k)$ is controllable

(3.) conclude (A,B) is controllable if and only if each $(\bar{A}_k,\bar{B}_k)$ is controllable.

Because of the identities (2.23) and (2.24) presented in Section 2.1.1, subsystem k is controllable if and only if subsystem N-k is. Precisely, the controllability matrix $\left[\bar{B}_k,\bar{A}_k\,\bar{B}_k,\ldots,A_k^{n-1}\bar{B}_k\right]$ is the complex conjugate of the controllability matrix $\left[\bar{B}_{N-k},\bar{A}_{N-k}\bar{B}_{N-k},\ldots,\bar{A}_{N-k}^{n-1}\bar{B}_{N-k}\right]$. Hence the two controllability matrices have equal rank, and so $(\bar{A}_k,\bar{B}_k)$ is controllable if and only if $(\bar{A}_{N-k},\bar{B}_{N-k})$ is. This means that in step (2.) above, controllability must be examined only for $(\bar{A}_k,\bar{B}_k)$, $k=0,1,\ldots,\left\lceil\frac{N}{2}\right\rceil$ where $\left\lceil\frac{N}{2}\right\rceil$ denotes the largest integer in $\frac{N}{2}$.

By considering subsystems k and N-k together, it is possible to obtain a test involving purely real matrices. Define the linear transformation $T_\ell : C^{2\ell\times 2\ell} \to C^{2\ell\times 2\ell}$ by

$$T_\ell = \begin{bmatrix} I_\ell & I_\ell \\ jI_\ell & -jI_\ell \end{bmatrix} \qquad (2.97)$$

Then for α and β in $\mathbb{R}^{\ell \times p}$, it is to be noted that

$$T_\ell \begin{bmatrix} \alpha+j\beta & 0 \\ 0 & \alpha-j\beta \end{bmatrix} T_p^{-1} = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} \tag{2.98}$$

It is of interest to determine if the pair $(\tilde{A}_k, \tilde{B}_k)$ is controllable, where

$$\tilde{A}_k = \begin{bmatrix} \bar{A}_k & 0 \\ 0 & \bar{A}_{N-k} \end{bmatrix} \tag{2.99}$$

and $\tilde{B}_k$ is defined analogously. But from (2.23) and (2.24), $\tilde{A}_k$ and $\tilde{B}_k$ are of the form $\begin{bmatrix} \alpha+j\beta & 0 \\ 0 & \alpha-j\beta \end{bmatrix}$. Hence $(\tilde{A}_k, \tilde{B}_k)$ is controllable if and only if the <u>real</u> pair $(T_n \tilde{A}_k T_n^{-1}, T_n \tilde{B}_k T_m^{-1})$ is controllable. The test for controllability of (A,B) is, therefore, reduced to testing for the controllability of $\left\lceil \frac{N}{2} \right\rceil + 1$ real subsystems of order 2n.

An interesting special case of the above procedure is when the subsystems are scalars, i.e. A and B are composed of 1×1 blocks. In this case, the transformed subsystem pair $(\bar{A}_k, \bar{B}_k)$ is controllable if and only if $\bar{B}_k$ is nonzero. Thus controllability of (A,B) is equivalent to the DFT of the top row of B having no zero elements. Since $\bar{A}$ and $\bar{B}$ are diagonal, the image of $\bar{A}\,\bar{B}$ is contained in $\bar{B}$. The controllable subspace, therefore, is particularly easy to determine. A vector x will lie in the controllable subspace if and only if $\bar{x}_k$ is zero whenever $\bar{B}_k$ is zero.

Observability for the pair (C,A) follows by duality from considering controllability for (A',C').

2.4.2 Minimal Realizations

Given a transfer matrix $H(s)$, the realization problem is to determine a linear system defined by the triple (A,B,C) such that the transfer matrix of the system $C(sI-A)^{-1}B$ equals $H(s)$. A minimal realization is one that has the lowest possible dimension for the system matrix A. The McMillan degree [53] of a transfer function is the order of any minimal realization of the transfer function.

Consider a block circulant transfer function $H(s) \in \mathbb{R}^{pN\times mN}$. That is, $H(s)$ is partitioned into N^2 blocks $H_k(s)$ according to the pattern of (2.1) and each $H_k(s)$ is an element of $\mathbb{R}^{p\times m}$. Then $\bar{H}(s) = \Phi_p^{-1} H(s) \Phi_m$ is a complex block diagonal transfer matrix. Since they are related by a linear transformation, $H(s)$ has a finite-dimensional realization if and only if $\bar{H}(s)$ has one. If $\bar{H}(s)$ has a finite-dimensional realization, then each block $\bar{H}_k(s)$ on the diagonal is also realizable by a finite-dimensional linear system. Conversely, assume each block $\bar{H}_k(s)$ has a realization $(\bar{A}_k,\bar{B}_k,\bar{C}_k)$ where $\bar{A}_k \in \mathbb{R}^{n\times n}$, $\bar{B}_k \in \mathbb{R}^{n\times m}$, $\bar{C}_k \in \mathbb{R}^{p\times n}$. Then a realization of $\bar{H}(s)$ is given by $(\bar{A},\bar{B},\bar{C})$ where $\bar{A} = \mathrm{diag}(\bar{A}_k)$, $\bar{B} = \mathrm{diag}(\bar{B}_k)$, $\bar{C} = \mathrm{diag}(\bar{C}_k)$. Now

$$
\begin{aligned}
H(s) &= \Phi_p \, \bar{H}(s) \, \Phi_m^{-1} \qquad (2.100)\\
&= \Phi_p \left[\bar{C}(sI-\bar{A})^{-1}\bar{B} \right] \Phi_m^{-1}\\
&= \left(\Phi_p \bar{C} \, \Phi_n^{-1} \right) \left(sI - \Phi_n \bar{A} \, \Phi_n^{-1} \right)^{-1} \left(\Phi_n \bar{B} \, \Phi_m^{-1} \right)\\
&= C\left(sI-A\right)^{-1}B
\end{aligned}
$$

for $A = \Phi_n \bar{A} \Phi_n^{-1}$, $B = \Phi_n \bar{B} \Phi_m^{-1}$, $C = \Phi_p \bar{C} \Phi_n^{-1}$. If the block diagonal matrices $\bar{A}$, $\bar{B}$, $\bar{C}$ satisfy the identities (2.23) and (2.24), then A, B, C will be real-valued and block circulant. Since the original transfer function H(s) is a matrix of polynomials in s with real coefficients, the blocks $\bar{H}_k(s)$ will satisfy (2.23) and (2.24), i.e.

$$\mathrm{Re}[\bar{H}_k(s)] = \mathrm{Re}[\bar{H}_{N-k}(s)] \tag{2.101}$$

$$\mathrm{Im}[\bar{H}_k(s)] = -\mathrm{Im}[H_{N-k}(s)] \tag{2.102}$$

Therefore, the triple $(\bar{A}_{N-k}, \bar{B}_{N-k}, \bar{C}_{N-k})$ that realizes $\bar{H}_{N-k}(s)$ can be taken to be the complex conjugate of the triple $(\bar{A}_k, \bar{B}_k, \bar{C}_k)$ that realizes $\bar{H}_k(s)$. In this way, the transfer function H(s) can be realized by the real, block circulant triple (A, B, C). Summarizing, in order to determine if the block circulant transfer function H(s) has a finite-dimensional realization and, if so, to find such a realization, the following procedure may be employed:

(1.) use the FFT to obtain $\bar{H}(s)$

(2.) test if each block $H_k(s)$ has a finite-dimensional realization

-- if so, continue with step (3.)

-- if not, conclude that H(s) does not have such a realization

(3.) determine a realization $(\bar{A}_k, \bar{B}_k, \bar{C}_k)$ of $\bar{H}_k(s)$ for $k=0,1,\ldots\left\lceil\frac{N}{2}\right\rceil$

(4.) use (2.23) and (2.24) to obtain $(\bar{A}_{N-k}, \bar{B}_{N-k}, \bar{C}_{N-k})$ from $(\bar{A}_k, \bar{B}_k, \bar{C}_k)$

(5.) use the FFT to obtain (A, B, C), a realization of H(s)

It is possible to avoid having to realize a complex transfer func-

tion $\bar{H}_k(s)$ by considering the transfer functions $\bar{H}_k(s)$ and $\bar{H}_{N-k}(s)$ together and using the linear transformation T defined in Section 2.3.1. The procedure is straightforward and the details are omitted.

A circulant transfer function has been shown to have a circulant realization. But is this circulant realization minimal? No, not unless each realization $(\bar{A}_k, \bar{B}_k, \bar{C}_k)$ of $\bar{H}_k(s)$ is minimal. In general, the $\bar{H}_k(s)$ will not all have the same McMillan degree, and thus not all of the realizations will be minimal.

Proposition 2.1 The block circulant transfer function H(s) has a minimal realization as a circulant system if and only if the diagonal blocks $\bar{H}_k(s)$ all have the same McMillan degree.

The following example demonstrates what can go wrong.

Example 2.1 Consider the circulant transfer function $H(s) = \begin{bmatrix} \frac{s+1}{s^2} & \frac{1}{s^2} \\ \frac{1}{s^2} & \frac{s+1}{s^2} \end{bmatrix}$.

Then

$$\Phi^{-1} H(s) \Phi = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \frac{s+1}{s^2} & \frac{1}{s^2} \\ \frac{1}{s^2} & \frac{s+1}{s^2} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{s+2}{s^2} & 0 \\ 0 & \frac{1}{s} \end{bmatrix}$$

In this case, $\bar{H}_o(s)$ has a second-order realization and $\bar{H}_1(S)$ has a first order realization. Thus H(s) does not have any circulant minimal realization.

Even in the cases where no minimal circulant realization exists, however, there are nonminal circulant realizations.

2.4.3 Pole Allocation and State Reconstruction

Since pole allocation and state reconstruction are known to be dual problems, only the pole allocation problem will be explicitly dealt with here. All the results, of course, are applicable to the state reconstruction or observer problem.

The pole allocation problem is to determine a full-state feedback map $G : x(t) \to u(t)$ such that the resulting closed-loop system matrix (A-BG) has specified eigenvalues. It is well known that the closed-loop eigenvalues can be specified arbitrarily if and only if the system is controllable. Actually, to be precise, if only a real-valued gain matrix G is permitted, then the specified closed-loop eigenvalues must occur in complex conjugate pairs. Consider then a controllable circulant pair (A,B), and suppose that the desired closed-loop eigenvalues are the set Λ. It is assumed that if the complex number α is in Λ, then $\alpha^* \in \Lambda$.

Since (A,B) is controllable, the transformed pair $(\bar{A}, \bar{B})$ is controllable, and so is each pair $(\bar{A}_k, \bar{B}_k)$. Hence it is possible to assign the poles of each transformed subsystem. Let λ_k denote the set of closed loop poles which are to be associated with the k^{th} transformed subsystem.

Of course, $\Lambda = \bigcup_{k=o}^{N-1} \lambda_k$. Given λ_k, the pole placement problem is solved for each pair $(\bar{A}_k, \bar{B}_k)$ to yield the complex feedback gain matrix $\bar{G}_k$. If $\bar{G} = \text{diag}(\bar{G}_k)$, then $G = \Phi_m \bar{G} \Phi_n^{-1}$ solves the overall problem, i.e. the set of eigenvalues of (A-BG) equals Λ. The feedback matrix obtained by this procedure will, in general, be complex. In order to obtain a real-valued G, the transformed subsystem gains $\bar{G}_k$ must obey (2.23) and (2.24). This means the closed-loop poles of subsystem N-k must be the complex conjugates of the closed-loop poles of subsystem k, i.e. $\lambda_{N-k} = \{\alpha^* | \alpha \in \lambda_k\}$. If the sets λ_k are chosen such that this requirement is met, then the gains $\bar{G}_k$ will satisfy (2.23) and (2.24). The resulting state feedback gain matrix G will be real-valued and block circulant. The procedure for pole placement for circulant systems is

(1.) use the FFT to obtain $(\bar{A}, \bar{B})$

(2.) choose the sets λ_k such that $\lambda_{N-k} = \{\alpha^* | \alpha \in \lambda_k\}$ and $\Lambda = \bigcup_{k=o}^{N-1} \lambda_k$

(3.) solve the pole placement problem for each subsystem, thereby obtaining $\bar{G}_k$

(4.) use the FFT to obtain G

Just as in the test for controllability or the procedure for obtaining a minimal realization, pole placement can be done without considering complex systems by treating subsystems k and N-k together. Once again, however, the details are omitted.

In the case of scalar subsystems, the pole allocation problem becomes

particularly easy. Suppose the desired closed-loop poles are λ_k, k=0,1,...,N-1. Note that it is possible to specify not only these eigenvalues but also which particular closed-loop eigenvector, as given by (2.7), is associated with each eigenvalue. Assuming it is desired to associate λ_o with ϕ_o, etc., then the transformed closed-loop system matrix is

$$\bar{A} - \bar{B}\,\bar{G} = \Phi_n^{-1} A \Phi_n - \Phi_n^{-1} B \Phi_m \Phi_m^{-1} G \Phi_n \qquad (2.103)$$

$$= \begin{pmatrix} \bar{a}_0 - \bar{b}_0\bar{g}_0 & & & \\ & \bar{a}_1 - \bar{b}_1\bar{g}_1 & & \\ & & \bar{a}_2 - \bar{b}_2\bar{g}_2 & \\ & & & \ddots \\ & & & & \bar{a}_{N-1} - \bar{b}_{N-1}\bar{g}_{N-1} \end{pmatrix}$$

$$= \begin{pmatrix} \lambda_0 & & & \\ & \lambda_1 & & \\ & & \lambda_2 & \\ & & & \ddots \\ & & & & \lambda_{N-1} \end{pmatrix}$$

The transformed feedback gains are simply

$$\bar{g}_k = \frac{\bar{a}_k - \lambda_k}{\bar{b}_k} \quad , \quad k=0,1,\ldots,N-1 \qquad (2.104)$$

Note that this requires all of the $\bar{b}_k$ to be nonzero. But this is exactly the condition for controllability of the system as seen in Section 2.3.1. The scalar case, therefore, clearly illustrates both the necessity and the sufficiency of controllability for pole allocation.

2.5 Decomposition of Lyapunov and Riccati Equations via the Spatial Transformation

Large computational savings can arise in the solution of Lyapunov and Riccati equations for circulant systems. These savings are obtained by using the transformation introduced in Section 2.1.2 to decompose the original equation into a series of uncoupled lower-order equations of the same type. The lower-order problems are then solved independently, and the solutions are combined to yield the overall solution. Only Lyapunov and Riccati equations are explicitly considered, but the same technique works for other (e.g. Sylvester) equations.

Consider first the time-varying Lyapunov equation

$$\frac{d}{dt} P(t) = AP(t) + P(t)A' + Q \tag{2.105}$$

where A, Q, and P(0) are all block circulant matrices of order N. Pre- and post-multiplying by Φ_n^{-1} and Φ_n, respectively, yields

$$\frac{d}{dt}[\Phi_n^{-1} P(t) \Phi_n] = [\Phi_n^{-1} A \Phi_n][\Phi_n^{-1} P(t) \Phi_n] + [\Phi_n^{-1} P(t) \Phi_n][\Phi_n^{-1} A' \Phi_n] +$$

$$+ [\Phi_n^{-1} Q \Phi_n]$$

$$\frac{d}{dt} \overline{P(t)} = \bar{A}\, \overline{P(t)} + \overline{P(t)}\, (\bar{A}') + \bar{Q} \tag{2.106}$$

Since A is circulant, so is A'. Hence $\overline{(A')}$ as well as $\bar{A}$ and $\bar{Q}$, is block diagonal. Together with the fact that $\overline{P(0)}$ is block diagonal, this implies that $\overline{P(t)}$ is also block diagonal for all t. In other words, P(t) is a block circulant matrix. Of course, the conclusion that P(t) must be block circulant can be arrived at by purely physical reasoning.

Recall that the matrix $\overline{(A')}$ in (2.106) is not equal to $\bar{A}'$, but by (2.32), equals $\bar{A}^*$. Therefore (2.106) is equivalent to the following set of uncoupled complex-valued Lyapunov equations:

$$\frac{d}{dt}\bar{P}_k(t) = \bar{A}_k\bar{P}_k(t) + \bar{P}_k(t)\bar{A}_k^*(t) + Q_k \quad , k = 0,1,\ldots, N-1 \tag{2.107}$$

The solution P(t) of the original equation (2.105) is simply

$$P(t) = \Phi_n\bar{P}(t)\Phi_n^{-1} \tag{2.108}$$

where $\bar{P}(t) = \text{diag}[\bar{P}_k(t)]$.

The Riccati equation is decomposed in a completely analogous fashion. Consider the equation

$$\frac{d}{dt}P(t) = A\,P(t) + P(t)A' + Q - P(t)C'R^{-1}CP(t) \tag{2.109}$$

where the matrices A,C,Q,R, and P(0) are all block circulant of order N. Following the same approach as that used for the Lyapunov equation yields

$$\frac{d}{dt}\bar{P}(t) = \bar{A}\,\bar{P}(t) + \bar{P}(t)\,(\bar{A}') + \bar{Q} - \bar{P}(t)(\bar{C}')\,(\overline{R^{-1}})\,\bar{C}\,\bar{P}(t) \tag{2.110}$$

Thus, P(t) is a block circulant matrix, as was clear from the symmetry of the physical problem underlying (2.109). Moreover, since $(\overline{R^{-1}}) = (\bar{R})^{-1}$, the transformed Riccati equation may be written as the following set of lower-order equations:

$$\frac{d}{dt}\bar{P}_k(t) = \bar{A}_k\bar{P}_k(t) + \bar{P}_k(t)\bar{A}_k^* + \bar{Q}_k - \bar{P}_k(t)\bar{C}^*\bar{R}^{-1}\bar{C}\,\bar{P}_k(t) \quad , k=0,1,\ldots,N-1 \tag{2.111}$$

The quantity $P(t) = \Phi_n\bar{P}(t)\Phi_n^{-1}$ is the solution of (2.109).

If the matrices A,C,Q,R in the Lyapunov and Riccati equations are all real-valued matrices, then the solution must be real-valued also. Thus, if the lower-order complex-valued Lyapunov and Riccati equations (2.105) and (2.109) are solved for $k = 0,1,\ldots,\left\lceil\frac{N}{2}\right\rceil$, then the identities (2.23) and (2.24) can be used to obtain the remaining $N - \left\lceil\frac{N}{2}\right\rceil - 1$ terms.

Lyapunov and Riccati equations for circulant systems have been shown to decompose into a set of uncoupled lower-order Lyapunov and Riccati equations, respectively. These lower-order equations, however, are complex-valued equations. Real-valued equations suitable for solution using standard computer software can be obtained by considering the k^{th} and $(N-k)^{th}$ equations together. In the case of the Lyapunov equation, these two equations can be trivially written as a complex Lyapunov equation of order 2n:

$$\frac{d}{dt}\begin{bmatrix}\bar{P}_k(t) & 0\\ 0 & \bar{P}_{N-k}(t)\end{bmatrix} = \begin{bmatrix}\bar{A}_k & 0\\ 0 & \bar{A}_{N-k}\end{bmatrix}\begin{bmatrix}\bar{P}_k(t) & 0\\ 0 & \bar{P}_{N-k}(t)\end{bmatrix} +$$

$$+\begin{bmatrix}\bar{P}_k(t) & 0\\ 0 & \bar{P}_{N-k}(t)\end{bmatrix}\begin{bmatrix}\bar{A}_k & 0\\ 0 & \bar{A}_{N-k}\end{bmatrix}+\begin{bmatrix}\bar{Q}_k & 0\\ 0 & \bar{Q}_{N-k}\end{bmatrix} \tag{2.112}$$

Pre- and post-multiplying by the matrices T_n and T_n^{-1}, respectively, that were introduced in Section 2.3.1 yields:

$$\frac{d}{dt}\begin{bmatrix} \mathrm{Re}\bar{P}_k(t) & \mathrm{Im}\bar{P}_k(t) \\ -\mathrm{Im}\bar{P}_k(t) & \mathrm{Re}\bar{P}_k(t) \end{bmatrix} = \begin{bmatrix} \mathrm{Re}\bar{A}_k & \mathrm{Im}\bar{A}_k \\ -\mathrm{Im}\bar{A}_k & \mathrm{Re}\bar{A}_k \end{bmatrix} \begin{bmatrix} \mathrm{Re}\bar{P}_k(t) & \mathrm{Im}\bar{P}_k(t) \\ -\mathrm{Im}\bar{P}_k(t) & \mathrm{Re}\bar{P}_k(t) \end{bmatrix} +$$

$$+ \begin{bmatrix} \mathrm{Re}\bar{P}_k(t) & \mathrm{Im}\bar{P}_k(t) \\ -\mathrm{Im}\bar{P}_k(t) & \mathrm{Re}\bar{P}_k(t) \end{bmatrix} \begin{bmatrix} \mathrm{Re}\bar{A}_k' & -\mathrm{Im}\bar{A}_k' \\ \mathrm{Im}\bar{A}_k' & \mathrm{Re}\bar{A}_k' \end{bmatrix} + \begin{bmatrix} \mathrm{Re}\bar{Q}_k & \mathrm{Im}\bar{Q}_k \\ -\mathrm{Im}\bar{Q}_k & \mathrm{Re}\bar{Q}_k \end{bmatrix} \qquad (2.113)$$

where (2.23) and (2.24) have been used. In its present form, (2.113) is not a Lyapunov equation, since the driving matrix and the unknown matrix are not symmetric. But (2.26) and (2.27) demand that $\mathrm{Re}\bar{P}_k(t) = \mathrm{Re}\bar{P}_k'(t)$ and $\mathrm{Im}\bar{P}_k(t) = -\mathrm{Im}\bar{P}_k'(t)$, so (2.113) becomes

$$\frac{d}{dt}\begin{bmatrix} \mathrm{Re}\bar{P}_k(t) & \mathrm{Im}\bar{P}_k(t) \\ \mathrm{Im}\bar{P}_k'(t) & \mathrm{Re}\bar{P}_k(t) \end{bmatrix} = \begin{bmatrix} \mathrm{Re}\bar{A}_k & \mathrm{Im}\bar{A}_k \\ -\mathrm{Im}\bar{A}_k & \mathrm{Re}\bar{A}_k \end{bmatrix} \begin{bmatrix} \mathrm{Re}\bar{P}_k(t) & \mathrm{Im}\bar{P}_k(t) \\ \mathrm{Im}\bar{P}_k'(t) & \mathrm{Re}\bar{P}_k(t) \end{bmatrix} +$$

$$+ \begin{bmatrix} \mathrm{Re}\bar{P}_k(t) & \mathrm{Im}\bar{P}_k(t) \\ \mathrm{Im}\bar{P}_k'(t) & \mathrm{Re}\bar{P}_k(t) \end{bmatrix} \begin{bmatrix} \mathrm{Re}\bar{A}_k & \mathrm{Im}\bar{A}_k \\ -\mathrm{Im}\bar{A}_k & \mathrm{Re}\bar{A}_k \end{bmatrix}' + \begin{bmatrix} \mathrm{Re}\bar{Q}_k & \mathrm{Im}\bar{Q}_k \\ \mathrm{Im}\bar{Q}_k' & \mathrm{Re}\bar{Q}_k \end{bmatrix} \qquad (2.114)$$

a standard Lyapunov equation of order 2n. The set of complex-valued Riccati equations (2.111) can be converted into real-valued equations analogous to (2.114) in exactly the same way.

It is quite clear that the algebraic Lyapunov and Riccati equations which give the steady-state solution to (2.105) and (2.109) can also be decomposed into complex-valued algebraic equations and then combined into real-valued equations of order 2n.

Just how large are the computational savings associated with decomposing a circulant Lyapunov or Riccati equation? In either case, standard solution techniques require $O(n^3N^3)$ operations, i.e., the number of operations is proportional to $(nN)^3$ for large values of (nN). In the circulant case, the DFT takes $O(n^2N\log_2 N)$ operations, and the solution of each transformed equation takes $O(n^3)$ operations. The total computations required are, therefore, $O(n^2N\log N + n^3N)$. Thus, for a fixed subsystem order, the necessary computations increase only as $N\log_2 N$ in the circulant case. This presents a tremendous saving over the $O(N^3)$ operations generally required.

2.6 Summary and Conclusions

Circulant systems have been introduced in this chapter, and their properties have been developed. The key element here has been the use of the DFT to diagonalize a circulant matrix. For a circulant system, this transformation yields a set of uncoupled spatial subsystems. The diagonalizing property of the DFT was exploited to study system theoretic issues in Section 2.4, and to decompose Lyapunov and Riccati equations in Section 2.5.

Both symmetric and antisymmetric tridiagonal systems can be imbedded in circulant systems by the methods described in Section 2.2. Section 2.3 contained some examples of circulant and tridiagonal systems, and also discussed some stability results of Dickerson and Erickson for circulant systems in view of the diagonalization of circulants by the DFT. Throughout this chapter, only continuous-time systems were explicitly considered, but the same techniques also work in discrete time.

Because of the decomposition of Lyapunov and Riccati equations, control and estimation problems for circulant, large-scale systems can be solved very efficiently off-line. The on-line implementation of these solutions, however, employs a centralized controller or estimator. Chapter 3 will consider using the spatial frequency domain for

- efficient implementation of centralized processors
- design of decentralized processors

In both cases, the fact that circulant matrices are diagonalized by the DFT will be crucial.

For a large-scale system to be circulant, it is necessary that all of its subsystems are identical. In an actual large-scale system, it might be the case that the subsystems are similar, but not identical. One could consider approximating the large-scale system by a circulant system in such a case, however, because of the tremendous computational advantages associated with circulant systems. Of course, the resulting controller or estimator will be suboptimal. If the primary interest is in interactions among subsystems, as opposed to the detailed behavior of individual subsystems, then it is expected that the approximate circulant model could be very helpful. Some preliminary results along this line are found in Chapter 3.

CHAPTER 3

CONTROL AND ESTIMATION FOR CIRCULANT SYSTEMS

The linear-quadratic regulator problem and its dual filtering problem are now considered for circulant systems. Section 3.1 deals with the centralized regulator and uses the spatial transformation to decompose the problem and to obtain an efficient on-line implementation of the control law. Section 3.2 attacks the fixed-structure decentralized control problem. Both optimal and suboptimal decentralized control gains are considered in this section. Using the imbedding procedure introduced in Section 2.2, centralized and decentralized control laws for a rectangular membrane are computed in Section 3.3 to illustrate the preceding development. Some preliminary work toward using circulant controllers for a general large-scale system is presented in Section 3.4. The filtering problem is then treated in Section 3.5 by duality. The chapter concludes with a summary and brief discussion in Section 3.6.

The decomposition of the centralized circulant control problem in Section 3.1 is essentially just the finite-dimensional analog of the procedure used by Melzer and Kuo [30] to decompose Toeplitz control problems (see Section 5.1.2). Where the spatial transform concepts are used in a fundamentally new way is in Section 3.2, more specifically, in Section 3.2.3, to obtain suboptimal decentralized feedback gains from the optimal centralized gains. From the perspective of the spatial frequency response, this problem is observed to be analogous to the design of finite-impulse response digital filters from infinite-impulse filters. Thus, some digital filter design techniques are adapted in Section 3.2.3 to the design of decentralized controllers.

3.1 Centralized Control

The vehicle used for studying centralized control of circulant systems will be the standard linear-quadratic regulator problem. For the circulant system (2.2), the regulator problem consists of the determination of the input time function u(t) which minimizes the quadratic cost functional

$$J = \frac{1}{2} x'(T)Fx(T) + \frac{1}{2}\int_0^T x'(t)Q(t)x(t) + u'(t)R(t)u(t)\, dt \qquad (3.1)$$

where $F \geq 0$ and for all $t \in [0,T]$, $Q(t) \geq 0$ and $R(t) > 0$. The optimizing input u(t) can be expressed in linear state-variable feedback form as

$$u(t) = - G^c(t)x(t) \qquad (3.2)$$

where $G^c(t)$ is a time-varying gain matrix. The centralized gain is given by

$$G^c(t) = R^{-1}(t)B'K(t) \qquad (3.3)$$

where K(t) is the unique symmetric matrix solving the Riccati equation

$$\frac{d}{dt} K(t) = - K(t)A - A'K(t) - Q(t) + K(t)B\, R^{-1}(t)B'K(t) \qquad (3.4)$$

subject to $K(T) = F$.

The problem is called a regulator problem because the objective is to choose the input so as to regulate the state near zero. The weighting matrices, F,Q(t) and R(t) reflect the tradeoff between deviations of the state from zero and the level of control used to reduce these deviations. Consider

first the state weighting matrix $Q(t)$, and suppose it is partitioned into N^2 $n \times n$ submatrices;

$$Q(t) = \begin{pmatrix} Q_{11}(t) & Q_{12}(t) & \cdots & Q_{1N}(t) \\ Q_{21}(t) & Q_{22}(t) & \cdots & Q_{2N}(t) \\ \vdots & \vdots & & \vdots \\ Q_{N1}(t) & Q_{N2}(t) & & Q_{NN}(t) \end{pmatrix} \tag{3.5}$$

The state vector $x(t)$ of the circulant system is composed of N substates, x_i, $i = 0,1,\ldots,N-1$, and the weighting matrix for each substate x_i is $Q_{ii}(t)$. As discussed in Chapter 2, all of the subsystems are identical. Therefore, it is quite reasonable to propose that all the blocks $Q_{ii}(t)$ on the diagonal of $Q(t)$ be equal. In general, the block $Q_{ik}(t)$ is the cross-weighting between subsystems i and k; i.e., the term $x_i'(t)Q_{ik}(t)x_k(t)$ appears in the integrand of (3.1). The dynamic interactions between subsystems i and k, however, depend only on $(i-k) \bmod N$. Hence, it is proposed the $Q_{ik}(t)$ should be a function of only $(i-k) \bmod N$. That is to say, in the case of circulant systems, it is physically reasonable to restrict the state weighting matrix $Q(t)$ to being a block circulant matrix. Similar arguments, of course, apply to the weighting matrices F and $R(t)$. Therefore, throughout this chapter, all of the weighting matrices in the quadratic cost functional (3.1) will be taken to be block circulant matrices.

Since all of the matrices in the Riccati equation (3.4) are block circulant, the analysis of Section 2.5 can be applied to decompose (3.4)

into the following set of Riccati equations of order n:

$$\frac{d}{dt}\bar{K}_i(t) = -\bar{K}_i(t)\bar{A}_i - \bar{A}_i^*\bar{K}_i(t) - \bar{Q}_i(t) + \bar{K}_i(t)\bar{B}_i\bar{R}_i^{-1}(t)\bar{B}_i^*\bar{K}_i(t) \tag{3.6}$$

subject to the terminal condition $\bar{K}_i(T) = \bar{F}_i$. Moreover, the gain matrix $G^c(t)$ is also circulant and can be computed as

$$G^c(t) = \Phi_m \bar{G}^c(t) \Phi_n^{-1} \tag{3.7}$$

where the i^{th} block $\bar{G}_i^c(t)$ of the block diagonal matrix $\bar{G}^c(t)$ is given by

$$\bar{G}_i^c(t) = \bar{R}_i^{-1}(t)\bar{B}_i^*\bar{K}_i(t) \quad . \tag{3.8}$$

It is clear that the circulant quadratic regulator requires much less off-line computational effort to compute the gain matrix $G^c(t)$ than is required in the general case.

The on-line implementation of this feedback law can also be performed efficiently by using the FFT. Consider the discrete-time version of the regulator problem where at time k the input u(k) equals $-G^c(k)x(k)$. Since $G^c(k)$ is a circulant matrix, this matrix multiplication can be performed by

(i) using the FFT to compute the $\bar{x}_i(k)$ from the $x_i(k)$

(ii) computing $\bar{u}_i(k) = -\bar{G}_i^c(k)\bar{x}_i(k)$

(iii) using the inverse FFT to compute the $u_i(k)$ from the $\bar{u}_i(k)$.

Figure 3.1 illustrates the structure of this centralized controller.

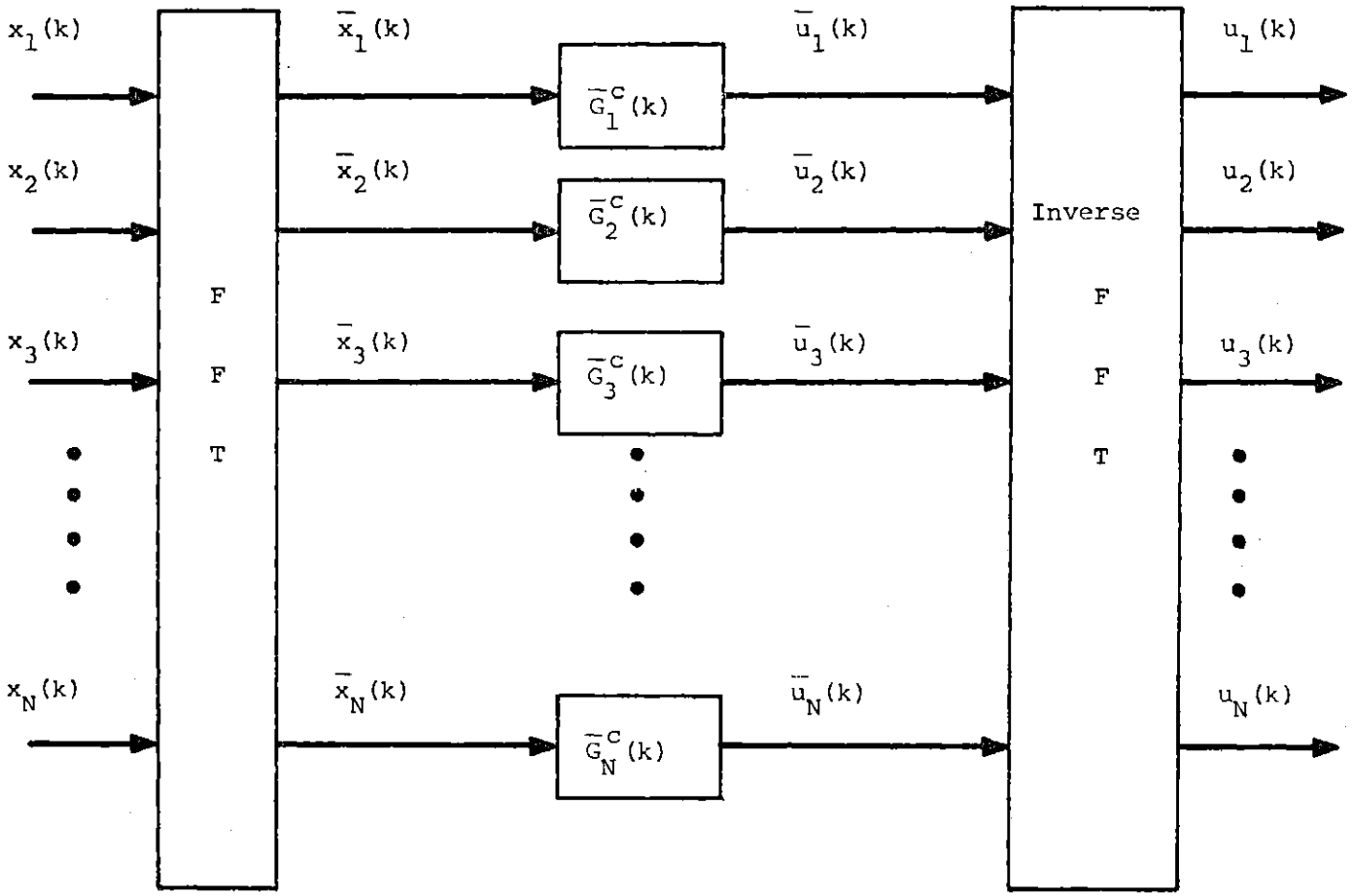


Figure 3.1 By using the FFT, the circulant regulator has an efficient on-line implementation employing parallel processing.

The computational advantages of the circulant regulator can be more clearly understood by recalling from (2.20) that the transformed circulant system is composed of N uncoupled subsystems,

$$\frac{d}{dt}\bar{x}_i(t) = \bar{A}_i\bar{x}_i(t) + \bar{B}_i\bar{u}_i(t) \qquad i = 0,1, \ldots ,N-1 \qquad (3.9)$$

Specifying the circulant cost matrices $F,Q(t)$, and $R(t)$ is the same as specifying the cost matrices $\bar{F}_i,\bar{Q}_i(t)$, and $\bar{R}_i(t)$ for the individual transformed subsystems. This means that restricting the cost matrices in (3.1) to being circulant is equivalent to proposing independent subproblems for each of the N transformed subsystems. The minimizing control $\bar{u}_i(t)$ for the i^{th} subproblem is just

$$\bar{u}_i(t) = -\bar{G}_i^c(t)\bar{x}_i(t) \qquad (3.10)$$

where the $\bar{G}_i^c(t)$ are given by (3.8).

A time-invariant controller can be found by considering the cost functional

$$J = \frac{1}{2}\int_0^\infty x'(t)Q\,x(t) + u'(t)Ru(t)\,dt \qquad (3.11)$$

where $Q \geq 0$ and $R > 0$. Under the conditions that (A,B) is controllable and $(A,Q^{\frac{1}{2}})$ is observable, the minimizing $u(t)$ is given by

$$u(t) = -G^c\,x(t) \qquad (3.12)$$

where the constant feedback gain G^c equals $-R^{-1}B'K$ and K is the unique positive definite matrix solving the algebraic Riccati equation

$$0 = -A'K - KA - Q + KBR^{-1}B'K \qquad (3.13)$$

As discussed in Section 2.4.1, (A,B) controllable and $(A,Q^{\frac{1}{2}})$ observable implies $(\bar{A}_i,\bar{B}_i)$ controllable and $(\bar{A}_i,\bar{Q}_i^{\frac{1}{2}})$ observable for all i. The low-order algebraic Riccati equations

$$0 = -\bar{A}_i^*\bar{K}_i - \bar{K}_i\bar{A}_i - \bar{Q}_i + \bar{K}_i\bar{B}_i\bar{R}_i^{-1}\bar{B}_i^*\bar{K}_i \tag{3.14}$$

therefore, have unique positive definite solutions $\bar{K}_i$. Of course, the solution K of (3.13) can be easily obtained from the $\bar{K}_i$.

The time-invariant state feedback law (3.12) is centralized in that each subinput $u_i(t)$ will, in general, depend upon every substate $x_k(t)$, k = 0,1, ..., N-1. The decentralized linear state feedback problem is the subject of the next section.

3.2 Decentralized Control

3.2.1 Introduction

Under a decentralized state feedback control law, each subinput is a function of some fixed subset of the substates. Only linear, non-dynamic feedback maps will be considered here. This means that the structure of the decentralized controller is fixed so as to require the input $u(t)$ to be of the form

$$u(t) = - G^d x(t) \tag{3.15}$$

where the decentralized feedback matrix G^d is block circulant and subject to the constraints

$$G_i^d = 0 \qquad \forall \; i \notin \alpha \subset \{0,1, \ldots, N-1\} \tag{3.16}$$

Under these conditions, the subinput $u_o(t)$ is just a linear combination of the substates $x_i(t)$, $i \in \alpha$. In fact, for any k, the subinput $u_k(t)$ is the same linear combination of the substates $x_{(i+k) \bmod N}(t)$, $i \in \alpha$.

The set α specifies which substates can be used in the computation of the individual subinputs. The usual situation is that the number of elements in α will be much less than N. As an example, consider the case of decentralized control with only nearest neighnor feedback. This corresponds to $\alpha = \{0,1,N-1\}$, and so the gain matrix G^d is constrainted to have the form

$$G^d = \begin{pmatrix} G_o^d & G_{N-1}^d & \cdots & G_i^d \\ G_1^d & G_o^d & G_{N-1}^d & \\ & G_1^d & \ddots & \\ & & & G_{N-1}^d \\ G_{N-1}^d & & G_i^d & G_o^d \end{pmatrix} \qquad (3.17)$$

Given the set α, the remaining problem is to determine the feedback gains G_i^d, $i \in \alpha$. Section 3.2.2 proposes choosing the gains which minimize the infinite-horizon cost functional (3.12). The approach taken in §3.2.2 is to obtain a set of necessary conditions for the optimal decentralized gains. Suboptimal controllers are considered in Section 3.2.3. The spatial transform domain is used to obtain suboptimal decentralized feedback gains from the optimal centralized gains of Section 3.1

3.2.2 Optimal Decentralized Controller

It is desired to find the feedback matrix G^d which minimizes the cost functional (3.12) subject to the constraints (3.16). Substituting $u(t) = -G^d x(t)$ into (3.12) yields the quadratic cost functional

$$J = \frac{1}{2}\int_o^\infty x'(t)[Q + G^{d'}RG^d]\, x(t)\, dt \qquad (3.18)$$

Minimization of (3.18) will yield a result which, in general, depends upon the initial state $x(0)$. In order to obtain a decentralized feedback matrix G^d which is not a function of the initial state, it is assumed that $x(0)$ is a zero-mean random variable with circulant covariance Σ_o. The problem then is to find the G^d (if it exists) which minimizes the

expected value of the cost (3.18).

Many authors have solved various versions of closely related problems [54] - [61]. In [61] it is shown that

$$E\{J\} = \frac{1}{2} \text{trace } [P(Q + G^{d'} R G^{d})] \tag{3.19A}$$

$$= \frac{1}{2} \text{trace } [K\Sigma_o] \tag{3.19B}$$

where the matrices P and K obey

$$(A - BG^{d}) P + P(A - BG^{d})' + \Sigma_o = 0 \tag{3.20}$$

$$(A - BG^{d})'K + K(A - BG^{d}) + (Q + G^{d'} RG^{d}) = 0 \quad . \tag{3.21}$$

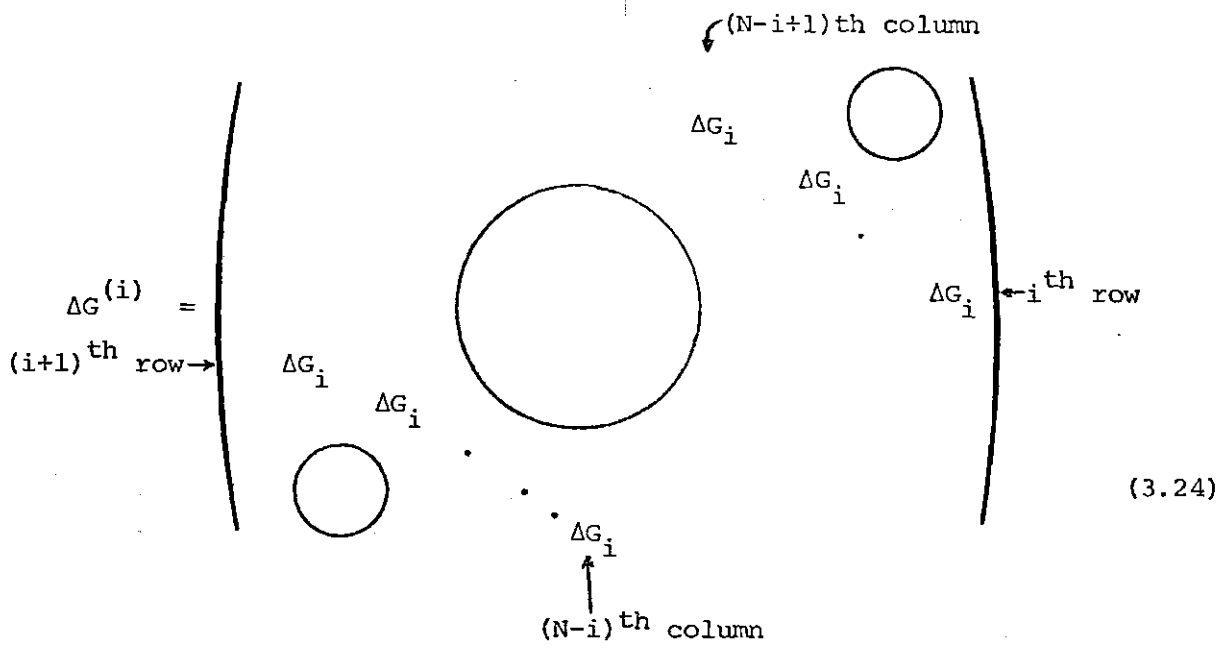
It is necessary that the optimal decentralized gain G^{d} satisfy [ ]:

$$\frac{\partial}{\partial \varepsilon} E[J(G_i^{d} + \Delta G_i]\Big|_{\varepsilon=0} = 0 \qquad \forall \ \Delta G_i \ \forall \ i \in \alpha \quad . \tag{3.22}$$

Using (3.19), it can be shown [61] that the necessary condition (3.22) is equivalent to

$$\text{trace } \{ P[- KB\Delta G^{(i)} - \Delta G^{(i)} B'K + G^{d'} R\Delta G^{(i)} + \Delta G^{(i)'} RG^{d}]\} = 0 \tag{3.23}$$

where $\Delta G^{(i)}$ is used to denote



(3.24)

The condition (3.23) can be manipulated further and written componentwise as

$$\sum_{k=0}^{N-1} P_{-i-k} \left[\sum_{\ell=0}^{N-1} G^{d'}_{-\ell} R_{k-\ell} - K_\ell B_{k-\ell} \right] = 0 \qquad \forall \; i \in \alpha . \tag{3.25}$$

The necessary conditions (3.20), (3.21), (3.25) for the optimal decentralized feedback gain G^d are seen to be a set of coupled nonlinear equations.

From the Lyapunov equation (3.20), the closed-loop system is stable if and only if P is positive definite. If $(A - BG^d)$ is unstable, then G^d is certainly <u>not</u> minimizing, even though it satisfies the necessary conditions, since the expected value of the cost (3.18) is infinite. It may be the case that no stabilizing gain G^d exists, and hence the optimization problem is ill-posed. This fact is discussed in the sequel. From (3.22),

the matrix K is clearly positive definite and has the interpretation of the cost matrix for the closed-loop system, i.e.,

$$\frac{1}{2} x'(0)Kx(0) = \frac{1}{2} \int_0^\infty x'(t) [Q + G^{d'} R G^d] \, x(t)dt \tag{3.26}$$

In contrast to the centralized case and the corresponding Riccati equation, the necessary conditions in the decentralized case may have more than one solution, such that P and K are positive definite.

The necessary conditions (3.20), (3.21), (3.25) can also be written in the transform domain since K and P are both circulant. The resulting equations are the finite-dimensional analog of the necessary condition for decentralized control of Toeplitz systems obtained by Chu [31]. Chu's results will be discussed in Section 5.1.2. The two Lyapunov equations (3.20) and (3.21) decompose in the transform domain as was shown in Chapter 2,

$$(\bar{A}_k - \bar{B}_k \bar{G}_k^d)^* \bar{K}_k + \bar{K}_k (\bar{A}_k - \bar{B}_k \bar{G}_k^d) + (\bar{Q}_k + \bar{G}_k^{d*} \bar{R}_k \bar{G}_k^d) = 0 \tag{3.27}$$

$$(\bar{A}_k - \bar{B}_k \bar{G}_k^d) \bar{P}_k + \bar{P}_k (\bar{A}_k - \bar{B}_k \bar{G}_k^d)^* + (\bar{\Sigma}_o)_k = 0 \tag{3.28}$$

for k = 0,1, ..., N-1. After some tedius algebraic manipulations, the third necessary condition (3.25) may be written as

$$\sum_{k=0}^{N-1} \bar{P}_k (\bar{G}_k^{d*} \bar{R}_k - \bar{K}_k \bar{B}_k) \, W_N^{-ik} = 0 \qquad \forall \; i \in \alpha \tag{3.29}$$

It is noteworthy that these necessary conditions remain coupled in the transform domain. This is because the constraint $G_i^d = 0 \quad \forall \; i \in \alpha$ introduces a dependency among the transformed gains $\bar{G}_k^d$. For example, if $\alpha = \{0\}$, then all the $\bar{G}_k^d$ must be equal. It is impossible, therefore, to

solve the decentralized problem by considering isolated, transformed subproblems as was the case for the centralized problem.

Determinining the optimal gain matrix G^d from the necessary conditions (3.20), (3.21), (3.25) or (3.27) - (3.29) is a formidable problem. This observation in the general (as opposed to circulant) case led Kosut [56] to consider suboptimal decentralized feedback gains. Kosut proposed two suboptimal controllers -- both defined in terms of the optimal centralized controller. His basic idea was to approximate (in some sense) the optimal gain matrix G^c by a decentralized gain matrix G^d. For example, one proposal was to choose G^d so as to minimize the sum of the squares of $G^c - G^d$. This philosophy of using the decentralized gains to approximate the centralized gains will be the basis for the suboptimal decentralized controllers considered in Section 3.2.3.

Before leaving the optimal decentralized controller, it is interesting to note that even if (A,B) is controllable and $(A,Q^{\frac{1}{2}})$ is detectable, it might happen that no stabilizing G^d exists which satisfies the constraints (3.16). This is shown by the following rather trivial example.

Example 3.1. Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and suppose $\alpha = \{1\}$. The gain matrix then is of the form $G^d = \begin{pmatrix} 0 & G_1^d \\ G_1^d & 0 \end{pmatrix}$. But for any choice of G^d, the closed-loop matrix $A - BG^d$ is unstable.

When the B matrix is diagonal, however, and (A_o, B_o) is controllable, then a stabilizing gain will exist if local feedback is allowed; i.e., $0 \in \alpha$. The restriction that B is diagonal means that each subinput $u_k(t)$ only directly influences subsystem k. If (A_o, B_o) is controllable, then it

is possible to choose G_o^d so that the measure μ [as defined by (2.71)] of $A_o - B_o G_o^d$ is as negative as desired. In particular, G_o^d can be chosen such that the sufficient stability condition (2.75) holds.

3.2.3 Suboptimal Decentralized Controller

Since obtaining the optimal decentralized gains from the associated necessary conditions is such a difficult task, it is useful to consider suboptimal decentralized gains which are much more readily computed. Kosut [56] has proposed choosing suboptimal decentralized gains on the basis of some approximation to the centralized gains. This approach will now be expanded for circulant systems by making heavy use of the diagonalizing spatial transformation defined in Chapter 2.

In order to gain insight into the approximation of the centralized controller, consider a circulant system (A,B) with scalar subsystems and centralized feedback matrix G^c. The closed-loop system matrix $(A-BG^c)$ has the transform $(\bar{A} - \bar{B}\bar{G}^c)$ that is a diagonal matrix with elements $(\bar{A}_i - \bar{B}_i\bar{G}_i^c)$. These elements are just the eigenvalues of the closed-loop system matrix. Since the $\bar{A}_i$ and $\bar{B}_i$ are fixed by the system, the closed-loop eigenvalues are located by the $\bar{G}_i^c$. The decentralized closed-loop system matrix $(A - BG^d)$ has transform $(\bar{A} - \bar{B}\bar{G}^d)$ and eigenvalues $(\bar{A}_i - \bar{B}\,\bar{G}_i^d)$. If the transformed decentralized gains $\bar{G}_i^d$ closely approximate the transformed centralized gains $\bar{G}_i^c$, then the decentralized and centralized closed-loop eigenvalues will be closely matched. The performance of the decentralized system, in this case, would be quite similar to the optimal centralized system.

It is very important to realize that the above claim that the performance of two closed-loop circulant systems will be similar if the feedback

gains are approximately equal, is not true for a general noncirculant system. Consider the following example:

Example 3.2. Let A, B, G^1 be $n \times n$ matrices

$$A = \begin{pmatrix} 1 & 1 & & & \\ & 1 & 1 & & \\ & & \ddots & \ddots & \\ & & & 1 & 1 \\ & & & & 1 \end{pmatrix}, \quad B = G^1 = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ & & & & 1 \end{pmatrix}$$

The closed-loop system matrix is

$$A-BG^1 = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{pmatrix}$$

and has all n eigenvalues equal to zero. Let G^2 be a second feedback matrix, close to G^1,

$$G^2 = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ 10^{-n} & & & 1 \end{pmatrix}$$

Example 3.2 (cont'd)

The resulting closed-loop system matrix is

$$A\text{-}BG^2 = \begin{pmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ 10^{-n} & & & 0 \end{pmatrix}$$

and has as its n eigenvalues the n roots of 10^{-n}. All these eigenvalues have magnitude 10^{-1} -- not particularly close to zero.

The point of this example is that small perturbations of the elements of an ill-conditioned matrix can result in large changes in the matrix's eigenvalues. Thus, approximating the centralized gain G^c by a decentalized gain G^d does not, in general, yield similar closed-loop systems and is not an especially good approach. For circulant systems -- at least, for circulant systems with scalar subsystems -- this difficulty cannot occur, as was discussed previously. The centralized and decentralized closed-loop poles are $(\bar{A}_i - \bar{B}_i\bar{G}_i^c)$ and $(\bar{A} - \bar{B}_i\bar{G}_i^d)$; if $\bar{G}_i^d$ is close to $\bar{G}_i^c$, then the corresponding poles are also close. In the case of vector subsystems, if the $\bar{G}_i^d$ are close to the $\bar{G}_i^c$, then the blocks $(\bar{A}_i - \bar{B}_i\bar{G}_i^d)$ on the diagonal of the transformed closed-loop system matrix will be close to the blocks $(\bar{A}_i - \bar{B}_i\bar{G}_i^c)$. Since these are low-order blocks, the centralized

closed-loop system. As is well known, the side lobes of the rectangular and decentralized closed-loop systems should be quite similar.

The decentralized control problem, therefore, is being viewed as an approximation of the set $\{\bar{G}_i^c\}$ by the transformed decentralized gains $\{\bar{G}_i^d\}$. This problem is almost identical to the design of finite impulse response (FIR) digital filters from infinite impulse response (IIR) filters [62]. In digital-signal processing, it often happens that the ideal desired frequency response of a filter has a unit sample response that is of infinite duration. Because of implementation considerations, however, it may be desired to have an FIR filter that has a frequency response closely approximating that of the ideal IIR filter. There are a variety of techniques for constructing FIR filters from IIR filters [62]. The use of some of these techniques for solving the decentralized control problem will now be considered.

The well known windowing technique is a popular and straightforward method for designing FIR filters from IIR filters. In this approach, the unit sample response $h_F(i)$ of the FIR filter is expressed as the product of the IIR filter unit sample response $h_I(i)$ times a window $w(i)$,

$$h_F(i) = h_I(i) \cdot w(i) \tag{3.30}$$

where the window is identically zero for all i outside some interval (say) $0 \leq i \leq M$. The frequency response of the resulting FIR filter is the convolution of the IIR filter frequency response and the frequency response of the window. The frequency response of the window consists of a main lobe and decreasing side lobes. When convolved with the frequency response of the IIR filter, this yields a FIR filter having a

frequency response which is a smeared version of that of the IIR filter. The Blackman, Hamming, Hanning, triangular, and rectangular windows are all described in [62]. These windows have a variety of desirable properties which makes them useful in the design of FIR filters.

Return now to the decentralized control problem, and suppose G^c is the optimal centralized feedback matrix for some quadratic cost functional. Also, suppose α is $\{0,1, \ldots, M\}$ where $M < N$. Then a suboptimal decentralized controller can be obtained from windowing by defining

$$G_i^d = w(i)G_i^c \tag{3.31}$$

where $w(i)$ is an appropriate window function. The transform $\bar{G}_i^d$ of these decentralized gains will be the convolution of the transform $\bar{w}(i)$ of the window and the transform $\bar{G}_i^c$ of the centralized gains. The transformed gain $\bar{G}_{i_o}^d$, therefore, will be a smearing of the gains $\bar{G}_i^c$ about i_o,

$$\bar{G}_i^d = \sum_{k=0}^{N-1} \bar{w}(k)\ \bar{G}_{i-k}^c \tag{3.32}$$

It is expected that $\bar{G}_{i_o}^d$ will be close to $\bar{G}_{i_o}^c$ if

(i) the transformed centralized gains do not vary too rapidly as a function of i

(ii) $\bar{w}(0)$ is near one and the other $\bar{w}(i)$ are near zero.

If the G_i^c are slowly varying, then their inverse transform, $\bar{G}_i^c$, will be concentrated around $i = 0$. This means that the windowing approach can

be expected to be most successful when the system interactions and penalty matrices are spatially localized. For example, a system characterized by nearest-neighbor interactions would be a good candidate for using windowing to obtain decentralized control gains.

If $\bar{G}_i^d = \bar{G}_i^c \; \forall \; i$, then the transformed window $\bar{w}(i)$ is $\delta_{i,o}$ and so the window itself is $w(i) = 1 \; \forall \; i$. Of course, this does not yield a decentralized controller. For any window that does yield a decentralized controller, some of the $w(i)$ will be nonzero when i is nonzero. For $i \neq 0$, a nonzero $w(i)$ is part of a side lobe of the transformed window [62]. As the height of the side lobes increases, the decentralized gains $\bar{G}_i^d$ become a more smeared version of the centralized gains $\bar{G}_i^c$. A window having smaller side lobes, therefore, is to be preferred to one having larger side lobes.

Example 3.3. This example will be used throughout the remainder of this section to illustrate the various frequency-domain techniques for obtaining a decentralized feedback gain G^d. Let

$$A = \begin{pmatrix} -4\frac{1}{2} & \frac{1}{2} & 3\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -4\frac{1}{2} & \frac{1}{2} & 3\frac{1}{2} \\ 3\frac{1}{2} & \frac{1}{2} & -4\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 3\frac{1}{2} & \frac{1}{2} & -4\frac{1}{2} \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Then, after taking the spatial transformation,

Example 3.3 (cont'd)

$$\bar{A} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -8 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -8 \end{pmatrix}, \quad \bar{B} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Open-loop poles are at 0, -8, -2, -8. Suppose the optimal closed-loop pole locations are $-\varepsilon$, -10, -10, -10 where ε is a small positive constant. The transformed centralized feedback matrix is therefore

$$\bar{G}^c = \begin{pmatrix} \varepsilon & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

and hence,

$$G^c = \begin{pmatrix} \frac{1}{4}\varepsilon+3 & \frac{1}{4}\varepsilon-2 & \frac{1}{4}\varepsilon+1 & \frac{1}{4}\varepsilon-2 \\ \frac{1}{4}\varepsilon-2 & \frac{1}{4}\varepsilon+3 & \frac{1}{4}\varepsilon-2 & \frac{1}{4}\varepsilon+1 \\ \frac{1}{4}\varepsilon+1 & \frac{1}{4}\varepsilon-2 & \frac{1}{4}\varepsilon+3 & \frac{1}{4}\varepsilon-2 \\ \frac{1}{4}\varepsilon-2 & \frac{1}{4}\varepsilon+1 & \frac{1}{4}\varepsilon-2 & \frac{1}{4}\varepsilon+3 \end{pmatrix}$$

The decentralized feedback is restricted to nearest-neighbor feedback -- i.e., $\alpha = \{0,1,3\}$ or, equivalently, $G_2^d = 0$. The first candidate decentralized controller will be obtained by windowing with a

Example 3.3 (cont'd)

rectangular window.

(I.) RECTANGULAR WINDOWING

The rectangular window $w_R(i)$ is defined by

$$w_R(i) = \begin{cases} 1 & , i = 0,1,3 \\ 0 & , i = 2 \end{cases}$$

The decentralized control gain is

$$G^R = \begin{pmatrix} \frac{1}{4}\varepsilon+3 & \frac{1}{4}\varepsilon-2 & 0 & \frac{1}{4}\varepsilon-2 \\ \frac{1}{4}\varepsilon-2 & \frac{1}{4}\varepsilon+3 & \frac{1}{4}\varepsilon-2 & 0 \\ 0 & \frac{1}{4}\varepsilon-2 & \frac{1}{4}\varepsilon+3 & \frac{1}{4}\varepsilon-2 \\ \frac{1}{4}\varepsilon-2 & 0 & \frac{1}{4}\varepsilon-2 & \frac{1}{4}\varepsilon+3 \end{pmatrix}$$

and the transformed gain is

$$\bar{G}^R = \begin{pmatrix} \frac{3}{4}\varepsilon-1 & 0 & 0 & 0 \\ 0 & \frac{1}{4}\varepsilon+3 & 0 & 0 \\ 0 & 0 & -\frac{1}{4}\varepsilon+7 & 0 \\ 0 & 0 & 0 & \frac{1}{4}\varepsilon+3 \end{pmatrix}$$

The resulting closed-loop poles are then $\left(1 - \frac{3}{4}\varepsilon\right)$, $\left(-11-\frac{1}{4}\varepsilon\right)$, $\left(-9 + \frac{1}{4}\varepsilon\right)$, $\left(-11- \frac{1}{4}\varepsilon\right)$ --- <u>unstable</u> for small ε .

In Example 3.3, using a rectangular window gives rise to an unstable

closed-loop system. As is well known, the side lobes of the rectangular window are relatively large and not at all insignificant [62]. Other windows, with a less abrupt transition to zero, have smaller side lobes.

Example 3.3 (continued)

(II) TRIANGULAR (BARTLETT) WINDOWING

The triangular window is more smoothly tapered than the rectangular window,

$$w_T(i) = \begin{cases} 1 & , i = 0 \\ \frac{1}{2} & , i = 1,3 \\ 0 & , i = 2 \end{cases}$$

This results in the decentralized gain matrix

$$G^T = \begin{pmatrix} \frac{1}{4}\varepsilon+3 & \frac{1}{8}\varepsilon-1 & 0 & \frac{1}{8}\varepsilon-1 \\ \frac{1}{8}\varepsilon-1 & \frac{1}{4}\varepsilon+3 & \frac{1}{8}\varepsilon-1 & 0 \\ 0 & \frac{1}{8}\varepsilon-1 & \frac{1}{4}\varepsilon+3 & \frac{1}{8}\varepsilon-1 \\ \frac{1}{8}\varepsilon-1 & 0 & \frac{1}{8}\varepsilon-1 & \frac{1}{4}\varepsilon+3 \end{pmatrix}$$

and transformed gain

$$\bar{G}^T = \begin{pmatrix} \frac{1}{2}\varepsilon+1 & 0 & 0 & 0 \\ 0 & \frac{1}{4}\varepsilon+3 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & \frac{1}{4}\varepsilon+3 \end{pmatrix}$$

Example 3.3 (cont'd.)

The closed-loop poles are now $(-1 - \frac{1}{2}\epsilon)$, $(-11 - \frac{1}{4}\epsilon)$, (-7), $(-11 - \frac{1}{4}\epsilon)$, -- stable.

The idea of this section is to choose the decentralized gain matrix G^d on the basis of approximating the optimal centralized gain G^c. An alternative to the windowing technique is the following least-squares type of approach. It is desired that the diagonal blocks of the transformed decentralized closed-loop system closely match the diagonal blocks of the transformed centralized system. Consider then the cost functional

$$J_{WLS} = \sum_{i=0}^{N-1} d_i \left\| (\bar{A}_i - \bar{B}_i \bar{G}_i^c) - (\bar{A}_i - \bar{B}_i \bar{G}_i^d) \right\|^2 \tag{3.33}$$

$$= \sum_{i=0}^{N-1} d_i \left\| \bar{B}_i (\bar{G}_i^d - \bar{G}_i^c) \right\|^2$$

The d_i are simply weights that are assigned to the various blocks. The dominant or critical blocks, naturally, would be assigned a larger weight. A suboptimal decentralized controller can be obtained by using the G^d which minimizes the cost functional J_{WLS}. The minimizing gain could be obtained from an iterative procedure such as Newton's method implemented on a digital computer.

The case of the weighted least-squares approach for scalar subsystems will be examined further. If the subsystems are scalars, then the functional (3.33) can be written as

$$J_{WLS} = \sum_{i=0}^{N-1} d_i \, (\bar{G}_i^d - \bar{G}_i^c)(\bar{G}_i^d - \bar{G}_i^c)^* \tag{3.34}$$

where the $\bar{B}_i$ have been incorporated into the d_i. If all the d_i in (3.34) are equal, then the result from Fourier series is that the minimizing G^d is given by

$$G_i^d = \begin{cases} G_i^c & , \; i \in \alpha \\ 0 & , \; i \in \alpha \end{cases} \tag{3.35}$$

This is nothing but using a rectangular window to obtain G^d from G^c. Kosut [56] refers to this particular decentralized gain as the "minimum norm" controller because this choice of G^c minimizes $\| G^c - G^d \|$.

Assigning all the weights d_i to be equal, however, may not result in a good closed-loop system. The reasoning here is as follows. Suppose the centralized closed-loop system has a very slow pole $(\bar{A}_o - \bar{B}_o \bar{G}_o^c) = \lambda_o$ and a very fast pole $(\bar{A}_1 - \bar{B}_1 \bar{G}_1^c) = \lambda_1$. For the decentralized closed-loop system to approximate the centralized closed-loop system, it is much more important that $(\bar{A}_o - \bar{B}_o \bar{G}_o^d)$ be close to λ_o than it is that $(\bar{A}_1 - \bar{B}_1 \bar{G}_1^d)$ be close to λ_1 -- if for no other reason than the stability of the closed-loop system. Indeed, this is exactly the situation in Example 3.3 when a rectangular window is used. The four decentralized closed-loop poles all differ from their centralized counterparts by the same amount, $1 + \frac{1}{4}\varepsilon$. For the very slow pole at $-\varepsilon$, this difference results in an unstable mode. It is clear, therefore, that the choice of identically equal d_i is not particularly wise in general. Rather, the d_i should be chosen so that the critical elements

$(\bar{A}_i - \bar{B}_i \bar{G}_i^c)$ are more closely approximated by the decentralized feedback than some other elements.

Example 3.3 (continued)

In this part of the example, the constant ε is taken to be 1 for computational simplicity.

(III.) WEIGHTED LEAST-SQUARES

Consider the assignment of weights

$$d_i = \begin{cases} 10 & , i=0 \\ 1 & , i=1,2,3 \end{cases}$$

Then the cost functional to be minimized is

$$J_{WLS} = \left\{ 10(1 - \bar{G}_o^w)^2 + (2 - \bar{G}_1^w)^2 + (8 - \bar{G}_2^w) + (2 - \bar{G}_3^w)^2 \right\}$$

where it is assumed that $G_1^w = G_3^w$. The problem then is

$$\min_{G_o^w, G_1^w} \left\{ 10(1 - G_o^w - 2G_1^w)^2 + 2(2 - G_o^w)^2 + (8 - G_o^w + 2G_1^w)^2 \right\}$$

After some tedious calculations, the minimizing gains are found to be

$$G_o^w = 2.6211$$

$$G_1^w = -0.9814$$

The closed-loop poles in this case are: (- 0.6584), (- 10.6211), (- 6.5839), (- 10.6211).

There are various other computer-based design procedures for FIR filters that could be tried for designing decentralized feedback gains. The frequency-sampling design and equripple approximation are two such procedures that will be considered here. The idea of the frequency-sampling design is this: if an FIR filter has an impulse response h(i) that is zero outside $0 \leq i \leq L-1$, then the frequency response of the filter can be exactly specified at L frequencies. The filter response at any other frequency is then fixed from these L frequency samples. In the case of the decentralized control problem, the corresponding approach is to exactly specify k transformed feedback gains (if α has k members). For scalar subsystem, this is equivalent to specifying k closed-loop poles. This specification results in a unique decentralized gain G^d, and so there are no remaining degrees of freedom left for the other N-k transformed feedback gains. There is a very real danger here that the unspecified N-k gains may result in closed-loop poles that are far from their desired locations, but the frequency-sampling technique does have the advantage of allowing the exact placement of some critical closed-loop poles.

Example 3.3 (continued)

(IV.) FREQUENCY-SAMPLING DESIGN

Closed-loop poles 0,1,3 are specified to be $-\epsilon$, -11, -11, respectively. This means that the transformed gains are

$$\bar{G}_o^{FS} = \epsilon$$

$$\bar{G}_1^{FS} = 3$$

$$\bar{G}_3^{FS} = 3$$

Example 3.3 (cont'd)

and $\bar{G}_2^{FS}$ is unspecified. Since G_2^{FS} is constrained to zero, it is possible to solve for G_o^{FS}, G_1^{FS}, G_3^{FS} from the three specified transformed gains. $\bar{G}_2^{FS}$ may then be computed. Since $G_1^{FS} = G_3^{FS}$,

$$\bar{G}_o^{FS} = \varepsilon \rightarrow \quad G_o^{FS} + 2G_1^{FS} = \varepsilon$$

$$\bar{G}_1^{FS} = 3 \rightarrow \quad G_o^{FS} = 3$$

$$\therefore \quad G_1^{FS} = \frac{3 - \varepsilon}{2} \quad .$$

Transformed gain $\bar{G}_2^{FS}$ is just

$$\bar{G}_2^{FS} = G_o^{FS} - G_i^{FS} + G_2^{FS} - G_3^{FS}$$

$$= 6 - \varepsilon \ .$$

The closed-loop poles are $(-\varepsilon)$, (-11), $(-8+\varepsilon)$, (-11) .

In the equiripple approximation of an ideal frequency response, the filter designer specifies a passband and stopband tolerance and the length of the filter impulse response [62]. An iterative scheme is then employed to find the digital filter with the narrowest transition region for the given tolerances. The design is called an "equiripple filter" because the ripples in the passband (stopband) frequency response all have the same height. An interpretation of this procedure is that rather than exactly specifying the frequency response at selected frequencies (as in the frequency-sampling approach), one specifies an acceptable interval in which the frequency response must lie. With this interpretation, the equiripple

or "interval specification" approach can be applied to the decentralized control problem. The idea is that the control designer specifies an interval for each closed-loop pole, in the case of scalar subsystems. Presumably slower or critical poles will have smaller specified intervals than other less crucial poles. In all cases, of course, to guarantee stability, the intervals should be entirely in the left-half plane.

Example 3.3 (continued)

(V.) EQUIRIPPLE DESIGN (INTERVAL SPECIFICATION)

Suppose pole 0 is required to be very near $-\varepsilon$, and the other three poles are required to be between - 9 and - 11. In fact, suppose pole 0 must be exactly at $-\varepsilon$. Then the conditions on the transformed feedback gains are

$$\bar{G}_o^E = \varepsilon$$

$$1 \leq \bar{G}_1^E \leq 3$$

$$7 \leq \bar{G}_2^E \leq 9$$

$$1 \leq \bar{G}_1^E \leq 3$$

In terms of the gains G_o^E and G_1^E (recall that $G_2^E = 0$ and $G_3^E = G_1^E$),

$$G_o^E + 2G_1^E = \varepsilon$$

$$1 \leq G_o^E \leq 3$$

$$7 \leq G_o^E - 2G_1^E \leq 9$$

Clearly, these constraints cannot all be satisfied simultaneously. This illustrates a difficulty with this method; if the specified

Example 3.3 (cont'd.)

intervals are too small, then there will not be any decentralized gain which meets the specifications. By enlarging the intervals, an acceptable gain can be found. Suppose now that poles 1,2,3 be required to be between $-8\frac{1}{2}$ and $-11\frac{1}{2}$. Then

$$\bar{G}_o^E = \varepsilon$$

$$\frac{1}{2} \leq \bar{G}_1^E \leq 3\frac{1}{2}$$

$$6\frac{1}{2} \leq \bar{G}_2^E \leq 9\frac{1}{2}$$

$$\frac{1}{2} \leq \bar{G}_1^E \leq 3\frac{1}{2}$$

Hence

$$G_o^E + 2G_1^E = \varepsilon$$

$$\frac{1}{2} \leq G_o \leq 3\frac{1}{2}$$

$$6\frac{1}{2} \leq G_o - 2G_1 \leq 9\frac{1}{2}$$

One satisfactory choice is $G_o^E = \frac{13}{4} + \varepsilon$ and $G_1^E = -\frac{13}{8}$. The resulting closed-loop poles are $(-\varepsilon)$, $(-11\frac{1}{4} + \varepsilon)$, $(-8\frac{1}{2} + \varepsilon)$, $(-11\frac{1}{4} + \varepsilon)$.

It is important to realize that there is no guarantee of stability for any of the decentralized control system designs proposed here. This was also true of the suboptimal controllers proposed by Kosut [56]. Of course, there is no guarantee that any stabilizing decentralized gains exist, as was pointed out in Section 3.2.2 when the optimal decentralized controller was discussed. The problem is that the set α may be too small (or not contain

the right members) to stabilize the system. The frequency domain approaches discussed in this subsection, however, provide some insight as to how α should be chosen so that the transformed decentralized gains $\bar{G}_i^d$ approximate the transformed centralized gains $\bar{G}_i^c$. This is an important consideration, since the control designer typically has some freedom in specifying which substates are known to each controller, i.e. where communication can take place. By using the transformed centralized gains $\bar{G}_i^c$, it is felt that one can make judgments as to which indices should be included in α.

The next section applies some of these ideas on decentralized control to an example of the circulant control of a rectangular membrane.

3.3 Computer Example: Circulant Control of a Rectangular Membrane

The physical system to be considered in this section is a vibrating membrane having fixed edges [46]. The vibrational motion is assumed to be described by the two-dimensional wave equation

$$\frac{\partial^2 v}{\partial t^2} - c^2 \frac{\partial^2 v}{\partial y^2} - b^2 \frac{\partial^2 v}{\partial z^2} = E(y,z,t) \qquad (3.36)$$

where $0 < y < 1$, $0 < z < 1$, and $E(y,z,t)$ is an external force applied to the membrane. The approach taken here will be to use difference quotients to approximate the two spatial partial derivatives in (3.36). An alternate approach would be to use modal analysis to design a controller. In [63], Creedon and Lindgren use modal techniques for the control of the vibrational motion of a thin rectangular plate (as opposed to a membrane). This work, as applied to the control of a deformable mirror, is successful in obtaining a centralized control. In the present section, decentralized as well as centralized laws are obtained.

The wave equation (3.36) is discretized at N points in the y direction and M points in the z direction. Letting N=9 and M=14, the displacement $d_{i,j}(t)$ is simply

$$d_{i,j}(t) = v\left(\frac{i}{10}, \frac{j}{15}, t\right) \tag{3.37}$$

for i=1, ..., 9 and j = 1, ..., 14. For control purposes, it is assumed that there is an ideal actuator located at each of the points on the discretized rectangular grid. If $u_{i,j}(t)$ is the applied force at the point $y = \frac{i}{10}$ and $j = \frac{j}{15}$, then the finite difference approximation of (3.36) is (taking c=b=0.1),

$$\ddot{d}_{i,j}(t) = 1.00\left[d_{i+1,j}(t) - 2d_{i,j}(t) + d_{i-1,j}(t)\right] + \tag{3.38}$$

$$+ 2.25\left[d_{i,j+1}(t) - 2d_{i,j}(t) + d_{i,j-1}(t)\right] + u_{i,j}(t)$$

By letting $x_{i,j}(t) = \begin{pmatrix} d_{i,j}(t) \\ \dot{d}_{i,j}(t) \end{pmatrix}$ and lexicographically ordering the points yields the lumped-parameter model (2.95) where now

$$F_0 = \begin{pmatrix} 0 & 1 \\ -3.25 & 0 \end{pmatrix} \tag{3.39A}$$

$$F_1 = \begin{pmatrix} 0 & 0 \\ 2.25 & 0 \end{pmatrix} \tag{3.39B}$$

$$F_2 = \begin{pmatrix} 0 & 0 \\ 1.00 & 0 \end{pmatrix} \tag{3.39C}$$

As discussed in Section 2.2, this system can be imbedded in a circulant

system of order $8(N+1)(M+1) = 1200$. This circulant system is composed of circulant blocks and is the basis of the control laws now obtained.

Before addressing the control problem, however, it is somewhat striking to note that when a tridiagonal system is imbedded in a circulant system (see Section 2.2), the circulant system can be unstable, even if the original tridiagonal system is stable. For the circulant system being considered here, the system matrix can be put in block diagonal form with 2×2 blocks by taking two spatial transforms. If this is done, the initial (upper lefthand corner) block is easily found to be

$$\begin{pmatrix} 0 & 1 \\ +3.25 & 0 \end{pmatrix}$$

obviously unstable.

A centralized linear-quadratic regulator will now be obtained for this (doubly) circulant system. A steady-state regulator is desired, and the weighting matrices Q and R are both chosen equal to the identity, for simplicity. Employing two spatial transforms, the (1200×1200) Riccati equation is decomposed into (2×2) complex-valued Riccati equations, as shown in (2.110). Solving these low-order Riccati equations and taking two inverse spatial transformations yields the solution of the original equation.

All the centralized closed-loop poles, of course, are in the left-half-plane. The real parts of these eigenvalues are all between -2.3 and -0.5. Almost all of the poles with real parts less than -1 are purely real. Conversely, essentially all the poles with real parts greater than

-1 are underdamped, having nonzero imaginary components. The poles that have real parts around -0.5 are the most underdamped — some of the imaginary parts of these poles are as large as 3.1.

Two decentralized controllers were also computed for this system. Both controllers were obtained from the centralized solution by using the windowing technique as described in Section 3.2.3. A square and a triangular window were both used, and the allowable feedback was taken to be local feedback and feedback from the two nearest neighbors on both sides and in both directions. Also, the optimal centralized local feedback gains on position and velocity were 0.870 and 1.498, respectively. For the first nearest neighbors in the y direction, the two gains were 0.659 and 0.339, and for the second neighbors, 0.369 and 0.165. Also, in the z direction, the gains on the first nearest neighbor were 0.453 and 0.210, and 0.363 and 0.156 for the second-nearest neighbor.

The square-windowed decentralized controller uses these gains directly; all other feedback gains are identically zero. This results in a stable closed-loop system with the closed-loop eigenvalues having real parts from -1.67 to -0.317. The decentralized system, therefore, is slower than the optimal centralized one. Also, the poles with the largest real parts (those around -0.32) are quite underdamped, having imaginary components as large as 5.2. Also, in this decentralized case, purely real poles are found to the right of the -1 point, again in contrast to the centralized case. The slowest real pole, in fact, is equal to -0.472.

A second decentralized controller is obtained by using a triangular window. This decentralized controller is obtained by multiplying the

local gains by 1, the nearest-neighbor gains by 2/3, and the second nearest-neighbor gains by 1/3. Once again, the resulting closed-loop system is found to be stable. The real parts of these poles lie between -1.71 and -0.086. The relatively slow pole at -0.08, in fact, is purely real. Most of these closed-loop poles have nonzero complex parts, and the complex parts range up to 5.6.

In comparing these three control laws, the centralized controller produces a closed-loop system that is faster than either of the two decentralized closed-loop systems. The performance of the square-windowed decentralized controller, however, is quite comparable to that of the centralized controller and could well be acceptable. The triangular-windowed decentralized controller, on the other hand, is substantially slower than the other two. This might be compensated for by multiplying (say) 3/2, 1, 1/2 instead of 1, 2/3, 1/3 when computing the windowed gains. What this does is preserve the triangular shape of the window while increasing the level of local feedback — thereby, it is hoped, increasing the speed of response of the closed-loop system.

It is interesting to conjecture why the centralized controller should be quicker to respond that the decentralized controllers are. The centralized controller at any subsystem is aware of disturbances which are propagating toward the subsystem. It can, therefore, apply the appropriate control to compensate for this disturbance before the disturbance significantly affects the local subsystem. Decentralized controllers, however, are limited to only much more localized information, and hence may not take any appropriate control action until the disturbance is upon the local subsystem.

In conclusion, it is noted that the assumption that control actuators were present at every subsystem is not at all necessary. What this assumption allows, however, is consideration of only second-order subsystems. If one desired to employ a finer spatial discretization, for example, the resulting system could still be imbedded in a circulant. The dimension of the subsystems, it should be realized, will have to increase from two.

3.4 Circulant Control of Large-Scale Systems

The linear-quadratic control problem for circulant systems was shown in Section 3.1 to have efficient solutions both on-line and off-line. The off-line savings were due to the fact that the problem decomposes into smaller, independent subproblems. The on-line savings were a result of using the FFT to implement the multiplication of a circulant matrix times a vector. The purpose of this section is to propose using a circulant controller for a general large-scale system, thereby realizing at least some of the computational benefits just mentioned.

For the large-scale system

$$\frac{d}{dt}\, x(t) = A_L x(t) + B_L u(t) \tag{3.40}$$

the control problem of interest is to determine the input $u(t)$ so as to minimize the quadratic cost functional

$$J_L = x'(T)F_L x(T) + \int_0^T x'(t)Q_L(t)x(t) + u'(t)R_L(t)u(t)dt \tag{3.41}$$

where $F_L \geq 0$, $Q_L(t) \geq 0$, and $R_L(t) > 0$. McClamroch [64] has adduced the question of what happens if, instead of $A_L, B_L, F_L, Q_L(t)$ and $R_L(t)$, the matrices A_a, B_a, F_a, $Q_a(t)$ and $R_a(t)$, respectively, are used to compute a control law for the system (3.40). Let J_a denote the optimal cost of the approximate system (A_a, B_a) with weighting matrices F_a, $Q_a(t)$, and $R_a(t)$, and let $K_a(t)$ be the solution of the Riccati equation

$$\frac{d}{dt}K_a(t) = -K_a(t)A_a - A_a'K_a(t) - Q_a(t) + K_a(t)B_aR_a^{-1}(t)B_a'K_a(t) \qquad (3.42)$$

$$K_a(T) = F_a$$

The control $u(t) = -R_a^{-1}(t)B_a' K_a(t)x(t)$ is then applied to the original large-scale system (3.40). McClamroch is now able to bound the cost J_L when this suboptimal control is applied. For any real number ρ satisfying

$$(\rho-1)[K_a(t)B_aR_a^{-1}(t)B_a'K_a(t) + Q_a(t)] +$$

$$-\rho\, K_a(t)[(A_L - A_a) - (B_L-B_a)R_a^{-1}(t)B_a' K_a(t)] +$$

$$-[(A_L-A_a) - (B_L-B_a)R_a^{-1}(t)B_a' K_a(t)]'K_a(t)\,\rho +$$

$$-K_a(t)B_aR_a^{-1}(t)[R_L(t)-R_a(t)]R_a^{-1}(t)B_a' K_a(t) + (Q_a-Q_L) \geq 0 \qquad (3.43)$$

$$\forall \quad 0 \leq t \leq T$$

and

$$(\rho-1)F_a - (F - F_a) \geq 0 \qquad (3.44)$$

the following condition holds for all $x(0)$:

$$J_L \leq \rho\, J_a \qquad (3.45)$$

Thus, the cost when using a control designed for the approximate system on the original system is bounded by ρ times the cost for the approximate system.

Bailey and Ramapriyan [20] simplify the imposing conditions (3.43) and (3.44) for the infinite-horizon problem. They also impose the additional restrictions that $B_a = B_L$, $Q_a = Q_L$, and $R_a = R_L$. Bailey

and Ramapriyan study the dual problem also, and thereby obtain a lower, as well as an upper, bound for J_L. After some manipulation, condition (3.43) reduces to finding the two smallest values of such that

$$(1-\frac{1}{\rho}) \; [K_a B_L R_L^{-1} B_L' K_a + Q_L] \pm [K_a(A_L - A_a) + (A_L - A_a)' K_a] \geq 0 \qquad (3.46)$$

Bailey and Ramapriyan show that these two values of ρ are given by the maximum (σ_M) and the minimum (σ_m) eigenvalues of

$$\left| [K_a B_L R_L^{-1} B_L' K_a + Q_L]^{-1} [K_a(A_L - A_a) + (A_L - A_a)' K_a] \right| \qquad (3.47)$$

The conclusion is that if $-1 < \sigma_m \leq \sigma_M < 1$, then the feedback designed from the approximate system is <u>stabilizing</u> and the cost satisfies

$$(1+\sigma_m) J_a \leq J_L^0 \leq J_L \leq \left(\frac{1}{1-\sigma_m}\right) J_a \qquad (3.48)$$

where J_L^0 denotes the optimal cost for the original problem.

One might very well ask at this point what the above machinations have done to reduce the computational load. The important observation is that if the approximating matrix A_a is purposefully chosen to be circulant (and the other matrices B_L, Q_L, R_L are also circulant), then (3.48) can provide quite a useful result. First, the Riccati equation (3.42) giving the K_a may be decomposed as shown in Chapter 2. Second, the control gain ($-R_L^{-1} B_L' K_a$) is circulant, and therefore can be efficiently implemented using the FFT. The remaining problem of computing the eigenvalues of (3.47), however, is still a difficult task. It should be

noted that rather than inverting the matrix $[K_a B_L R_L^{-1} B_L' K_a + Q_L]$ multiplying, and then solving for the eigenvalues, one can instead solve the so-called "generalized eigenvalue problem", i.e., in the present case, find scalars λ and x such that

$$\lambda[K_a B_L R_L^{-1} B_L' K_a + Q_L]x = [K_a(A_L - A_a) + (A_L - A_a)' K_a]x \qquad (3.49)$$

the matrix on the LHS of (3.49) is a circulant, and this may present some computational advantages. Finally, to ensure stability, all that is needed is that the eigenvalues σ_m and σ_M have magnitude less than one.

Developing bounds on the performance index that are simpler than either (3.43) and (3.44) or (3.47) is an area where fruitful research could be conducted. Another issue concerns the approximating circulant system matrix A_a. The above analysis assumed that A_a was given. How this matrix should be chosen is a crucial question and warrants further investigation.

One class of systems for which it seems particularly easy to obtain a good circulant approximation is a finite string of subsystems having a Toeplitz system matrix. The idea here is simply to tie together the two ends of the system, thereby turning the Toeplitz system matrix into a circulant one. The tridiagonal systems examined in Section 2.2 fall into this class. These tridiagonal systems could be imbedded in higher-order circulant systems, so the utility of a circulant approximation is questionable. But, in the general case of a finite string described by a Toeplitz matrix, the circulant approximation could be very useful. The question of how well the circulant approximation works for designing feedback gains is still open.

3.5 The Dual Filtering Problem

All the results of this chapter on the linear-quadratic control problem for circulant systems are also applicable to the dual filtering problem. The stochastic circulant system under consideration is

$$\frac{d}{dt} x(t) = A'x(t) + w(t) \tag{3.50}$$

$$y(t) = B' x(t) + v(t) \tag{3.51}$$

where $x(0)$, $w(t)$ and $v)t)$ are all independent, zero-mean, Gaussian random variables, and $E\, x(0)x'(0) = \Sigma(0)$, $E\, w(t)w'(\tau) = Q(t)\delta(t-\tau)$, and $E\, v(t)c'(\tau) = R(t)\delta(t-\tau)$. The matrices $A,B,\Sigma(0)$, $R(t)$, $Q(t)$ are all assumed to be block circulant. The argument used in Section 3.1 to justify the use of only block circulant weighting matrices can be applied here to show why it is physically reasonable to demand that covariance matrices are all block circulant. The filtering problem is to compute the Bayesian estimate $\hat{x}(t|t)$ of $x(t)$ given the observations $y(\tau)$ from time zero to t, i.e.,

$$\hat{x}(t|t) = E[x(t)|\ y(\tau),\ \ 0 \leq \tau \leq t] \tag{3.52}$$

The solution to this problem is, of course, the Kalman filter

$$\frac{d}{dt} \hat{x}(t|t) = A'\hat{x}(t|t) + P(t)BR^{-1}(t)[y(t)-B'\hat{x}(t|t)] \tag{3.53}$$

where the error covariance $P(t)$ is given by the Riccati equation

$$\frac{d}{dt} P(t) = A'P(t) + P(t)A + Q(t) - P(t)BR^{-1}(t)\ B'P(t) \tag{3.54}$$

subject to $P(0) = \Sigma(0)$. Comparing (3.4) and (3.54), it is obvious that if $\Sigma(0) = F$, then the error covariance $P(t)$ equals the cost-to-go $K(T-t)$ for all $t \in [0,T]$. Thus, the filter gain matrix $P(t)B'R^{-1}(t)$ is just the transpose of the feedback matrix $G^c(T-t)$. It is clear that this filtering problem is the dual of the optimal control problem addressed in Section 3.1.

The circulant filtering problem in the transform domain, then, consists of N independent subproblems. The required off-line computation is greatly reduced; only the lower-order Riccati equations (3.6) need to be solved. The FFT can be used to reduce the on-line computations just as in the case of the centralized optimal controller. That is, a structure analogous to that of Figure 3.1 can be used to implement the update step of the discrete-time Kalman filter. The point here is that the circulant filter enjoys all of the computational advantages of the circulant regulator.

A time-invariant filter can be obtained if the covariances $Q(t)$ and $R(t)$ are constant and observations are available from the infinite past. The filter gain is PBR^{-1} where P satisfies the algebraic Riccati equation (3.14). It is to be noted that the resulting estimate is the minimum variance estimate of $x(t)$, i.e., the gain PBR^{-1} minimizes the error covariance over the set of all filter gains. This observation is the key to understanding the filtering dual of the decentralized controllers discussed in Section 3.2.

The dual of the decentralized control problem will now be considered. The dual filter produces an estimate $\hat{x}(t)$ according to

$$\frac{d}{dt}\hat{x}(t) = A'\hat{x}(t) + G^f[y(t)-\hat{x}(t)] \tag{3.55}$$

where the filter gain G^f is block circulant and subject to the constraints

$$G_i^f = 0 \quad \forall \quad i \in \alpha \tag{3.56}$$

This filter has a fixed structure, and the problem is to choose G^f to minimize the resulting error covariance P. The error covariance can be computed as a function of the gain G^f from the Lyapunov equation

$$(A - BG^{f'})'P + P(A-BG^f) + (Q + G^f R G^{f'}) = 0 \tag{3.57}$$

provided $(A - BG^{f'})$ is a stability matrix. The optimal G^f minimizes

$$\frac{1}{2} \text{ trace } [P] = \frac{1}{2} \text{ trace } [P \cdot I] \tag{3.58}$$

where I is the identity. Comparing the cost functional (3.58) with (3.19B) and the Lyapunov equations (3.57) and (3.21), it is obvious that the problem of computing the optimal fixed-structure filter gain G^f is equivalent to the decentralized control problem of Section 3.2.2. The necessary identifications are $G^f = G^{d'}$ and $\Sigma_o = I$. The equivalence is completed by defining K from

$$(A - BG^{f'})K + K(A - BG^{f'})' + I = 0 \tag{3.59}$$

and then associating this K and the error covariance P with the matrices P and K, respectively, in the decentralized control problem, (3.20) and (3.21). Therefore, from (3.25), the optimal filter gain G^f satisfies

$$\sum_{k=0}^{N-1} K_{-i-k} \sum_{\ell=0}^{N-1} \left[G^f_{-\ell} R_{k-\ell} - P_\ell B_{k-\ell} \right] = 0 \qquad \forall \; i \in \alpha \qquad (3.60)$$

where K and P are given by (3.57) and (3.59). Alternatively, these necessary conditions can be written in the transform domain as (3.27) - (3.29).

All the comments in Section 3.2.2 concerning existence and uniqueness of solutions of the necessary conditions are relevant here. Obtaining the optimal fixed-structure filter gain G^f, therefore, is quite difficult. A suboptimal gain, however, can be found using the frequency-domain techniques presented in Section 3.2.3. These frequency-domain techniques attempt to approximate the spatial transform of the optimal centralized filter gain by a suboptimal filter gain that satisfies the constraints (3.56).

It is not entirely clear that the fixed-structure filter considered here should be referred to as a decentralized filter. For the dual optimal control problem, this interpretation was very natural, since each subinput could be computed by a local controller. The fixed-structure filter (3.55), on the other hand, appears centralized. Consider the discrete-time version of this filter. Then the propagation step is centralized; in fact, the propagation step is exactly the same as for the optimal centralized filter. For a general system matrix, the propagated estimate of each substate involves the estimates of <u>all</u> the other substates. The measurement update step, however, is decentralized. Consequently, local filters can be used to compute the updated estimate at

subsystem k from the observations $y_{(i+k) \bmod N}$, $i \in \alpha$. Only in that the measurement update step is decentralized may the fixed-structure filter be said to be a decentralized filter. However, if the system matrix A is sparse, then the propagation step involves limited inter-subsystem communication. For example, a tridiagonal system matrix means that the propagation step can be implemented by simply having each subsystem transmit its own substate estimate to its two nearest neighbors.

3.6 Summary and Discussion

The main theme of this chapter has been the use of the spatial frequency domain for design of controllers and estimators for circulant systems. Both the centralized and decentralized cases were considered. In the centralized case, the spatial transform decomposed the original problem into a number of low-order problems of the same type. Also, the centralized solution was shown to have an efficient on-line implementation employing the FFT. Optimal and suboptimal decentralized regulators were covered in Section 3.2. The most interesting results here are in Section 3.2.3, where suboptimal decentralized regulators are designed from the optimal centralized regulator. The key step in this development is the use of the spatial frequency domain to draw the analogy between the problem at hand and the design of finite-impulse response digital filters from infinite-impulse response filters. Section 3.3 provides an example of centralized and decentralized circulant controllers for a rectangular membrane having fixed edges. The dual filtering problem is covered in Section 3.5.

Section 3.4 represents some first steps in attempting to obtain the off-line or on-line advantages of a circulant control system for a general large-scale system. A fundamental problem here is the choice of an appropriate circulant system matrix A_a to approximate the large-scale system matrix A. An immediate candidate for the circulant approximation A_a is a matrix consisting of the average subsystem interactions, i.e., the i^{th} block of A_a is given by

$$(A_a)_i = \frac{1}{N} \sum_{k=0}^{N-1} A_{i+k,k} \tag{3.61}$$

where the indices are modulo N. Intuitively, if $A_{i+k,k}$ is close to $(A_a)_i$ for each k, then this approach would work well. Therefore, perhaps a functional such as the variance

$$V_i = \frac{1}{N} \sum_{k=0}^{N-1} \| A_{i+k,k} - (A_a)_i \| \tag{3.62}$$

will be important in determining when the circulant control law is stabilizing, or in bounding the suboptimal cost. These suggestions are just speculative, and much work remains to be done before one can say whether or not the average interactions (3.61) and the variance (3.62) are useful quantities.

CHAPTER 4

THE FIXED-INTERVAL SMOOTHING PROBLEM

4.1 Introduction

The fixed-interval (FI) smoothing problem is of particular interest in post-experimental data analysis and has received considerable attention [65], [66]. This study of the smoothing problem, however, has quite a different motivation; namely, some recent results of Attasi [28] on discrete-time filtering for Toeplitz systems (defined in Chapter 5). Actually Attasi's work is concerned with recursive estimation for images, but it is shown in Chapter 5 how his results apply to Toeplitz systems. Attasi has shown that the update cycle of the discrete-time Kalman filter for some Toeplitz systems is equivalent to a smoothing problem. Thus the update operation may be implemented by a smoother -- in particular, the Mayne-Fraser two-filter smoother [41], [42]. The implications of this result for filtering of large-scale systems are examined in detail in Chapter 5. The purpose of this chapter is to carefully study the two-filter smoother.

Smoothing refers to estimating a state vector at a time point intermediate to a span of measurements. Consequently, there is an essential element of noncausality in smoothing since some of the measurements occur in the future. Fixed interval smoothing involves measurements over a given, fixed time interval. Estimates of the state are desired throughout this time interval.

Consider the continuous-time linear dynamic system

$$\frac{d}{dt} x(t) = A(t)x(t) + w(t) \tag{4.1}$$

with observations

$$y(t) = C(t)x(t) + v(t) \tag{4.2}$$

where $x(o) = x_o$ and

$$E\, x_o = E\, w(t) = E\, v(t) = 0,\ E\, x_o x_o' = \Sigma(o),$$

$$E\, w(t)w'(\tau) = Q(t)\ \delta(t-\tau),\ E\, v(t)v'(\tau) = R(t)\ \delta(t-\tau),$$

$$E\, x_o w'(t) = E\, x_o v'(t) = E\, w(t)v'(\tau) = 0$$

Also it is assumed that all the random variables x_o, $w(t)$, $v(t)$ are Gaussian. The FI smoothing problem is to compute, for all $t \in [0,T]$, the conditional expectation of $x(t)$ given the observations over $[0,T]$, i.e. the smoothed estimate is

$$\hat{x}_s(t) = E\{x(t) \,|\, y(\tau),\ 0 \leq \tau \leq T\} \tag{4.3}$$

and the corresponding smoothed error covariance is

$$P_s(t) = E\left\{\left[x(t) - \hat{x}_s(t)\right]\left[x(t) - \hat{x}_s(t)\right]'\right\} \tag{4.4}$$

As is well-known, the estimate $\hat{x}_s(t)$ is a linear functional of the observations and is also the maximum *a posteriori* estimate and the linear least-squares estimate.

In order to gain some insight into the smoothing problem, consider the time-invariant version of the smoothing problem with observations over the interval $(-\infty, +\infty)$. In the sequel, this will be referred to as the time-invariant infinite-lag smoothing problem. Also, it is assumed here that x and y are scalar random processes and

$$y(t) = x(t) + v(t) \tag{4.5}$$

The Wiener filter will be used to examine the relationship between past and future observations in estimating x(t).

The Wiener filter provides an estimate of x(t) given $y(\tau)$, $\tau \leq t$, in terms of a convolution integral as

$$\hat{x}(t|t) = \int_{-\infty}^{t} h(t-\tau)y(\tau)d\tau \tag{4.6}$$

where the filter impluse response obeys the Wiener-Hopf Equation,

$$R_{xy}(t-\sigma) = \int_{-\infty}^{t} h(t-\tau)R_{y}(\tau-\sigma)d\tau \ , \ \sigma \leq t \tag{4.7}$$

with $R_{xy}(t) = E[x(t)y(o)]$ and $R_{y}(t) = E[y(t)y(o)]$. A similar anti-causal expression can be obtained to provide an estimate $\hat{x}_{r}(t|t)$ of x(t) from future observations,

$$\hat{x}_{r}(t|t) = \int_{t}^{\infty} h(t-\tau)y(\tau)d\tau \tag{4.8}$$

where the reversed-time filter obeys the Wiener-Hopf equation

$$R_{xy}(t-\sigma) = \int_{t}^{\infty} h_{r}(t-\tau)R_{y}(\tau-\sigma)d\tau \ , \ \sigma \geq t \tag{4.9}$$

It is easily shown that the cross-correlation function between x(t) and y(t) is an even function of time,

$$R_{xy}(t) = E[x(t)y(o)] \tag{4.10}$$

$$= E[x(t)x(o) + x(t)v(o)]$$

$$= E[x(t)x(o)]$$

$$= E[x(-t)x(o)]$$

$$= R_{xy}(-t)$$

This fact can now be used to relate the filter impulse responses h(t) and $h_r(t)$,

$$\int_{-\infty}^{t} h(t-\tau)R_y(\tau-\sigma)d\tau = R_{xy}(t-\sigma) \quad , \sigma \leqslant t \tag{4.11}$$

$$= R_{xy}(t-s) \quad , s-t = t-\sigma$$

$$= \int_{t}^{\infty} h_r(t-\tau)R_y(\tau-s)d\tau \quad , \text{from (4.9)}$$

$$= \int_{-\infty}^{-t} h_r(t+\mu)R_y(\mu+s)d\mu \quad , \mu = -\tau$$

$$= \int_{-\infty}^{t} h_r(-t+\tau)R_y(s-2t+\tau)d\tau \quad , \tau = 2t+\mu$$

$$= \int_{-\infty}^{t} h_r(-t+\tau)R_y(\tau-\sigma)d\tau$$

$$\Rightarrow h(t-\tau) = h_r(\tau-t) \tag{4.12}$$

Equation (4.12) says that the weight $h(-\tau)$ given to $y(\tau)$ in forming $\hat{x}(o|o)$ is the same as the weight $h_r(\tau)$ given to $y(-\tau)$ in forming $\hat{x}_r(o|o)$. In this sense, the same linear filter is used to estimate $x(t)$ from either the past or the future observations. The only place that the assumption $x(t)$ and $y(t)$ are scalar processes is used is for the relation $R_x(t)=R_x(-t)$. Whenever $R_x(t) = R_x'(t)$, it follows that $R_x(t)$ is an even function of time, and so the same proof will work for vector processes in this case. The vector case is addressed more completely in Section 4.3.5. As far as their relative performance, it is straightforward to show that both estimates $\hat{x}(t|t)$ and $\hat{x}_r(t|t)$ have the same mean-square error. Therefore the past and future contain equal amounts of information about $x(t)$, and one would expect equal weightings on both when forming the smoothed estimate.

That this is exactly the case can be seen from the Wiener smoother. The smoothed estimate is

$$\hat{x}_s(t) = \int_{-\infty}^{+\infty} h_s(t-\tau)y(\tau)d\tau \qquad (4.13)$$

where

$$\mathcal{F}\{h(t)\} = \frac{\mathcal{F}\{R_{xy}(t)\}}{\mathcal{F}\{R_y(t)\}}$$

and the operator $\mathcal{F}\{\cdot\}$ is the Fourier transform. Since $R_{xy}(t)$ and $R_y(t)$ are even functions of t, their transforms will be purely real. Thus the transform of $h(t)$ will also be real, and so $h(t)$ must be even. This proves that the past and future contribute equally to the smoothed estimate

of $x(t)$.

In Section 4.2, the Mayne-Fraser two-filter smoother will be presented. Section 4.2 provides an historical review of the two-filter smoother discussing the work of Mayne, Fraser, and Mehra. The two-filter smoother gives the smoothed estimate as a combination of a forward and a backward estimate. Both estimates come from Kalman filters. A surprising fact, however, is that in the infinite-lag case when the state dimension equals one, the steady-state covariance of the backward is always larger than the covariance of the forward filter. (See (4.40) for the backwards covariance.) This seems to contradict the previous development where the past and future observations were seen to be equally valuable in estimating $x(t)$.

The reason for this apparent contradiction is that the Mayne-Fraser two-filter smoother has a built-in asymmetry that is absent from the original problem. In Section 4.3, it is shown that this asymmetry is due to the way in which the *a priori* information enters into the estimate of $x(t)$. New forms of the two-filter smoother will be presented which are symmetric with respect to forward- and reversed-time. It is to be hoped that the analysis and discussion of Section 4.3 will provide a clear understanding of how future observations are used in the FI smoothing problem.

Section 4.4 uses one of the new forms of the two-filter smoother to analyze reduced-order smoothers and to perform a sensitivity analysis.

This introductory section closes with two well-known results from probability theory. The first deals with combining estimates that have independent errors.

Proposition 4.1. Let x, y_1 and y_2 be zero-mean Gaussian random variables, and let $\hat{x}_1$ and $\hat{x}_2$ be the Bayesian (maximum likelihood) estimates of x given y_1 and y_2, respectively, with associated error covariances P_1 and P_2. If the errors $x-\hat{x}_1$ and $x-\hat{x}_2$ are independent, then the Bayesian (maximum likelihood) estimate of x given both y_1 and y_2 is

$$\hat{x} = P[P_1^{-1}\hat{x}_1 + P_2^{-1}\hat{x}_2] \tag{4.14}$$

where the error covariance P is given by

$$P = [P_1^{-1} + P_2^{-1}]^{-1} \tag{4.15}$$

proof: see Schweppe [67].

By abuse of terminology, the estimates $\hat{x}_1$ and $\hat{x}_2$ are often referred to as independent estimates. The second result is simply the formula for the conditional expectation of a Gaussian random variable.

Proposition 4.2. Let x be a Gaussian random variable with mean m and covariance Σ, and let y be an observation of x,

$$y = Hx + v \tag{4.16}$$

where v is a zero-mean Gaussian random variable with covariance R that is

independent of x. The Bayesian estimate of x given y is

$$\hat{x} = E[x|y] \tag{4.17}$$

$$= P[(H'R^{-1})y + \Sigma^{-1}m]$$

where the error covariance P satisfies

$$P = [(H'R^{-1}H)^{-1} + \Sigma^{-1}]^{-1} \tag{4.18}$$

and it is assumed that the rank of H equals the dimension of x.

proof: see Schweppe [67].

It is noted that Proposition 4.2 can be interpreted as meaning that the Bayesian estimate of x given y equals the maximum likelihood estimate of x given both y and the *a priori* mean m and covariance Σ. This interpretation is obtained by forming two independent maximum likelihood estimates of x, one based on y and one based on m and Σ. Combining these two maximum likelihood estimates by Proposition 4.1 yields (4.17) and (4.18).

4.2 Historical Review of the Two-Filter Smoother

The first person to express the solution of the FI smoothing problem as a combination of two estimates was David Mayne [41] in 1966. The system under consideration is a discrete-time analog of (4.1) and (4.2),

$$x(k+1) = \Phi(k+1,k)x(k) + w(k) \tag{4.19}$$

$$y(k) = C(k)x(k) + v(k) \tag{4.20}$$

where $x(o) = x_o$ and

$$E\, x_o = E\, w(k) = E\, v(k) = 0 \quad , \quad E\, x_o x_o' = \Sigma_o$$

$$E\, w(i)w'(k) = Q(k)\delta_{i,k} \quad , \quad E\, v(i)v'(k) = R(k)\delta_{i,k}$$

$$E\, x_o w'(k) = E\, x_o v'(k) = E\, w(i)v'(k) = 0$$

Mayne's starting point was the conditional probability density of the states given the observations,

$$p(x(o),\ldots,x(T)\,|\,y(o),\ldots,y(T)) = \frac{p(y(o),\ldots,y(T)\,|\,x(o),\ldots,x(T))\,p(x(o),\ldots,x(T))}{p(y(o),\ldots,y(T))} \tag{4.21}$$

Because of the independence of the observation noise process $\{v(k)\}$, the likelihood function $p(y(o),\ldots,y(T)\,|\,x(o),\ldots,x(T))$ may be written as

$$p(y(o),\ldots,y(T)\,|\,x(o),\ldots,x(T)) = \prod_{k=o}^{T} p(y(k)\,|\,x(k))$$

$$= K_1 \prod_{k=o}^{T} \exp\left\{-\frac{1}{2}\,||\,y(k) - C(k)x(k)\,||^2_{R^{-1}(k)}\right\} \tag{4.22}$$

Also, since the sequence $\{x(k)\}$ is a Markov process,

$$p(x(o),\ldots,x(T)) = p(x(o))\prod_{k=o}^{T-1} p(x(k+1)\,|\,x(k))$$

$$= K_2 \exp\left\{-\frac{1}{2}||\,x(o)\,||^2_{\Sigma_o^{-1}}\right\} \prod_{k=o}^{T-1} \exp\left\{-\frac{1}{2}||\,x(k+1) - \Phi(k+1,k)\,||^2_{Q^{-1}(k)}\right\} \tag{4.23}$$

Therefore, substituting into (4.21) and realizing that $p(y(o),\ldots,y(T))$ is just a normalization constant yields

$$p(x(o),\ldots,x(T)|y(o),\ldots,y(T)) = K_3 \exp\Big\{-\frac{1}{2}||x(o)||^2_{\Sigma_o^{-1}} + \\ -\frac{1}{2}\sum_{k=o}^{T}||y(k) - C(k)x(k)||^2_{R^{-1}(k)} + \\ -\frac{1}{2}\sum_{k=o}^{T-1}||x(k+1) - \Phi(k+1,k)x(k)||^2_{Q^{-1}(k)} \tag{4.24}$$

where K_1, K_2, and K_3 are constants. The optimal smoothed estimates $\{\hat{x}(k|T)\}$ maximize the conditional density given in (4.24).

Consider now the negative of the exponent in the right-hand-side of (4.24),

$$J(x(o),\ldots,x(T)) = \frac{1}{2}||x(o)||^2_{\Sigma_o^{-1}} + \frac{1}{2}\sum_{k=o}^{T}||y(k) - C(k)x(k)||^2_{R^{-1}(k)} + \\ + \frac{1}{2}\sum_{k=o}^{T-1}||x(k+1) - \Phi(k+1,k)x(k)||^2_{Q^{-1}(k)} \tag{4.25}$$

The smoothed estimates can be obtained from the <u>minimization</u> of the functional J. Mayne's approach to this minimization was to consider some fixed r between 0 and T and to define

$$J_r(x(r)) = \min_{\{x(k)|k\neq r\}} J(x(o),\ldots,x(T)) \tag{4.26}$$

$J_r(x(r))$ will be a quadiatic form in $x(r)$, and therefore given $J_r(x(r))$ it is an easy matter to compute $\hat{x}(r|T)$ and $P(r|T)$. Hence the problem of

interest is the determination of an expression for $J_r(x(r))$.

Mayne decomposes the minimization over $\{x(k) | k \neq r\}$ into two separate minimizations -- one over $\{x(o),\ldots,x(r-1)\}$ and the other over $\{x(r+1),\ldots,x(T)\}$. Thus

$$J_r(x(r)) = J_{o,r}(x(r)) + J_{r,T}(x(r)) \tag{4.27}$$

where

$$J_{o,r}(x(r)) = \min_{\{x(o),\ldots,x(r-1)\}} \left\{ \frac{1}{2} ||x(o)||^2_{\Sigma_o^{-1}} + \right.$$

$$\left. + \frac{1}{2}\sum_{k=o}^{r-1} ||y(k) - C(k)x(k)||^2_{R^{-1}(k)} + \frac{1}{2} \sum_{k=o}^{r-1} ||x(k+1) - \Phi(k+1,k)x(k)||^2_{Q^{-1}(k)} \right\} \tag{4.28}$$

and

$$J_{r,T}(x(r)) = \min_{\{x(r+1),\ldots,x(T)\}} \left\{ \frac{1}{2} \sum_{k=r}^{T} ||y(k) - C(k)x(k)||^2_{R^{-1}(k)} + \right.$$

$$\left. + \frac{1}{2} \sum_{k=r}^{T-1} ||x(k+1) - \Phi(k+1,k)x(k)||^2_{Q^{-1}(k)} \right\} \tag{4.29}$$

Both $J_{o,r}(x(r))$ and $J_{r,T}(x(r))$ are quadratic forms in $x(r)$, (say)

$$J_{o,r}(x(r)) = \frac{1}{2} x(r)' F_{o,r} x(r) + g'_{o,r} x(r) + h_{o,r} \tag{4.30}$$

$$J_{r,T}(x(r)) = \frac{1}{2} x(r)' F_{r,T} x(r) + g'_{r,T} x(r) + h_{r,t} \tag{4.31}$$

Therefore,

$$J_r(x(r)) = \frac{1}{2}\, x(r)'(F_{o,r} + F_{r,T})x(r) + (g_{o,r} + g_{r,T})'x(r) +$$

$$+ (h_{o,r} + h_{r,T}) \tag{4.32}$$

and so the smoothed estimate and covariance are simply

$$\hat{x}_s(r) = -(F_{o,r} + F_{r,T})^{-1}(g_{o,r} + g_{r,T}) \tag{4.33}$$

$$P_s(r) = (F_{o,r} + F_{r,T})^{-1} \tag{4.34}$$

What remains is to determine recursive expressions for $J_{o,r}(x(r))$ and $J_{r,T}(x(r))$.

First consider $J_{o,r}(x(r)$ defined by (4.28). Note that this is just the cost functional one would minimize to obtain the maximum *a posteriori* estimate $\hat{x}(r|r-1)$, i.e.

$$\frac{1}{2}\, x(r)'F_{o,r}x(r) + g'_{o,r}x(r) = \frac{1}{2}[x(r) - \hat{x}(r|r-1)]'P^{-1}(r|r-1) \cdot$$

$$\cdot\, [x(r) - \hat{x}(r|r-1)] + \text{constant} \tag{4.35}$$

Therefore,

$$F_{o,r} = P^{-1}(r|r-1) \tag{4.36}$$

$$g_{o,r} = P^{-1}(r|r-1)\hat{x}(r|r-1) \tag{4.37}$$

and $h_{r,o}$ is of no real interest. Moreover, the Kalman filter provides a recursive computation of $\hat{x}(r|r-1)$ and $P(r|r-1)$. Thus recursive expressions for $F_{o,r}$ and $g_{o,r}$ are available.

Second, consider $J_{r,T}(x(r))$. Clearly,

$$J_{r,T}(x(r)) = \min_{\{x(r+1),\ldots,x(T)\}} \left\{ \frac{1}{2} \sum_{k=r}^{T} \left| \left| y(k) - C(k)x(k) \right| \right|^2_{R^{-1}(k)} + \right.$$

$$\left. + \frac{1}{2} \sum_{k=r}^{T} \left| \left| x(k+1) - \Phi(k+1,k)x(k) \right| \right|^2_{Q^{-1}(k)} \right\}$$

$$= \min_{\{w(r),\ldots,w(T-1)\}} \left\{ \frac{1}{2} \sum_{k=r}^{T} \left| \left| y(k) - C(k)x(k) \right| \right|^2_{R^{-1}(k)} + \right.$$

$$\left. + \frac{1}{2} \sum_{k=r}^{T-1} \left| \left| w(k) \right| \right|^2_{Q^{-1}(k)} \right\} \qquad (4.38)$$

subject to the constraint $x(k+1) = \Phi(k+1,k)x(k) + w(k)$, $k=r,\ldots,T-1$. But (4.38) is just a linear-quadratic optimal control problem and can be solved using dynamic programming from T backwards to r. The well-known solution to this problem yields a recursion for $F_{r,T}$ and $g_{r,T}$ ($h_{r,T}$ is of no interest) in terms of $F_{r+1,T}$, $g_{r+1,T}$, and $y(r)$. Thus there exist recursive relations for $J_{o,r}(x(r))$ and $J_{r,T}(x(r))$, and so $\hat{x}(r|T)$ and $P(r|T)$ can be found from (4.33) and (4.34).

This approach to the FI smoothing problem is easily extended to the continuous-time case. The details may be found in [41].

Mayne interprets this solution of the smoothing problem as a combination of two estimates. One estimate is based on past observations and is obtained from Kalman filtering; optimal control theory is used to obtain a second estimate from future observations. From (4.28), the estimate

based on past observations is a Bayesian estimate, but Mayne does not say what kind of estimate the one based on future observations is. In Section 4.3, this second estimate will be shown to be a maximum likelihood estimate. The idea of expressing the smoothed estimate as a linear combination of the two estimates was pursued in 1967 by Donald Fraser [42] for both continuous-time and discrete-time.

One of Fraser's two estimates is based on past observations. This estimate and the corresponding covariance are just the outputs of a standard Kalman filter working forward over the data. Fraser's second estimate is obtained from a backwards Kalman filter, i.e. a filter operating on future observations from T to the present time t. The idea is to then combine these two estimates using the formulas (4.14) and (4.15) for the optimal combination of independent estimates.

The appropriate continuous-time backward filter is [42]

$$-\frac{d}{dt}\hat{x}_b(t) = -A(t)\hat{x}_b(t) + P_b(t)C'(t)R^{-1}(t)[y(t) - C(t)\hat{x}_b(t)] \tag{4.39}$$

$$-\frac{d}{dt}P_b(t) = -A(t)P_b(t) - P_b(t)A'(t) + Q(t) + \\ - P_b(t)C'(t)R^{-1}(t)C(t)P_b(t) \tag{4.40}$$

where $P_b^{-1}(T) = 0$ and $\lim_{t \to T} [P_b^{-1}(t)\hat{x}_b(t)] = 0$. The interpretation given by Fraser and Potter [68] is that $\hat{x}_b(t)$ is, "...the best estimate of the state at time t based upon all the measurements from time t to the end of

the data interval." The terminal condition of an infinite covariance matrix is intended to reflect complete uncertainty about the state estimate at time T because of the complete lack of information about $x(T)$. Thus no terminal estimate can be made; only the limit of $P_b^{-1}(t)\hat{x}_b(t)$ can be specified. Because of these initial conditions, the filter must be implemented in the so-called "information filter" form [68].

The smoothed estimate is formed by combining the "independent" past and future estimates according to (4.14) and (4.15), viz.

$$\hat{x}_s(t) = P_s(t)[P^{-1}(t|t)\hat{x}(t|t) + P_b^{-1}(t)\hat{x}_b(t)] \tag{4.41}$$

$$P_s(t) = [P^{-1}(t|t) + P_b^{-1}(t)] \tag{4.42}$$

This is the same formula as Mayne's (4.33) and (4.34) in continuous-time if $P_b^{-1}(t) = F_{t,T}$ and $P_b^{-1}(t)\hat{x}_b(t) = -g_{t,T}$. Rather than showing these two equalities, Fraser's method of proof consists of re-deriving the smoothing formulas of Rauch, Tung, Striebel [69] from (4.39)-(4.42) and the usual Kalman filter equations. This is certainly a valid method of proof, and it does show that the smoothed estimate is given by (4.41) and (4.42). What is not clear, however, is <u>why</u> the estimates $\hat{x}(t|t)$ and $\hat{x}_b(t)$ can be combined by (4.41) and (4.42) or <u>why</u> the estimate $\hat{x}_b(t)$ is given by the backward Kalman filter (4.39) and (4.40). It would be desirable to answer these questions starting from basic principles.

Raman Mehra [43] attempts to clarify these points in his doctoral thesis. First consider the backward filtering equations (4.39) and (4.40). By multiplying the state equation (4.1) times -1, Mehra obtains

$$-\frac{d}{dt}x(t) = [-A(t)]x(t) - w(t) \qquad (4.43)$$

He then applies the usual Kalman filter equations to this backwards system by letting $\tau=T-t$ and thereby obtains (4.39) and (4.40). This same argument was later adopted by Kailath and Frost [70]. It is incorrect, however, because "future" (with respect to τ) values of the driving noise w are correlated with the present state (see Ljung and Kailath [71] where this observation was first made). That is, (4.43) is not a usual state-space realization. Therefore, it is not possible to blindly apply the Kalman filter to (4.43) and obtain the backward filter (4.39) and (4.40).

Mehra also addresses the question of independence of the estimates $\hat{x}(t|t)$ and $\hat{x}_b(t)$. His approach is to write the differential equations for the forward error $\tilde{x}(t|t)$ and the backward error $\tilde{x}_b(t)$,

$$\frac{d}{dt}\tilde{x}(t|t) = [A(t) - K(t)C(t)]\tilde{x}(t|t) + w(t) - K(t)v(t) \qquad (4.44)$$

$$-\frac{d}{dt}\tilde{x}_b(t) = [-A(t) - K_b(t)C(t)]\tilde{x}_b(t) - w(t) - K_b(t)v(t) \qquad (4.45)$$

Equation (4.44) is integrated from 0 to t while (4.45) is integrated from T to t. Thus Mehra points out that $\tilde{x}(t|t)$ depends on $\{w(\tau), v(\tau) \mid 0 < \tau \leq t\}$

and $\tilde{x}_b(t)$ depends on $\{w(\tau), v(\tau) \mid t < \tau \leq T\}$ -- two independent sets of noises. Is this sufficient for the conclusion that $\tilde{x}(t|t)$ and $\tilde{x}_b(t)$ are independent? Obviously not, $\tilde{x}(t|t)$ and $\tilde{x}_b(t)$ may be dependent because of their initial values. For example, the random variable $x(o)$ is correlated with (in fact, equal to) $\tilde{x}(o|o)$ and therefore with $\tilde{x}(t|t)$. Is $\tilde{x}_b(T)$ also correlated with $x(o)$? Mehra can't say because at this point in his development he has not specified any initial values for $\tilde{x}_b(T)$ or $P_b(T)$. It should be clear that without such a specification, the independence of $\tilde{x}(t|t)$ and $\tilde{x}_b(t)$ is indeterminate. Nevertheless, Mehra prematurely declares that they are independent because they are functions of independent sets of noises. The independence of these two estimates will be examined further in Section 4.3.

The behavior of the the two-filter smoother when there are errors in the various model parameters (such as the system matrix or initial covariance) was also considered by Mehra [43]. Following the work of Nishimura [72] and Fitzgerald [73], Mehra performs a sensitivity analysis to obtain an expression for the covariance of the errors of the forward and backward Kalman filters. In order to obtain the smoothed error covariance, he combines these two covariances assuming the forward and backward errors of the mismatched filters are uncorrelated! Of course this is not the case, and in Section 4.4 an expression for the smoothed error covariance is found which includes the correlation between the forward and backward errors.

In summary, this section has presented the two-filter smoother as

developed by Mayne [41] and Fraser [42]. This solution of the FI smoothing problem is unique compared to the Rauch, Tung, Striebel [69] smoother, and others [66], in that it is not given as a correction to the Kalman filter estimate at the same point. Rather, it takes the form of a combination of two optimal linear filter estimates. The work of Mehra [43] was primarily directed toward deriving this smoother from basic principles. Unfortunately, as discussed above, this derivation is incorrect. It must be realized, in fairness to Mehra, that the reversed-time Markov models which are so crucial to the development in the next section, were not available when he performed the work in [43]. Section 4.3 presents a second attempt at obtaining the two-filter smoother from basic principles by carefully considering the use of future observations for estimating the present state.

4.3 A New Solution to the Fixed-Interval Smoothing Problem

4.3.1 Motivation

When forming the smoothed Bayesian estimate of $x(t)$, there are three separate sets of information,

(i.) *a priori* data, $E\, x(t) = 0$ and $E\, x(t)x'(t) = \Sigma(t)$

(ii.) past observations, $\{y(\tau) \mid 0 \leq \tau \leq t\}$

(iii.) future observations, $\{y(\tau) \mid t < \tau \leq T\}$

Intutitvely, the smoothed estimate should incorporate each of these sets exactly once. One of the main contributions of Section 4.3 will be to show how this incorporation takes place in the two-filter smoother.

The filtered estimate $\hat{x}(t|t)$ is based on the *a priori* data and the past observations. This is easily obtained via the Kalman filter. Also, it is a simple matter to form an estimate of $x(t)$ from just the *a priori* data, i.e. the estimate is zero and the covariance is $\Sigma(t)$. What is less well-known, however, is how to use future observations in forming an estimate of $x(t)$. To this end, reversed-time Markov models will be introduced in the next subsection. When combined with the Kalman filter, these reversed-time models yield the expression for a Bayesian estimate of $x(t)$ based on *a priori* data and future observations.

Only the continuous-time problem is considered in Section 4.3. The analogous results for the discrete-time version are presented in Appendix B. Note that with respect to the continuous-time problem, the present observation $y(t)$ is a linear measurement of $x(t)$ corrupted by additive noise having an infinite covariance. Thus the isolated observation $y(t)$ contains no information about the process $x(t)$. This remark is purely formal, of course, as is the entire development of this section. These arguments can be made rigorous through the use of Ito calculus, but for ease of presentation and understanding, a formal development is deemed perferable. The future observations, therefore, can be defined as $\{y(\tau)|t \leq \tau \leq T\}$, where now $y(t)$ is included in the future observations, without altering the analysis. This is in contrast to the discrete-time case where the present observation contains non-zero information and is the major cause of any differences between the equations of Section 3 and those of Appendix B.

4.3.2 Reversed-Time Markov Models

Essentially simultaneously in the summer of 1976, several authors introduced reversed-time Markov models [71], [74], [75]. Corresponding to the forward system of (4.1), consider the reversed-time model

$$-\frac{d}{dt}x_r(t) = [-A(t) - Q(t)\Sigma^{-1}(t)]x_r(t) + \xi(t) \qquad (4.46)$$

where

$$E\, x_r(T) = 0,\ E\, \xi(t) = 0,\ E\, x_r(T)x_r'(T) = \Sigma(T)$$

$$E\, \xi(t)\xi'(\tau) = Q(t)\delta(t-\tau)\ ,\ E\, \xi(t)x_r'(T) = 0$$

Equation (4.46) is meant to denote that the reversed-time process $x_r(t)$ propagates backwards from T to 0, i.e. $x_r(t)$ is a Gauss-Markov process in negative time.

Theorem 4.1. The stochastic process x(t) of (4.1) and the stochastic process $x_r(t)$ of (4.46) have the same covariance function.

proof: A simple algebraic proof is given in [71].

Since x and x_r are both zero-mean, Theorem 4.1 states that they are equivalent up to second order properties. If both processes are Gaussian, then they have the same joint probability density functions. Therefore, (4.1) and (4.46) can be viewed as two realizations of the same stochastic process. That is, x and x_r are stochastically indistinguishable. This does not mean, however, that x and x_r have the same sample paths. One

implication of this equivalence is that the observations $y_r(t)$,

$$y_r(t) = C(t)x_r(t) + v(t) \tag{4.47}$$

are stochastically equivalent to the observations $y(t)$ on $x(t)$. Another implication of this equivalence is any least-squares linear estimator of $x(t)$ is also a least-squares linear estimator for $x_r(t)$, and vice versa. That is, given any set of observations, the estimate of $x(t)$ is the same functional on these observations as is the estimate of $x_r(t)$. This is a key point in the development of the sequel.

Recall the time-varying Lyapunov equation describing the state covariance $\Sigma(t)$,

$$\frac{d}{dt}\Sigma(t) = A(t)\Sigma(t) + \Sigma(t)A'(t) + Q(t) \tag{4.48}$$

Using this equation, the reversed-time system matrix $-A(t)-Q(t)\Sigma^{-1}(t)$ may be written as

$$-A(t)-Q(t)\Sigma^{-1}(t) = \Sigma(t)A'(t)\Sigma^{-1}(t) + \left(-\frac{d}{dt}\Sigma^{-1}(t)\right)\Sigma(t) \tag{4.49}$$

The stability of the reversed-time system can now be examined from (4.49). Consider first the case of a time-invariant system in the steady-state. Then the reversed-time system matrix is simply $\Sigma A'\Sigma^{-1}$. But A must be a stability matrix and has the same eigenvalues as A'. Moreover, $\Sigma A'\Sigma^{-1}$ is just a similarity transformation of A' and therefore has the same eigenvalues. The conclusion then is that the reversed-time system matrix

$\Sigma A' \Sigma^{-1}$ is also a stability matrix and has the same eigenvalues as the forward system matrix A. Therefore the forward-time process x(t) and the reversed-time process $x_r(t)$ both have stable realizations.

In the time-varying case, it is necessary to consider the adjoint system of (4.1),

$$-\frac{d}{dt} p(t) = A'(t)p(t) \tag{4.50}$$

This system propagating backwards in time has the same stability properties as the original forward system. Let $z(t) = \Sigma(t)p(t)$. Then

$$-\frac{d}{dt} z(t) = \left[\Sigma(t)A'(t)\Sigma^{-1}(t) + \left(-\frac{d}{dt}\Sigma(t)\right)\Sigma^{-1}(t)\right] z(t) \tag{4.51}$$

The resulting reversed-time system matrix is, according to (4.49), the same as the system matrix for the reversed-time process x_r. The system (4.46) will have the same stability properties as (4.50), and hence as (4.1), if $z(t) = \Sigma(t)p(t)$ is a Lyapunov transformation [53]. For this transformation to be a Lyapunov transformation, the following conditions must hold [53]:

(i.) Σ has a continuous derivative

(ii.) Σ and $\frac{d}{dt}\Sigma$ are bounded

(iii.) there exists a constant m such that $0 < m \leq |\det \Sigma(t)|$, $\forall t$

Assuming these conditions are met, the forward- and reversed-time realizations (4.1) and (4.46) possess identical stability properties.

4.3.3 An Estimate Based on Future Observations Plus *A Priori* Information

The conditional expectation of $x_r(t)$ (or $x(t)$) given the future observations is denoted $\hat{x}_r(t|t)$,

$$\hat{x}_r(t|t) = E\{x_r(t)|y(\tau),\ t \leq \tau \leq T\} \tag{4.52}$$

The process $x_r(t)$ is a Gauss-Markov process in reversed-time as given by (4.46). Therefore, this Bayesian estimate can be computed from the Kalman filter for the reversed-time system model. Explicitly,

$$-\frac{d}{dt}\hat{x}_r(t|t) = [-A(t)-Q(t)\Sigma^{-1}(t)]\hat{x}_r(t|t) + \\ + K_r(t)[y(t)-C(t)\hat{x}_r(t|t) \tag{4.53}$$

$$-\frac{d}{dt}P_r(t|t) = [-A(t)-Q(t)\Sigma^{-1}(t)]P_r(t|t) + \\ + P_r(t|t)[-A(t)-Q(t)\Sigma^{-1}(t)]' + Q(t) + \\ - P_r(t|t)C'(t)R^{-1}(t)C(t)P_r(t|t) \tag{4.54}$$

$$K_r(t) = P_r(t|t)C'(t)R^{-1}(t) \tag{4.55}$$

where $\hat{x}_r(T|T) = 0$ and $P_r(T) = \Sigma(T)$. Note that the conditions at time T for this filter are finite, in contrast with the initial conditions of Fraser's backward filter (4.39) and (4.40).

The Bayesian estimate $\hat{x}_r(t|t)$ of $x(t)$ is a combination of *a priori*

information and the future observations. Of course, it is possible to estimate x(t) from only the *a priori* data,

$$\hat{x}_{a.p.}(t) = 0 \tag{4.56}$$

$$P_{a.p.}(t) = \Sigma(t) \tag{4.57}$$

Consider now a maximum likelihood estimate of x(t) based only on the future observations, i.e. an estimate that uses the future observations but <u>not</u> *a priori* data. Denote this estimate as $\hat{x}_{future}(t)$ and the covariance as $P_{future}(t)$. The error $x(t) - \hat{x}_{future}(t)$ in this estimate is solely due to the driving noise and the observation noise over the interval [t,T]. Therefore, the estimates $\hat{x}_{future}(t)$ and $\hat{x}_{a.p.}(t)$ have independent errors. This means that

$$\hat{x}_{future}(t) = x(t) + \nu(t) \tag{4.58}$$

where $E\,\nu(t)\nu'(t) = P_{future}(t)$ and $E\,x(t)\nu'(t) = 0$. Equation (4.58) treats the random variable $\hat{x}_{future}(t)$ as a noisy observation of x(t). By proposition 4.2, the Bayesian estimate of x(t) given $\hat{x}_{future}(t)$ (or, equivalently, given the future observations) is

$$\begin{aligned}\hat{x}_r(t|t) &= P_r(t|t)\left[P^{-1}_{future}(t)\,\hat{x}_{future}(t) + \Sigma^{-1}(t)\hat{x}_{a.p.}(t)\right] \\ &= P_r(t|t)\,P^{-1}_{future}(t)\,\hat{x}_{future}(t)\end{aligned} \tag{4.59}$$

$$P_r(t|t) = \left[P^{-1}_{future}(t) + \Sigma^{-1}(t) \right]^{-1} \tag{4.60}$$

An interesting feature of (4.59) and (4.60) is that they are invertible in that it is possible to solve for $\hat{x}_{future}(t)$ and $P_{future}(t)$ in terms of the other quantities. This yields

$$\hat{x}_{future}(t) = P_{future}(t) P^{-1}_r(t|t) \hat{x}_r(t|t) \tag{4.61}$$

$$P_{future}(t) = \left[P^{-1}_r(t|t) - \Sigma^{-1}(t) \right]^{-1} \tag{4.62}$$

Differentiating (4.61) and (4.62) with respect to -t yields differential equations for the maximum likelihood estimate $\hat{x}_{future}(t)$ and $P_{future}(t)$ propagating backwards from T. This is precisely what is done in the proof of

Theorem 4.2. The maximum likelihood estimate $\hat{x}_{future}(t)$ and covariance $P_{future}(t)$ are identically equal to Fraser's $\hat{x}_b(t)$ and $P_b(t)$. That is,

$$\hat{x}_b(t) = P_b(t) P^{-1}_r(t|t) \hat{x}_r(t|t) \tag{4.63}$$

$$P_b(t) = \left[P^{-1}_r(t|t) - \Sigma^{-1}(t) \right]^{-1} \tag{4.64}$$

proof: Explicit dependence on t is suppressed throughout this proof. (covariance) It is shown that P^{-1}_b equals $P^{-1}_r - \Sigma^{-1}$. At time T, $P^{-1}_r(T|T) - \Sigma^{-1}(T) = 0 = P^{-1}_b(T)$, so it suffices to show that the derivatives are equal.

proof: (contd.)

$$-\frac{d}{dt}P_b^{-1} = -P_b^{-1}\left[-\frac{d}{dt}P_b\right]P_b^{-1}$$

$$= P_b^{-1}A + A'P_b^{-1} - P_b^{-1}Q\,P_b^{-1} + C'R^{-1}C \qquad \text{from (4.40)}$$

$$= \left(P_r^{-1} - \Sigma^{-1}\right)A + A'\left(P_r^{-1} - \Sigma^{-1}\right) - \left(P_r^{-1} - \Sigma^{-1}\right)Q\left(P_r^{-1} - \Sigma^{-1}\right) +$$

$$+ C'R^{-1}C \qquad \text{by hypothesis}$$

$$= P_r^{-1}\left(A + Q\Sigma^{-1}\right) + \left(A + Q\Sigma^{-1}\right)'P_r^{-1} - P_r^{-1}Q\,P_r^{-1} + C'R^{-1}C +$$

$$- \Sigma^{-1}A - A'\Sigma^{-1} - \Sigma^{-1}Q\,\Sigma^{-1} \qquad \text{rearranging}$$

$$= -\frac{d}{dt}P_r^{-1} + \frac{d}{dt}\Sigma^{-1} \qquad \text{from (4.48) and (4.54)}$$

$$= -\frac{d}{dt}\left[P_r^{-1} - \Sigma^{-1}\right]$$

(estimate) The proof is completed by showing $P_b^{-1}\hat{x}_b$ equals $P_r^{-1}\hat{x}_r$. Once again, it suffices to demonstrate the equality of the derivatives since the quantities are equal at time T.

$$-\frac{d}{dt}\left[P_b^{-1}\hat{x}_b\right] = \left(-\frac{d}{dt}P_b^{-1}\right)\hat{x}_b + P_b^{-1}\left(-\frac{d}{dt}\hat{x}_b\right)$$

$$= \left(P_b^{-1}A + A'P_b^{-1} - P_b^{-1}Q\,P_b^{-1} + C'R^{-1}C\right)\hat{x}_b +$$

$$+ P_b^{-1}\left(-A\hat{x}_b + P_bC'R^{-1}\left[y - C\hat{x}_b\right]\right) \qquad \text{from (4.39) and (4.40)}$$

$$= \left(A' - P_b^{-1}Q\right)P_b^{-1}\hat{x}_b + C'R^{-1}y \qquad \text{combining terms}$$

proof: (contd.)

$$-\frac{d}{dt}\left[P_b^{-1}\hat{x}_b\right] = \left(A' - \left[P_r^{-1} - \Sigma^{-1}\right]Q\right) P_r^{-1}\hat{x}_r + C'R^{-1}y \qquad \text{by hypothesis}$$

$$= \left(\left[A + Q\Sigma^{-1}\right]' - P_r^{-1}Q\right) P_r^{-1}\hat{x}_r + C'R^{-1}y \qquad \text{rearranging terms}$$

$$= \left(P_r^{-1}\left[A + Q\Sigma^{-1}\right] + \left[A + Q\Sigma^{-1}\right]' P_r^{-1} - P_r^{-1}Q\,P_r^{-1} + C'R^{-1}C\right)\hat{x}_r +$$

$$+ P_r^{-1}\left(\left[-A - Q\Sigma^{-1}\right]\hat{x}_r + P_r C'R^{-1}\left[y - C\hat{x}_r\right]\right)$$

adding and subtracting $P_r^{-1}\left[A + Q\Sigma^{-1}\right]\hat{x}_r + C'R^{-1}C\hat{x}_r$

$$= \left(-\frac{d}{dt}P_r^{-1}\right)\hat{x}_r + P_r^{-1}\left(-\frac{d}{dt}\hat{x}_r\right) \qquad \text{from (4.53) and (4.54)}$$

$$= -\frac{d}{dt}\left[P_r^{-1}\hat{x}_r\right]$$

Q.E.D.

This result says that the *a priori* information can be "subtracted out" from the conditional expectation of x(t) to form the backward estimate. Moreover, this backward estimate is the maximum likelihood estimate of x(t). The conditional expectation comes from a reversed-time Kalman filter. Using this Kalman filter together with the Lyapunov equation for the state covariance has yielded a differential equation for the maximum likelihood estimate of x(t) based on future observations.

4.3.4 The Solution

Theorem 4.3. The smoothed Bayesian estimate and covariance satisfy

$$\hat{x}_s(t) = P_s(t)\left[P^{-1}(t|t)\hat{x}(t|t) + P_r^{-1}(t|t)\hat{x}_r(t|t)\right] \tag{4.65}$$

$$P_s(t) = \left[P^{-1}(t|t) + P_r^{-1}(t|t) - \Sigma^{-1}(t)\right] \tag{4.66}$$

proof: There are two realizations of the process $x(\tau)$, a forward-time realization (4.1) and a reversed-time realization (4.46). A third realization is introduced in this proof which combines (4.1) and (4.46) to propagate $x_c(t)$ forward and backward from time t. The process $x_c(\tau)$ is generated by

$$\frac{d}{d\tau} x_c(\tau) = A(\tau)x_c(\tau) + w(\tau) \quad , \tau > t$$

$$-\frac{d}{d\tau} x_c(\tau) = \left[-A(\tau) - Q(\tau)\Sigma^{-1}(\tau)\right] x_c(\tau) + \xi(\tau) \quad , \tau < t$$

where $E\, x_c(t) = 0$, $E\, w(\tau) = 0$, $E\, \xi(\tau) = 0$, $E\, x_c(t)x_c'(t) = \Sigma(t)$,

$$E\, w(\tau)w'(\sigma) = Q(\tau)\delta(\tau-\sigma), \quad E\, \xi(\tau)\xi'(\sigma) = Q(\tau)\delta(\tau-\sigma),$$

$$E\, w(\tau)x_c'(t) = E\, \xi(\tau)x_c'(t) = E\, w(\tau)\xi'(\sigma) = 0$$

These differential equations are meant to denote that $x_c(\tau)$ may be written as

$$x_c(\tau) = \begin{cases} \int_t^\tau A(\sigma)x_c(\sigma)d\sigma + \int_t^\tau dw(\sigma) + x_c(t) & , \tau > t \\ \int_t^\tau \left[-A(\sigma)-Q(\sigma)\Sigma^{-1}(\sigma)\right] x_c(\sigma)d\sigma + \int_t^\tau d\xi(\sigma) + x_c(t) & , \tau < t \end{cases}$$

The process $x_c(\tau)$ is easily shown to be stochastically equivalent to $x(\tau)$ and $x_r(\tau)$. Let $\hat{x}_{past}(t)$ and $P_{past}(t)$ be the maximum likelihood estimate of $x(t)$ and the error covariance given the past observations. By applying the same argument that was used in Section 4.3.3 for $\hat{x}_{future}(t)$ and $P_{future}(t)$ to $\hat{x}_{past}(t)$ and $P_{past}(t)$ one obtains

$$\hat{x}_{past}(t) = P_{past}(t)P^{-1}(t|t)\hat{x}(t|t)$$

$$P_{past}(t) = \left[P^{-1}(t|t) - \Sigma^{-1}(t)\right]^{-1}$$

Consider now the third realization, i.e. the process $x_c(\tau)$. The errors in the maximum likelihood estimates $\hat{x}_{past}(t)$ and $\hat{x}_{future}(t)$ are caused by $\{\xi(\tau), v(\tau) | 0 \leq \tau < t\}$ and $\{w(\tau), v(\tau) | t < \tau \leq T\}$, respectively. Therefore these estimates are independent estimates, and Proposition 4.1 can be used to obtain the maximum likelihood estimate of $x_c(t)$ (or $x(t)$) given all the observations,

$$\hat{x}_{ML}(t) = P_{ML}(t)\left[P_{past}^{-1}(t)\hat{x}_{past}(t) + P_{future}^{-1}(t)\hat{x}_{future}(t)\right]$$

$$P_{ML}(t) = \left[P_{past}^{-1}(t) + P_{future}^{-1}(t)\right]^{-1}$$

From Proposition 4.2, the smoothed Bayesian estimate is

$$\hat{x}_s(t) = P_s(t)P_{ML}^{-1}(t)\hat{x}_{ML}(t)$$

$$= P_s(t)\left[P_{past}^{-1}(t)\hat{x}_{past}(t) + P_{future}^{-1}(t)\hat{x}_{future}(t)\right]$$

$$= P_s(t)\left[P^{-1}(t|t)\hat{x}(t|t) + P_r^{-1}(t|t)\hat{x}_r(t|t)\right]$$

$$P_s(t) = \left[P_{ML}^{-1}(t) + \Sigma^{-1}(t)\right]^{-1}$$

$$= \left[P_{past}^{-1}(t) + P_{future}^{-1}(t) + \Sigma^{-1}(t)\right]^{-1}$$

$$= \left[P^{-1}(t|t) + P_r^{-1}(t|t) - \Sigma^{-1}(t)\right]$$

Q.E.D.

Aside: An alternate proof of Theorem 4.3 is to note that substitution of (4.63) into (4.41) yields (4.65) and substitution of (4.64) into (4.42) yields (4.66).

This theorem expresses the smoothed estimate as a combination of two filtered estimates

- one estimate from a forward Kalman filter for the forward system model
- one estimate from a reversed-time Kalman filter for the reversed-time model.

These estimates are not independent, however, because they both include the *a priori* information. The two different sets (past and future)

of observations may be said to be independent observations of $x_c(t)$ because the observation noise processes are independent. Schoute, et al. [76] have obtained very similar expressions to (4.65) and (4.66) for the smoothed estimate of a discrete-time system in the steady-state. Their approach is closely related to the approach taken here in that they employ a reversed-time model and Kalman filter.

One striking characteristic of the smoother Theorem 4.3 is the complete symmetry with respect to forward-time vs. reversed-time. Equations (4.65) and (4.66) are called symmetric because both estimates $\hat{x}(t|t)$ and $\hat{x}_r(t|t)$ are conditional expectations of $x(t)$ given the past and the future observations, respectively. This is certainly in contrast with the usual two-filter smoother. The symmetry between forward- and reversed-time will be developed and discussed further in Section 4.3.6.

Another important characteristic of this smoother is that both the forward and backward models used in forming the two Kalman filters are stable (assuming the original forward realization is stable). This feature will allow a sensitivity analysis in Section 4.4 that requires the integration of only stable differential equations.

Finally, it is noted that even though this development assumed the random process $x(t)$ was zero-mean, the case of a nonzero-mean process is easily handled. Letting $z(t)$ equal $x(t)$ minus the mean value of $x(t)$ yields a zero-mean process obeying the same state equation (4.1). Then taking as observations of $z(t)$ the observations $y(t)$ minus the mean value of $y(t)$ produces a smoothing problem of the form studied here. The

smoothed estimate of x(t) is simply the smoothed estimate of z(t) plus the mean of x(t).

4.3.5 The Linear Time-Invariant Infinite-Lag Case

For the special case of smoothing over the interval $(-\infty, +\infty)$ with a time-invariant system, it is possible to investigate the symmetry between forward- and reversed-time in detail. If the system matrix A is a scalar, then the reversed-time system matrix $\Sigma A'\Sigma^{-1}$ equals A. That is, from (4.46), the reversed-time realization is identical to the forward-time realization. Thus the two Kalman filters are identical, and so the two steady-state error covariances P and P_r are equal. From (4.65), it is clear that this implies the two estimates $\hat{x}(t|t)$ and $\hat{x}_r(t|t)$ are weighted equally in forming the smoothed estimate. This confirms the intuitive expectation of Section 4.1 that the future and past should be equivalent.

The following proposition extends the above result beyond the scalar case:

Proposition 4.3. The reversed-time system (4.46) equals the forward-time system (4.1) in the time-invariant infinite-lag case if and only if the autocorrelation function of x is symmetric, i.e. $\Sigma A'\Sigma^{-1} = A$ if and only if $R_x(t) = R_x'(t)$.

proof: (if) By assumption $R_x(t) = R_x'(t)$, hence

$$e^{At}\Sigma = \Sigma\, e^{A't} \qquad \forall t$$

$$\Rightarrow e^{At} = \Sigma\, e^{A't}\Sigma^{-1} = e^{\Sigma A'\Sigma^{-1}t}$$

$$\therefore A = \Sigma A'\Sigma^{-1}$$

(only if) $\quad R_x(t) = e^{At}\Sigma$

$$= e^{(\Sigma A'\Sigma^{-1}t)}\Sigma \qquad \text{(by assumption)}$$

$$= \Sigma\, e^{A'(t)}$$

$$= R_x'(t)$$

Q.E.D.

Of course $R_x(t) = R_x'(t)$ is equivalent to $R_x(t) = R_x(-t)$, the condition needed in Section 4.1 to show that the same Wiener filter can be used to estimate $x(t)$ from either the future or the past. From Proposition 4.3, whenever the autocorrelation function is even, the future and past observations are equally valuable in forming an estimate of the present.

Theorem 4.3 expresses the smoothed estimate of $x(t)$ as a linear combination of a causal filter estimate and an anti-causal filter estimate. In the LTI infinite-lag case, the Wiener filter and smoother can also be used to obtain these estimates. From Theorem 4.3, an identity between the Wiener filter and smoother can be derived.

Let $H_f(s)$ be the transfer function of the Wiener filter for estimating x(t). Then [77]

$$H_f(s) = \frac{1}{S_y^+(s)} \left\{ \frac{S_{xy}(s)}{S_y^-(s)} \right\}_+ \tag{4.67}$$

where $S_y(s)$ = bilateral Laplace transform of the autocorrelation function of y

$S_{xy}(s)$ = bilateral Laplace transform of the cross-correlation function between x and y

$$S_y^+(s) = \frac{\prod_{i=1}^{m} (s-z_i)}{\prod_{i=1}^{n} (s-p_i)} \quad \text{for } \{z_i\}_1^m \left(\{p_i\}_1^n \right) \text{ the LHP zeros (poles) of } S_y(s)$$

$$S_y^-(s) = \frac{S_y(s)}{S_y^+(s)} = S_y^+(-s)$$

and the operator $\{\cdot\}_+$ is uniquely defined by

$$\left\{F(s)\right\}_+ = \int_{0^-}^{+\infty} \left[\int_{-j\infty}^{+j\infty} \frac{F(s)e^{st}}{2\pi j} \, ds \right] e^{-st} dt$$

i.e. the transform of the causal part of the inverse transform of F(s).

The transfer function of the Wiener smoother is simply

$$H_s(s) = \frac{S_{xy}(s)}{S_y(s)} \tag{4.68}$$

From (4.65), the causal part of the Wiener smoother must equal $P_s P^{-1}$ times the Wiener filter since they are the same function of the observations. That is,

$$\left\{H_s(s)\right\}_+ = P_s P^{-1} H_f(s) \tag{4.69}$$

This completes the proof of the next proposition.

Proposition 4.4. $$P P_s^{-1} \left\{\frac{S_{xy}(s)}{S_y^-(s)}\right\}_+ = \frac{1}{S_y^+(s)} \left\{\frac{S_{xy}(s)}{S_y^-(s)}\right\}_+$$

Of course a similar identity exists between the anti-causal Wiener filter and the anti-causal part of the smoother. Proposition 4.4 says that the Wiener filter has the same poles as the causal part of the Wiener smoother; the zeros, however, are different.

4.3.6 Discussion

Two key ingredients of the two-filter smoother are the maximum likelihood estimates $\hat{x}_{past}(t)$ and $\hat{x}_{future}(t)$. Theorem 4.2 showed that $\hat{x}_{future}(t)$ equals Fraser's backward estimate $\hat{x}_b(t)$. There is also a backwards estimate $\hat{x}_{r_b}(t)$ of the reversed-time process $x_r(t)$. The estimate $\hat{x}_{r_b}(t)$ is based on the observations $\{y(\tau) | 0 \leq \tau \leq t\}$ and may be shown (by repeating the proof of Theorem 4.2) to equal $\hat{x}_{past}(t)$. The differential equations for $\hat{x}_b(t)$ and $P_{r_b}(t)$ are just Fraser's (4.39) and (4.40) with the reversed-time system matrix in place of the forward system matrix,

$$\frac{d}{dt}\hat{x}_{r_b}(t) = \left[-A(t)-Q(t)\Sigma^{-1}(t)\right]\hat{x}_{r_b}(t) + K_{r_b}(t)\left[y(t)-C(t)\hat{x}_{r_b}(t)\right] \tag{4.70}$$

$$\frac{d}{dt}P_{r_b}(t) = \left[-A(t)-Q(t)\Sigma^{-1}(t)\right]P_{r_b}(t) + P_{r_b}(t)\left[-A(t)-Q(t)\Sigma^{-1}(t)\right]' + Q(t) +$$

$$- P_{r_b}(t)C'(t)R^{-1}(t)C(t)P_{r_b}(t) \tag{4.71}$$

where $K_{r_b}(t) = P_{r_b}(t)C'(t)R^{-1}(t)$, $P_{r_b}^{-1}(o) = 0$, and $\lim_{t\to 0}\left[P_{r_b}^{-1}(t)\hat{x}_{r_b}(t)\right] = 0$. This filter must be implemented as an information filter because of the initial conditions.

Using the estimate $\hat{x}_{r_b}$, it is possible to obtain another version of the two-filter smoother

$$\hat{x}_s(t) = P_s(t)\left[P_{r_b}^{-1}(t)\hat{x}_{r_b}(t) + P_r^{-1}(t|t)\hat{x}_r(t|t)\right] \tag{4.72}$$

$$P_s(t) = \left[P_{r_b}^{-1}(t) + P_r^{-1}(t|t)\right]^{-1} \tag{4.73}$$

This is essentially Fraser's smoother (4.41) and (4.42) applied to the reversed-time realization instead of the usual forward realization. The *a priori* information is combined with the future observations to form one estimate; the second estimate is formed from the past observations alone. All of the smoothing algorithms presented by Sidhu and Desai [74] are in this same spirit -- they are obtained by applying a standard smoothing

algorithm to the reversed-time model.

Consider now the problem of filtering for a linear system where the initial covariance $\Sigma(o)$ is unknown when the observation begins. One would like to operate a Kalman filter to obtain the estimate $\hat{x}(t|t)$, but the initial conditions for the filter are unknown. At first glance, it appears that it is possible to propagate $\hat{x}_{r_b}(t)$ and then later, if the covariance $\Sigma(o)$ became known, use (4.17) and (4.18) to construct the filtered estimate $\hat{x}(t|t)$. Indeed, applying the smoothing formula (4.72) and (4.73) at time T, the filtered estimate equals the smoothed estimate and can be written as

$$\hat{x}(T|T) = P(T|T)P_{r_b}^{-1}(T)\hat{x}_{r_b}(T) \tag{4.74}$$

$$P(T|T) = \left[P_{r_b}^{-1}(T) + \Sigma^{-1}(T)\right]^{-1} \tag{4.75}$$

Certainly (4.74) and (4.75) seem to accomplish what was just proposed.

This scheme, however, does not quite work because of the equation for $\hat{x}_{r_b}(t)$. From (4.70), the computation of this estimate involves the reversed-time system matrix $-A(t)-Q(t)\Sigma^{-1}(t)$. That is to say, $\hat{x}_{r_b}(t)$ depends on the covariance Σ, and if the covariance were unknown, then it would be impossible to compute $\hat{x}_{r_b}$. The point here is that $\hat{x}_{r_b}(t)$ is a maximum likelihood estimate of $x_r(t)$ based on its "future" observations (future with respect to $-t$, i.e. $\{y(\tau)|t \geq \tau \geq 0\}$). When the covariance of $x(t)$ is $\Sigma(t)$ as used in the reversed-time realization, the two pro-

cesses $x(t)$ and $x_r(t)$ are stochastically equivalent. But if $\Sigma(o)$ is unknown, then one does not know which reversed-time model to use in the computation of $\hat{x}_{r_b}(t)$.

One conclusion from the above analysis is that the smoothing formulas developed here do not yield a simple change of initial conditions result for the filtered estimate. It will now be shown that such a result can be obtained for the smoothed estimate $x_s(o)$.

The problem of interest is to compute the smoothed estimate of $x(o)$ by first processing the observations assuming $\Sigma(o)$ is unknown and then correcting this result to account for the known value of $\Sigma(o)$. But this is exactly what Fraser's two-filter smoother (4.41) and (4.42) does. The observations are used to compute $\hat{x}_b(o)$, and then the smoothed estimate is just

$$\hat{x}_s(o) = P_s(o)P_b^{-1}(o)\hat{x}_b(o) \tag{4.76}$$

$$P_s(o) = \left[\Sigma^{-1}(o) + P_b^{-1}(o)\right]^{-1} \tag{4.77}$$

since the filtered estimate $\hat{x}(o|o)$ is just the *a priori* zero value with covariance $\Sigma(o)$.

A closely related problem is the following: suppose the smoothed estimate $\hat{x}_s^{\Pi}(o)$ of $x(o)$ is obtained assuming that $E\, x(o)x'(o) = \Pi(o)$. The true initial covariance is $\Sigma(o)$, however. Is it possible to obtain the optimal smoothed estimate $\hat{x}_s(o)$ from $\hat{x}_s^{\Pi}(o)$ without further reference to the observations? Note that this is truly a change of initial condi-

tions question. The answer to this question is yes, and the result is easily obtained from (4.76) and (4.77). Applying these equations with the incorrect initial covariance $\Pi(o)$ yields

$$\hat{x}_s^{\Pi}(o) = P_s^{\Pi}(o) P_b^{-1}(o) \hat{x}_b(o) \tag{4.78}$$

$$P_s^{\Pi}(o) = \left[\Pi^{-1}(o) + P_b^{-1}(o) \right]^{-1} \tag{4.79}$$

Thus given $\hat{x}_s^{\Pi}(o)$ and $P_s^{\Pi}(o)$, the backward quantities may be computed as

$$\hat{x}_b(o) = P_b(o) P_s^{\Pi}(o)^{-1} \hat{x}_s^{\Pi}(o) \tag{4.80}$$

$$P_b(o) = \left[P_s^{\Pi}(o)^{-1} - \Pi^{-1}(o) \right]^{-1} \tag{4.81}$$

Substituting into (4.76) and (4.77) yields the simple relationship

$$\hat{x}_s(o) = P_s(o) P_s^{\Pi}(o)^{-1} \hat{x}_s^{\Pi}(o) \tag{4.82}$$

$$P_s(o) = \left[P_s^{\Pi}(o)^{-1} - \Pi^{-1}(o) + \Sigma^{-1}(o) \right] \tag{4.83}$$

The very natural interpretation of this result is that the incorrect covariance $\Pi(o)$ is removed from the estimate $\hat{x}_s^{\Pi}(o)$ and then the correct covariance $\Sigma(o)$ is added. Equations (4.82) and (4.83) are much simpler than the change of initial conditions formula for smoothed estimates presented by Ljung and Kailath in [78].

An important implication of this last result deals with the reversed-time estimate $\hat{x}_r(t|t)$. Recall that $\hat{x}_r(t|t)$ is the filtered estimate of

the reversed-time process $x_r(t)$ given observations from T to t. Thus

$$\hat{x}_s(o) = \hat{x}_r(o|o) \tag{4.84}$$

Now suppose that the backwards Kalman filter is designed for the reversed-time system

$$-\frac{d}{dt} x_r^{\Pi}(t) = \left[-A(t)-Q(t)\Pi^{-1}(t)\right] x_r^{\Pi}(t) + \xi(t) \tag{4.85}$$

where $E\, x_r^{\Pi}(T) x_r^{\Pi}(T)' = \Pi(T)$ and

$$\frac{d}{dt}\Pi(t) = A(t)\Pi(t) + \Pi(t)A'(t) + Q(t) \tag{4.86}$$

instead of the reversed-time system (4.46). This amounts to replacing Σ, the true state covariance, with Π, a quantity which also obeys the Lyapunov equation. The output of this reversed-time Kalman filter, initialized at time T with covariance $\Pi(t)$, is just $\hat{x}_r^{\Pi}(t|t)$. The smoothed estimate $\hat{x}_s(o)$ can now be found from the output of this (incorrect) Kalman filter by using (4.82) and (4.83).

It should be realized that the time 0 is not special in this development, and (4.78)-(4.85) can all be appropriately altered to be valid for any time t. The smoothed estimate and covariance can then be expressed in terms of the reversed-time estimate $\hat{x}_r^{\Pi}(t)$ as

$$\hat{x}_s(t) = P_s(t)\left[P^{-1}(t|t)\hat{x}(t|t) + P_r^{\Pi}(t|t)^{-1}\hat{x}_r^{\Pi}(t|t)\right] \tag{4.87}$$

$$P_s(t) = \left[P^{-1}(t|t) + P_s^{\Pi}(t|t)^{-1} - \Pi^{-1}(t)\right] \tag{4.88}$$

Equations (4.87) and (4.88) are a generalization of (4.65) and (4.66) of Theorem 4.3 in that the covariance Σ has been replaced by an arbitrary function Π satisfying (4.86).

Consider the implementation implications of this last observation. Basically any legitimate covariance function can be used in the reversed-time system matrix and Kalman filter. This added flexibility may be quite useful, especially when the forward system is time-invariant. In this case, one could use the steady-state covariance and thereby attain a time-invariant reversed-time model. This eliminates some of the problems involved with directly implementing the reversed-time filter.

4.4 Sensitivity Analysis and Reduced Order Smoothers

Sensitivity analysis is concerned with the increase in the smoothed error covariance caused by using incorrect model parameters. For example, if a smoother is implemented with the system matrix $A^*(t)$ in place of the correct matrix $A(t)$, what is the resulting error covariance? Reduced order smoothing refers to smoothing with a model of lower dimensional than the actual system (4.1). Griffin and Sage [79] have treated the sensitivity problem, but not the reduced order smoother, for only discrete-time processes by considering the Rauch, Tung, Striebel smoother [69]. The analysis given here is thought to be the first correct treatment of the two-filter smoother and is performed for both continuous- and discrete-time (see Appendix B for the discrete-time results). The reduced order smoothing problem will be treated first, and then the results will

be specialized to perform sensitivity analysis. Throughout this section, explicit time dependence will often be suppressed for ease of presentation.

The model used by the reduced order smoother for the dynamics and observations is

$$\frac{d}{dt} x^* = A^* x^* + w^* \tag{4.89}$$

$$y^* = C^* x^* + v^* \tag{4.90}$$

where w^* and v^* are independent white noise processes with covariances Q^* and R^*, respectively and $E\, x^*(o)x^{*\prime}(o) = \Sigma^*(o)$. The superscript asterisk will be used consistently to denote model parameters as distinguished from the true system parameters. It is assumed that there is an output z of the actual system defined by

$$z = H\, x \tag{4.91}$$

which is approximated by the output z^* of the model,

$$z^* = H^* x^* \tag{4.92}$$

The only restrictions imposed on the model are that y^* and z^* have the same dimensions as y and z, respectively.

A smoothed estimate of the output is obtained as $\hat{z}^*_s = H^* \hat{x}^*_s$, and the question is, "What is the covariance of $z - \hat{z}^*_s$?" In order to determine a expression for this covariance, first the forward-time system and

filter are jointly analyzed, and then a similar analysis is performed for the reversed-time system and filter. Next, the correlation between the forward- and reversed-time estimates are obtained. All these results are finally combined to yield the output error covariance.

Forward-Time System and Filter - The model (4.89) and (4.90) can be used to design a reduced order Kalman filter,

$$\frac{d}{dt}\hat{x}^* = A^*\hat{x}^* + K^*[y - C^*\hat{x}^*] \tag{4.93}$$

$$K^* = P^*C^{*\prime}R^{*-1} \tag{4.94}$$

$$\frac{d}{dt}P^* = A^*P^* + P^*A^{*\prime} + Q^* - P^*C^{*\prime}R^{*-1}C^*P^* \tag{4.95}$$

where $\hat{x}^*(o) = 0$ and $P^*(o) = \Sigma^*(o)$. Notice that the input to this filter is the actual observations y, of course. By combining the estimate $\hat{x}^*$ with the actual state x, one obtains an augmented state vector having dynamics

$$\frac{d}{dt}\begin{bmatrix} x \\ \hat{x}^* \end{bmatrix} = \begin{bmatrix} A & 0 \\ K^*C & A^*-K^*C^* \end{bmatrix}\begin{bmatrix} x \\ \hat{x}^* \end{bmatrix} + \begin{bmatrix} u \\ K^*v \end{bmatrix} \tag{4.96}$$

The covariance of this augmented state is defined as

$$\begin{bmatrix} \Sigma & M \\ M' & N \end{bmatrix} = E\left\{\begin{bmatrix} x \\ \hat{x}^* \end{bmatrix}\begin{bmatrix} x' & \hat{x}^{*\prime} \end{bmatrix}\right\} \tag{4.97}$$

and must obey the Lyapunov equation

$$\frac{d}{dt}\begin{bmatrix}\Sigma & M\\ M' & N\end{bmatrix} = \begin{bmatrix}A & 0\\ K^*C & A^*-K^*C^*\end{bmatrix}\begin{bmatrix}\Sigma & M\\ M' & N\end{bmatrix} + \begin{bmatrix}\Sigma & M\\ M' & N\end{bmatrix}\begin{bmatrix}A & 0\\ K^*C & A^*-K^*C^*\end{bmatrix}' +$$

$$+ \begin{bmatrix}Q & 0\\ 0 & K^*RK^{*\prime}\end{bmatrix} \tag{4.98}$$

Reversed-Time System and Filter - The reduced order reversed-time system corresponding to the model (4.89) is

$$-\frac{d}{dt}x_r^* = \left[-A^* - Q^*\Sigma^{*-1}\right]x_r^* + \xi^* \tag{4.99}$$

where ξ^* is a white noise process with covariance Q^*, $E\, x_r^*(T)x_r^{*\prime}(T) = \Sigma^*(T)$, and Σ^* is given by

$$\frac{d}{dt}\Sigma^* = A^*\Sigma^* + \Sigma^*A^{*\prime} + Q^* \tag{4.100}$$

By analogy with the forward-time case, there exists a reduced-order reversed-time Kalman filter having gain K_r^* and producing the estimate $\hat{x}_r^*$. The reversed-time augmented system is

$$-\frac{d}{dt}\begin{bmatrix}x_r\\ \hat{x}_r^*\end{bmatrix} = \begin{bmatrix}-A-Q\Sigma^{-1} & 0\\ K_r^*C & -A^*-Q^*\Sigma^{*-1}-K_r^*C^*\end{bmatrix}\begin{bmatrix}x_r\\ \hat{x}_r^*\end{bmatrix} + \begin{bmatrix}\xi\\ K_r^*v\end{bmatrix} \tag{4.101}$$

Let the covariance be

$$\begin{bmatrix} \Sigma & M_r \\ M_r' & N_r \end{bmatrix} = E\left\{ \begin{bmatrix} x_r \\ \hat{x}_r^* \end{bmatrix} \begin{bmatrix} x_r' & \hat{x}_r^{*\prime} \end{bmatrix} \right\}$$

$$= E\left\{ \begin{bmatrix} x \\ \hat{x}_r^* \end{bmatrix} \begin{bmatrix} x_r' & \hat{x}_r^{*\prime} \end{bmatrix} \right\} \tag{4.102}$$

since x and x_r are stochastically indistinguishable. Then

$$-\frac{d}{dt} \begin{bmatrix} \Sigma & M_r \\ M_r' & N_r \end{bmatrix} = \begin{bmatrix} -A-Q\Sigma^{-1} & 0 \\ K_r^*C & -A^*-Q^*\Sigma^{*-1}-K_r^*C^* \end{bmatrix} \begin{bmatrix} \Sigma & M_r \\ M_r' & N_r \end{bmatrix} +$$

$$+ \begin{bmatrix} \Sigma & M_r \\ M_r' & N_r \end{bmatrix} \begin{bmatrix} -A-Q\Sigma^{-1} & 0 \\ K_r^*C & -A^*-Q^*\Sigma^{*-1}-K_r^*C^* \end{bmatrix}' + \begin{bmatrix} Q & \\ 0 & K_r^*RK_r^{*\prime} \end{bmatrix} \tag{4.103}$$

Cross-Correlation of $\hat{x}^*$ and $\hat{x}_r^*$ - The preceding analysis has shown how solving the two Lyapunov equations (4.98) and (4.103) yields the covariances of $\hat{x}^*$and $\hat{x}_r^*$ and their cross-correlations with x. Before one can obtain an expression for the smoothed error covariance, it is also necessary to know the cross-correlation of $\hat{x}^*$ and $\hat{x}_r^*$. The estimate $\hat{x}^*$ can be written in integral form from the variation of constants formula as

$$\hat{x}^*(t) = \int_o^t \Phi_{A^*-K^*C^*}(t,\sigma)K^*(\sigma)\ y(\sigma)\ d\sigma \tag{4.104}$$

where $\Phi_{A^*-K^*C^*}$ is the state transition matrix of the forward Kalman filter,

$$\frac{d}{dt}\Phi_{A^*-K^*C^*}(t,o) = \left[A^*(t) - K^*(t)C^*(t)\right]\ \Phi_{A^*-K^*C^*}(t,o) \tag{4.105}$$

with the identity initial condition $\Phi_{A^*-K^*C^*}(o,o) = I$. Similarly, the reversed-time estimate can be written

$$\hat{x}_r^*(t) = \int_T^t \Psi_{-A^*-Q^*\Sigma^{*-1}-K_r^*C^*}(t,\tau)K_r^*(\tau)\ y(\tau)\ d\tau \tag{4.106}$$

where

$$-\frac{d}{dt}\Psi_{-A^*-Q^*\Sigma^{*-1}-K_r^*C^*}(t,T) = \left[-A^*(t)-Q^*(t)\Sigma^{*-1}(t)\right]\Psi_{-A^*-Q^*\Sigma^{*-1}-K_r^*C^*}(t,T) \tag{4.107}$$

and $\Psi_{-A^*-Q^*\Sigma^{*-1}-K_r^*C^*}(T,T) = I$. Combining the integral expressions (4.104) and (4.106) and taking the expectation yields

$$E\ \hat{x}^*(t)\hat{x}_r^{*\prime}(t) = \int_o^t \int_T^t \Phi_{A^*-K^*C^*}(t,\sigma)K^*(\sigma)\ E\left[y(\sigma)y'(\tau)\right]\ K_t^{*\prime}(\tau)\cdot$$

$$\cdot\ \Psi'_{-A^*-Q^*\Sigma^{*-1}-K_r^*C^*}(t,\tau)\ d\tau d\sigma \tag{4.108}$$

The autocorrelation function of y is evaluated in the following lemma:

Lemma 4.1 For $\sigma \leq \tau$,

$$E\, y(\sigma)y'(\tau) = C(\sigma)\Psi_{-A-Q\Sigma^{-1}}(\sigma,t)\Sigma(t)\Phi_A'(\tau,t)C'(\tau) + R(\sigma)\delta(\sigma-\tau) \tag{4.109}$$

where $t \in [\sigma,\tau]$ and

$$\frac{d}{d\tau}\Phi_A(\tau,t) = A(\tau)\,\Phi_A(\tau,t) \quad , \quad \Phi_A(t,t) = I \tag{4.110}$$

$$-\frac{d}{d\sigma}\Psi_{-A-Q\Sigma^{-1}}(\sigma,t) = \left[-A(\sigma)-Q(\sigma)\Sigma^{-1}(\sigma)\right]\Psi_{-A-Q\Sigma^{-1}}(\sigma,t) \quad ,$$

$$\Psi_{-A-Q\Sigma^{-1}}(t,t) = I \tag{4.111}$$

i.e. Φ_A and $\Psi_{-A-Q\Sigma^{-1}}$ are the state transition matrices of the forward- and reversed-time systems, respectively.

proof: $E\, y(\sigma)y'(\tau) = E\left\{\left[C(\sigma)x(\sigma) + v(\sigma)\right]\left[C(\tau)x(\tau) + v(\tau)\right]'\right\}$

$$= C(\sigma)\, E\left\{x(\sigma)x'(\tau)\right\} C'(\tau) + E\left\{v(\sigma)v'(\tau)\right\}$$

$$= C(\sigma)\, E\left\{\left[\Psi_{-A-Q\Sigma^{-1}}(\sigma,t)\, x(t)\right]\left[\Phi_A(\tau,t)x(t)\right]'\right\} C'(\tau) + R(\sigma)\delta(\sigma-\tau)$$

$$= C(\sigma)\Psi_{-A-Q\Sigma^{-1}}(\sigma,t)\, \Sigma(t)\, \Phi_A'(\tau,t)\, C'(\tau) + R(\sigma)\delta(\sigma-\tau)$$

Q.E.D.

Equation (4.109) can now be substituted into (4.108). Note that the term involving the delta function drops out because of the limits on the double integral. Thus

$$E\,\hat{x}^*(t)\hat{x}_r^{*'}(t) = \int_o^t \int_t^T \Phi_{A^*-K^*C^*}(t,\sigma)K^*(\sigma)C(\sigma)\Psi_{-A-Q\Sigma^{-1}}(\sigma,t)\Sigma(t)\cdot$$

$$\cdot\,\Phi'_A(\tau,t)C'(\tau)K_r^{*'}(\tau)\,\Psi'_{-A^*-Q^*\Sigma^{*-1}-K_r^*C^*}(t,\tau)\,d\tau d\sigma \qquad (4.112)$$

This result is now expressed as

Lemma 4.2 $$E\,\hat{x}^*(t)\hat{x}_r^{*'}(t) = \alpha^*(t)\Sigma(t)\beta^*(t) \qquad (4.113)$$

where

$$\alpha^*(t) = \int_o^t \Phi_{A^*-K^*C^*}(t,\sigma)K^*(\sigma)C(\sigma)\Psi_{-A-Q\Sigma^{-1}}(\sigma,t)\,d\sigma \qquad (4.114)$$

$$\beta^*(t) = \int_t^T \Phi'_A(\tau,t)C'(\tau)K_r^{*'}(\tau)\Psi'_{-A^*-Q^*\Sigma^{*-1}-K_r^*C^*}(t,\tau)\,d\tau \qquad (4.115)$$

proof: Lemma 4.2 is an immediate consequence of (4.112).

The integral expressions for α^* and β^* may be replaced by differential equations.

Lemma 4.3 $$\frac{d}{dt}\alpha^*(t) = \left[A^*-K^*C^*\right]\alpha^*(t) + \alpha^*(t)\left[-A-Q\Sigma^{-1}\right] + K^*C \qquad (4.116)$$

$$-\frac{d}{dt}\beta^*(t) = \left[A\right]'\beta^*(t) + \beta^*(t)\left[-A^*-Q^*\Sigma^{*-1}-K_r^*C^*\right]' + C'K_r^{*'} \qquad (4.117)$$

with initial conditions $\alpha^*(o) = 0$ and $\beta^*(T) = 0$.

proof: Differentiating (4.114) with respect to t yields

$$\frac{d}{dt}\alpha^*(t) = \int_o^t \frac{d}{dt}\left\{\Phi_{A^*-K^*C^*}(t,\sigma)K^*(\sigma)C(\sigma)\Psi_{-A-Q\Sigma^{-1}}(\sigma,t)\right\}d\sigma + K^*(t)C(t)$$

$$= \int_o^t \left\{\left[A^*(t)-K^*(t)C^*(t)\right]\Phi_{A^*-K^*C^*}(t,\sigma)K^*(\sigma)C(\sigma)\Psi_{-A-Q\Sigma^{-1}}(\sigma,t) + \right.$$

$$\left. + \Phi_{A^*-K^*C^*}(t,\sigma)K^*(\sigma)C(\sigma)\Psi_{-A-Q\Sigma^{-1}}(\sigma,t)\left[-A(t)-Q(t)\Sigma^{-1}(t)\right]\right\}d\sigma$$

$$+ K^*(t)C(t)$$

$$= \left[A^*(t)-K^*(t)C^*(t)\right]\alpha^*(t) + \alpha^*(t)\left[-A(t)-Q(t)\Sigma^{-1}(t)\right] + K^*(t)C(t)$$

Equation (4.117) is obtained in a completely analogous fashion.

Q.E.D.

The cross-correlation between $\hat{x}^*(t)$ and $\hat{x}^*_r(t)$ is given by the relatively simpl expression (4.113) of Lemma 4.2 where α^* and β^* obey differential equations of the Lyapunov type. The coefficients of α^* and β^* in these equations are

- $A^*-K^*C^*$: forward-time filter matrix
- $-A-Q\Sigma^{-1}$: reversed-time system matrix
- A' : forward-time system matrix transposed
- $\left[-A^*-Q^*\Sigma^{*-1}-K^*_rC^*\right]'$: reversed-time filter matrix transposed.

Notice the symmetry. In the linear time-invariant infinite-lag case, all four of these matrices are stability matrices. The steady-state algebraic versions of (4.116) and (4.117) will, therefore, always have unique solutions.

Everything necessary for the evaluation of the smoothed output error covariance is now available. The next theorem puts it all together.

Theorem 4.4 The error covariance of the reduced-order smoothed output estimate is given by

$$\mathrm{cov}\left[z(t)-\hat{z}_s^*(t)\right] = H\Sigma H' - H^*\Sigma_s^*\left[P^{*-1}M' + P_r^{*-1}M_r'\right]H' - H\left[MP^{*-1} + M_rP_r^{*-1}\right]\Sigma_s^*H^{*'} +$$

$$+ H^*\Sigma_s^*\left[P^{*-1}NP^{*-1} + P_r^{*-1}\beta^{*'}\Sigma\alpha^{*'}P^{*-1} + P^{*-1}\alpha^*\Sigma\beta^*P_r^{*-1} + P_r^{*-1}N_rP_r^{*-1}\right]\Sigma_s^*H^{*'} \tag{4.118}$$

where Σ, M, N come from (4.97); M_r and N_r come from (4.102); and α^* and β^* come from (4.116) and (4.117).

proof: $\mathrm{cov}\left[z(t) - \hat{z}_s^*(t)\right] = \mathrm{cov}\left[H\,x - H^*\hat{x}_s^*\right]$

$$= \begin{bmatrix} H & -H^* \end{bmatrix} E\left\{\begin{bmatrix} x \\ \hat{x}_s^* \end{bmatrix}\begin{bmatrix} x' & \hat{x}_s^{*'} \end{bmatrix}\right\}\begin{bmatrix} H' \\ -H^{*'} \end{bmatrix}$$

$$= H\,E\left\{x\,x'\right\}H' - H^*\,E\left\{\hat{x}_s^*x'\right\}H' - H\,E\left\{x\,\hat{x}_s^{*'}\right\}H^{*'} +$$

$$+ H^*\,E\left\{\hat{x}_s^*\hat{x}_s^{*'}\right\}H^{*'}$$

(I.) $$E\left\{x\ x'\right\} = \Sigma$$

(II.) $$E\left\{\hat{x}_s^* x'\right\} = \Sigma_s^* E\left\{\left[P^{*-1}\hat{x}^* + P_r^{*-1}\hat{x}_r^*\right]x'\right\}$$

$$= \Sigma_s^*\left[P^{*-1}M' + P_r^{*-1}M_r'\right] \qquad \text{from (4.97) and (4.102)}$$

(III.) $$E\left\{\hat{x}_s^*\hat{x}_s^{*\prime}\right\} = \Sigma_s^* E\left\{\left[P^{*-1}\hat{x}^* + P_r^{*-1}\hat{x}_r^*\right]\left[P^{*-1}\hat{x}^* + P_r^{*-1}\hat{x}_r^*\right]'\right\}\Sigma_s^*$$

$$= \Sigma_s^*\left[P^{*-1}NP^{*-1} + P_r^{*-1}\beta^{*\prime}\Sigma\alpha^{*\prime}P^{*-1} + P^{*-1}\alpha^*\Sigma\beta^*P_r^{*-1} + P_r^{*-1}N_rP_r^{*-1}\right]\Sigma_s^*$$

from (4.97), (4.102) and (4.113)

$$\therefore \operatorname{cov}\left[z-\hat{z}_s^*\right] = H\Sigma H' - H^*\Sigma_s^*\left[P^{*-1}M' + P_r^{*-1}M_r'\right]H' - H\left[MP^{*-1} + M_rP_r^{*-1}\right]\Sigma_s^*H^{*\prime} +$$

$$+ H^*\Sigma_s^*\left[P^{*-1}NP^{*-1} + P_r^{*-1}\beta^{*\prime}\Sigma\alpha^{*\prime}P^{*-1} + P^{*-1}\alpha^*\Sigma\beta^*P_r^{*-1} + P_r^{*-1}N_rP_r^{*-1}\right]\Sigma_s^*H^{*\prime}$$

Q.E.D.

This theorem is the main result of this section. Certainly (4.118) is a messy expression for the error covariance. The only difficult part of the derivation, however, was finding an expression for $E\ \hat{x}^*\hat{x}_r^{*\prime}$. All the other steps were quite straight-forward.

Before leaving the reduced-order smoother for the sensitivity analysis, notice that the cross-correlation between $\hat{x}^*$ and x could have been evaluated in an analogous fashion to the way (4.112) was obtained. This yields

$$E\,\hat{x}^*(t)x'(t) = \int_0^t \Phi_{A^*-K^*C^*}(t,\sigma)K^*(\sigma)C(\sigma)\Psi_{-A-Q\Sigma^{-1}}(\sigma,t)\Sigma(t)\,dt$$

$$= \alpha^*\Sigma \tag{4.119}$$

But $E\,\hat{x}^*(t)x'(t) = M'(t)$ from (4.97). Therefore,

$$\alpha^* = M'\Sigma^{-1} \tag{4.120}$$

Similarly,

$$\beta^* = \Sigma^{-1}M_r \tag{4.121}$$

Since the Lyapunov equations (4.98) and (4.103) have to be solved and yield M and M_r, (4.120) and (4.121) allow the evaluation of the output covariance without the solution of the additional equations (4.116) and (4.117) for α^* and β^*.

The solution of the sensitivity analysis problem is easily obtained from Theorem 4.4 by simply setting H and H^* equal to the identity.

<u>Corollary 4.1</u> $\quad \mathrm{cov}\left[x(t) - \hat{x}^*(t)\right] = \Sigma - \Sigma_s^*\left[P^{*-1}M' + P_r^{*-1}M_r'\right] +$

$$- \left[MP^{*-1} + M_rP_r^{*-1}\right]\Sigma_s^* +$$

$$+ \Sigma_s^*\left[P^{*-1}NP^{*-1} + P_r^{*-1}\beta^{*\prime}\Sigma\alpha^{*\prime}P^{*-1} + P^{*-1}\alpha^*\Sigma\beta^*P_r^{*-1} + P_r^{*-1}N_rP_r^{*-1}\right]\Sigma_s^* \tag{4.122}$$

The expression for the smoothed error covariance could also be obtained by expressing the smoothed error as a linear combination of three errors -- forward error, reversed-time error, and *a priori* error,

$$x - \hat{x}_s^* = \Sigma_s^* \left[P^{*-1}(x - \hat{x}^*) + P_r^{*-1}(x - \hat{x}_r^*) - \Sigma^{*-1}(x) \right] \qquad (4.123)$$

The covariances of these errors and their cross-correlations can be found from (4.98), (4.103), and (4.113). By making these substitutions and performing some tedious algebraic manipulations, one is able to arrive at (4.122) from (4.123). The details are omitted.

In summary, this section has addressed the problem of FI smoothing using an incorrect model. In the case of a reduced-order smoother, the actual smoothed output error covariance is given in Theorem 4.4. A special case of this result is the senstivity analysis expression (4.122). To use either equation, it is necessary to solve the forward- and reversed-time Lyapunov equations (4.98) and (4.103). The quantities α^* and β^* also obey Lyapunov equations, but can be computed (perhaps more conveniently) from (4.120) and (4.121).

4.5 Conclusions

The two-filter smoother expresses the smoothed state estimate as a linear combination of two optimal estimates. One of the main contributions of this chapter has been to obtain this smoother from first principles. Other derivations of the two-filter smoother have proceeded by showing equivalence with some other smoothing algorithms. Because of these derivations, it has never been clear exactly what type of estimate the backwards estimate $\hat{x}_b(t)$ is. Therefore, perhaps more important than the derivation of a new two-filter smoother in this chapter is the in-

sight gained from this approach. The backwards estimate is simply the maximum likelihood estimate. The backwards filter comes from removing the *a priori* information from a reversed-time Kalman filter. This reversed-time Kalman filter, a key element throughout the chapter, is designed from a reversed-time realization of the state process. Other authors, e.g. [74], have used the reversed-time model to obtain smoothing formulas, but these results essentially just applied standard smoothing formulas to the reversed-time model. Section 4.3 used the reversed-time filter in conjunction with the forward filter to obtain the resulting expression for the smoothed estimate. It should be noted that some of the equations in Section 4.3 are quite similar to ones obtained by Ljung and Kailath [65] by using the relationship between linear least-squares estimation and scattering theory. The approach taken here seems to be a much more natural one for addressing the smoothing problem and yields the very simple change of initial conditions formula (4.82) and (4.83) for $\hat{x}_s(o)$.

The smoothing formulas presented here are symmetric with respect to forward-time vs. reversed-time. This is not meant to imply that the two filter error covariances $P(t)$ and $P_r(t)$ are equal, but rather the form of the smoother is the same for both the past and the future. For example, the two estimates $\hat{x}(t|t)$ and $\hat{x}_r(t|t)$ that are combined to produce the smoothed estimate are both conditional expectations of $x(t)$. As discussed in Section 4.1, intuitively there is an equivalence between past and future observations. Where the difference between forward- and

reversed-time become apparent was in Section 4.3.6 when the question of uncertain initial covariance was considered. The reversed-time system matrix $-A(t)-Q(t)\Sigma^{-1}(t)$ obviously depends on the state covariance; the forward system matrix $A(t)$, of course, is independent of Σ. Therefore, when considering change of initial covariance problems, the fact that the original system model is given in forward-time introduces a distinction between forward- and reversed-time. Using the forward model, one can form a maximum likelihood estimate of $x(o)$ which can be combined with the *a priori* data to provide a change of initial condition formula for the smoothed estimate $\hat{x}_s(o)$. There does not exist an analogous formula for the filtered estimate $\hat{x}(t|t)$ because the reversed-time system matrix is a function of the state covariance.

In order to implement the two-filter smoother given in Theorem 4.3, it is very convenient to use the information filter form of the forward- and reversed-time Kalman filters. This means one should compute $P^{-1}(t|t)$, $P_r^{-1}(t|t)$, $P^{-1}(t|t)\hat{x}(t|t)$, and $P_r^{-1}(t|t)\hat{x}_r(t|t)$ instead of the usual Kalman filter estimate and covariance. If these quantities are available, then only one matrix inversion is needed in the computation of the smoothed estimate -- the inverse of $\left[P^{-1}(t|t) + P_r^{-1}(t|t) - \Sigma^{-1}(t)\right]$ is all that is required.

The final contribution of this chapter is the analysis of reduced-order smoothers and the sensitivity of two-filter smoothers. The approach taken here is similar to that of Mehra [43] except that Mehra erroneously assumed the forward and backward filtered errors were uncorrelated. Hence the main contribution of Section 4.4 is Lemma 4.2 which gives the

cross-correlation between $\hat{x}^*(t)$ and $\hat{x}_r^*(t)$, the forward- and reversed-time estimates. Another important aspect of the analysis in Section 4.4 is that both Lyapunov equations (4.98) and (4.103) can correspond to stable systems. In particular, for the time-invariant infinite-lag problem, the forward- and reversed-time augmented systems are both stable.

CHAPTER 5

CONTROL AND ESTIMATION FOR TOEPLITZ SYSTEMS

5.1 Introduction to Toeplitz Systems

5.1.1 Definition

Toeplitz systems are infinite-dimensional, spatially-invariant linear systems. They are composed of an infinite number of identical subsystems and are the natural infinite-dimensional analog of circulant systems. Primary attention will be paid to stationary discrete-time Toeplitz systems, although continuous-time systems will also be used on occasion.

The dynamics of the k^{th} subsystem in a deterministic Toeplitz system are given by

$$x_k(i+1) = \sum_{\ell=-\infty}^{+\infty} A_{k-\ell}x_\ell(i) + B_{k-\ell}u_\ell(i) \qquad (5.1)$$

where $k=0,\pm1,\ldots$. The state of the k^{th} subsystem at time i is $x_k(i)\in\mathbb{R}^n$, and $u_k(i)\in\mathbb{R}^m$ is the local control. The k^{th} output is

$$y_k(i) = \sum_{\ell=-\infty}^{+\infty} C_{k-\ell}x_\ell(i) \qquad (5.2)$$

where $k=0,\pm1,\ldots$ and $y_k(i)\in\mathbb{R}^p$. For notational simplicity, the infinite-dimensional state, input, and output vectors are defined,

$$x(i) = \begin{pmatrix} \vdots \\ x_{-1}(i) \\ x_0(i) \\ x_1(i) \\ \vdots \end{pmatrix}, \quad u(i) = \begin{pmatrix} \vdots \\ u_{-1}(i) \\ u_0(i) \\ u_1(i) \\ \vdots \end{pmatrix}, \quad y(i) = \begin{pmatrix} \vdots \\ y_{-1}(i) \\ y_0(i) \\ y_1(i) \\ \vdots \end{pmatrix}$$

In terms of these vectors, the infinite set of equations represented by (5.1) and (5.2) are written as

$$x(i+1) = A\,x(i) + B\,u(i) \tag{5.3}$$

$$y(i) = C\,x(i) \tag{5.4}$$

where the Toeplitz system matrix is

$$A = \begin{pmatrix} \ddots & \vdots & & & \vdots & \iddots \\ & A_0 & A_{-1} & A_{-2} & A_{-3} & \\ & A_1 & A_0 & A_{-1} & A_{-2} & \\ & A_2 & A_1 & A_0 & A_{-1} & \\ & A_3 & A_2 & A_1 & A_0 & \\ \iddots & \vdots & & & \vdots & \ddots \end{pmatrix}$$

The Toeplitz input and output matrices B and C are of the same form. The mutual interaction between subsystems k and ℓ depends only on $k-\ell$, as can be seen from (5.1). It is in this sense that Toeplitz systems are called

spatially-invariant.

The z-transform can be used to decouple Toeplitz systems in very much the same way the DFT was used to decouple circulant systems in Chapter 2. The z-transform of the state vector x(i) is

$$x(i,z) = \mathcal{Z}[x_k(i)] = \sum_{k=-\infty}^{+\infty} x_k(i) z^{-k} \tag{5.5}$$

The transforms y(i,z) and u(i,z) are defined similarly. The transform of the system matrix is

$$A(z) = \mathcal{Z}[A_k] = \sum_{k=-\infty}^{+\infty} A_k z^{-k} \tag{5.6}$$

The $n \times m$ matrix B(z) and the $p \times n$ matrix C(z) are defined similarly from B and C.

The system dynamics (5.1) and output (5.2) equations can be rewritten in the z domain. Using the property of the z-transform for convolution sums [80] yields

$$x(i+1,z) = A(z)\ x(i,z) + B(z)\ u(i,z) \tag{5.7}$$

$$y(i,z) = C(z)\ x(i,z) \tag{5.8}$$

These equations indicate that the Toeplitz system is composed of indepen-

dent subsystems, where the subsystems are indexed by the complex variable z. Compare with (2.17) and (2.18), the subsystem dynamics and output for circulant systems. There is an important difference between Toeplitz and circulant systems, however. The Toeplitz subsystems are indexed by a continuous complex variable; the circulant subsystem index has only a finite number of values.

In the case of an unforced system, the Toeplitz matrix A is an operator mapping the sequence $\{x_k(i)\}$ into the sequence $\{x_k(i+1)\}$. For the operator A to be a bounded operator on the space of square summable sequences, the induced norm of A must be finite. Widom [81] has shown that the operator norm of A is related to the z-transform A(z) by

$$||A|| = \operatorname*{ess\,sup}_{z \in U} ||A(z)||_2 \tag{5.9}$$

where the set U is the unit circle in the complex plane. Since only bounded operators A are of interest as the system matrix for a Toeplitz system, it is assumed that the z transform of A does exist and that the region of convergence includes the unit circle. The region of convergence of A(z) is an annulus in the complex plane consisting of all z for which the defining sum (5.7) is absolutely convergent. Likewise, the transforms B(z) and C(z) are also assumed to exist and to have regions of convergence containing the unit circle. Melzer and Kuo [30] and Chu [31] have studied these systems and mistakenly claim that a sufficient condition for the existence of (say) A(z) is that $\lim_{k \to +\infty} A_k = 0$.

Of course this condition is neither necessary nor sufficient; what does suffice is that the sequence geometrically approach zero in at least one direction.

A further assumption that will be useful throughout this chapter is that all the z-transforms of the matrices used in the Toeplitz model are rational functions of z. This assumption is not necessary for much of the development, and it will be explicitly noted when the rationality assumption is used. What this assumption allows, however, is the construction of an efficient procedure for optimal linear filtering or optimal control. In summary,

Assumption 5.1. All the z-transforms of matrices used in Toeplitz models are assumed to exist, to have regions of convergence including the unit circle, and to be rational functions of the complex variable z.

Survey of Chapter 5 - The remainder of this section will present the results of Melzer and Kuo [30] and Chu [31] on the optimal control of continuous-time Toeplitz systems. Their work deals with the use of the z-transform for the off-line design of centralized and decentralized feedback controllers. Section 5.2 discusses Attasi's [28] work in recursive processing of noisy images. Attasi's model and filtering solution are presented and related to Toeplitz systems. This solution employs the z-transform to obtain an efficient on-line implementation of the filter. The estimation problem for general Toeplitz systems is then posed and solved in Section 5.3. The goal of this section is a filter which has an

efficient on-line implementation. The key to obtaining such an efficient processor is the use of the spatial dynamics of the Toeplitz systems as they appear in the spatial z-transform. The implications of this result for filtering of large-scale systems and some implementation issues of such a filter are also covered in Section 5.3. The optimal control problem is addressed in Section 5.4 as the dual of the Toeplitz estimation problem. For Toeplitz systems having a block diagonal input matrix, i.e., $B_k = B_o \delta_{k,o}$, the optimal control for each subsystem is expressed in a novel and interesting fashion. Section 5.5 contains some concluding remarks about the chapter.

5.1.2 Optimal Control via z-Transforms

Motivated by the infinite string of vehicles problem (see 2.1.3), Melzer and Kuo [30] investigated the optimal regulator problem for Toeplitz systems. They considered a quadratic performance index and obtained the centralized, full-state feedback solution. This solution was arrived at by using the z-transform to decompose the original problem into decoupled lower-order problems indexed by z.

Melzer and Kuo consider the continuous-time version of the dynamics (5.1),

$$\frac{d}{dt} x_k(t) = \sum_{i=-\infty}^{+\infty} A_{k-i} x_i(t) + B_{k-i} u_i(t) \qquad (5.10)$$

The optimal control problem is to find the sequence of control vectors which minimizes the performance index

$$J = \frac{1}{2} x'(T)F\, x(T) + \frac{1}{2} \int_0^T x'(t)Q\, x(t) + u'(t)R\, u(t)\, dt \quad (5.11)$$

for arbitrary initial conditions x(o) where F, Q, R are Toeplitz matrices. The notation $x'(t)Q\, x(t)$ is defined by

$$x'(t)Q\, x(t) = \sum_{i=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} x_i'(t) Q_{i-k} x_k(t) \quad (5.12)$$

Also, it is assumed that $F \geq 0$, $Q \geq 0$, and $R > 0$, i.e. for example

$$\sum_{i=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} x_i'\, F_{i-k}\, x_k \geq 0 \quad (5.13)$$

for all sequences $\{x_k\}$. By assumption 5.1, the matrices F, Q, R are assumed to have rational z-transforms.

This problem is the infinite-dimensional version of the standard linear-quadratic regulator problem. Using the maximum principle, Melzer and Kuo obtain a solution which is formally the same as the linear-quadratic regulator solution. In the z-domain, the optimal control at time t is just

$$u(t,z) = G(t,z)\, x(t,z) \quad (5.14)$$

when the feedback gain G(t,z) is given by

$$G(t,z) = -R^{-1}(z)B'(z^{-1})K(t,z) \quad (5.15)$$

K(t,z) is obtained from a Riccati equation,

$$\frac{d}{dt} K(t,z) = -K(t,z)A(z) - A'(z^{-1})K(t,z) - Q(z)$$

$$-K(t,z)B(z)R^{-1}(z)B'(z^{-1})K(t,z) \qquad (5.16)$$

where K(T,z) = F(z). Taking the inverse transform yields the optimal controller in the spatial or index domain,

$$u(t) = G(t)\ x(t)$$

$$= -R^{-1}B'\ K(t)\ x(t) \qquad (5.17)$$

where G(t) and K(t) are block Toeplitz matrices whose elements are given by the inverse z-transform of G(t,z) and K(t,z), respectively. The matrix K(t) is symmetric and positive semi-definite just as in the finite-dimensional case.

The solution of the Toeplitz regulator problem is, from the Riccati equation (5.16), decomposed into the solution of lower-order, independent subproblems. Each subinput, however, depends upon all the substates,

$$u_k(t) = \sum_{i=-\infty}^{+\infty} G_{k-i}(t)\ x_i(t) \qquad (5.18)$$

That is, this solution requires that the information of all the states be available to every subsystem. The efficient implementation of this operation will be discussed later. Chu [31] moved beyond this centralized problem and considered optimal regulation when the individual sub-

inputs were only a function of some "local" information.

In Chu's decentralized regulator, the information available to the k^{th} controller is assumed to be only the k^{th} suboutput $y_k(t)$. For example, if the information structure were knowledge of the state of the local subsystem and the two nearest neighbors, then

$$y_k(t) = \begin{pmatrix} x_{k-1}(t) \\ x_k(t) \\ x_{k+1}(t) \end{pmatrix} \tag{5.19}$$

i.e. $C_i = I$ for $i=0,1,-1$ and zero otherwise. The control is constrained to be a linear, nondynamic function of y_k,

$$u_k(t) = G_o y_k(t) \tag{5.20}$$

The design problem is to determine the $n \times p$ matrix G_o, thereby specifying the controller for all the subsystems. Chu considers the infinite-horizon version of the cost (5.1),

$$J = \frac{1}{2} \int_o^\infty x'(t) Q\, x(t) + u'(t) R\, u(t)\, dt \tag{5.21}$$

and proposes choosing G_o to minimize this cost. Note that even for controllable systems, there is no guarantee that a minimizing or stabilizing G_o exists because of the constraints imposed on the control system.

The closed-loop system is still Toeplitz and has dynamics

$$\frac{d}{dt} x(t) = [A + BGC]\, x(t) \tag{5.22}$$

or

$$x(t) = \exp\left\{[A + BGC]t\right\} x(o) \tag{5.23}$$

where G is a block diagonal Toeplitz matrix with diagonal element G_o. In the transformed domain,

$$x(t,z) = \exp\left\{[A(z) + B(z)G_oC(z)]t\right\} x(o,z) \tag{5.24}$$

The transformed version of the cost is

$$J = \frac{1}{2}\int_o^\infty <x'(t,z^{-1})Q(z)x(t,z)>_o + <u'(t,z^{-1})R(z)u(t,z)>_o \, dt \tag{5.25}$$

where $<\ >_o$ denotes the zeroth element of the inverse z-transform, i.e.

$$<x(t,z)>_o = \frac{1}{2\pi j}\oint x(t,z)z^{-1}\, dz \tag{5.26}$$

Substituting (5.24) into the expression for the cost yields

$$J = \frac{1}{2}\int_o^\infty <x'(t,z^{-1})\left\{Q(z) + C'(z^{-1})G_o' R(z)G_oC'(z)\right\}x(t,z)>_o \, dt$$

$$= \frac{1}{2}\int_o^\infty <x'(o,z^{-1})\exp\left\{[A(z^{-1}) + B(z^{-1})G_oC(z^{-1})]'t\right\}\left\{M(z)\right\} x$$

$$x \exp\left\{[A(z) + B(z)G_oC(z)]t\right\} x(o,z)>_o \, dt \tag{5.27}$$

where $M(z) = Q(z) + C'(z^{-1})G_o' R(z)G_oC(z)$. Interchanging integration with respect t and z and using the vector identity $x'y = tr(yx)$ gives

$$J = <\frac{1}{2}\, tr\left\{\int_o^\infty \exp[A_{CL}(z^{-1})t]'\, M(z)\exp[A_{CL}(z)t]x(o,z)x'(o,z)\,dt\right\}>_o \tag{5.28}$$

where $A_{CL}(z) = A(z) + B(z)G_oC(z)$ is the transformed closed-loop system matrix.

The minimization of J over G_o produces a result which from (5.28) will in general depend upon x(o,z). In order to eliminate this dependence, the initial disturbances are assumed to be random variables with zero mean and transformed covariance X(o,z). (see [55]). The optimization problem, then, is to choose G_o to minimize

$$J = <\frac{1}{2}\, tr \int_o^\infty \exp[A_{CL}(z^{-1})'t]M(z)\ \exp[A_{CL}(z)t]X(o,z)\ dt> \quad (5.29)$$

Chu gives the following necessary condition for the optimal feedback gain G_o:

$$<R(z)G_oC(z)P(z)C'(z^{-1})>_o + <B'(z^{-1})K(z)P(z)C'(z^{-1})>_0 = 0 \quad (5.30)$$

where K(z) and P(z) are given by

$$K(z)A_{CL}(z) + A_{CL}'(z^{-1})K(z) + M(z) = 0 \quad (5.31)$$

$$P(z)A_{CL}'(z^{-1}) + A_{CL}(z)P(z) + X(o,z) = 0 \quad (5.32)$$

Note that A_{CL} and M are functions of G_o, so that this necessary condition is a coupled set of nonlinear equations. The proof of this result is similar to the proof of the circulant case discussed in Section 3.2. The details may be found in [31].

This necessary condition for the optimal decentralized regulator

is not nearly as nice as the solution given by Melzer and Kuo for the centralized regulator. The z-transform does not decompose the decentralized problem. In both cases, it should be realized, the solutions must be obtained analytically and not numerically. This is because the indexing variable z is a continuous variable. This requirement makes the solution of either problem so difficult that the only examples published in the literature are for first and second order subsystems. Furthermore, the on-line aspects of these solutions have been neglected. Thus these solutions to the centralized and decentralized regulator problems are somewhat less than satisfactory. In particular, it is still of interest to determine an efficient solution of the regulator problem that is applicable to as general a class of Toeplitz systems as is possible.

Analogous results can be obtained for the discrete-time centralized and decentralized regulator problems. Also, by duality, the linear least-squares estimation problem for Toeplitz systems can be treated similarly. Attasi [28] has treated the estimation problem for a special type of Toeplitz system. The next section discusses Attasi's work with emphasis on his efficient filtering algorithm. With the insight gained from this algorithm, the general Toeplitz estimation problem is solved. The efficient solution of the optimal control problem is then obtained by duality.

5.2 Attasi's Work in Image Processing

The z-transform has been shown to be a vehicle for the off-line solution of linear-quadratic optimal control problems. The work of Melzer and Kuo discussed in 5.1.2 was directed toward the computation of optimal feedback gains via the z-transform. This transformation is also useful for considering the on-line solution of these problems. Attasi [28] has used z-transform techniques to obtain an efficient algorithm for filtering in a very special type of Toeplitz system. The overall objective of this chapter is to use the z-transform in both parts of control and estimation problems. That is, the transform domain will be used to obtain, off-line, a solution which has an efficient on-line implementation.

There has been much work in the image processing field directed toward the recursive estimation of discretized images from observations corrupted by additive noise [25]. In the search for computationally tractable estimation formulas, the use of two-parameter models for images has proved helpful [27]. Of primary interest in this section is the two-parameter model introduced by Attasi [28]. This model will be shown to give rise to a very efficient recursive estimation procedure.

Attasi has considered least-squares estimation of an image $z(i,j)$ under noisy observations

$$y(i,j) = z(i,j) + v(i,j) \tag{5.33}$$

where the image is generated by

$$z(i,j) = H\, x(i,j) \tag{5.34}$$

$$x(i,j) = F_1 x(i-1,j) + F_2 x(i,j-1) - F_1 F_2 x(i-1,j-1) + w(i-1,j) \tag{5.35}$$

with the restriction that $F_1 F_2 = F_2 F_1$. Here $x(i,j) \in \mathbb{R}^n$, $y(i,j) \in \mathbb{R}^p$ and $w(i,j)$ and $v(i,j)$ are independent, zero-mean, Gaussian white noise processes with covariances Q and R, respectively. Equations (5.33)-(5.35) are a stochastic realization of the doubly indexed sequence of vectors $\{y(i,j)\}$.

The estimation problem is to compute the estimate $\hat{x}(i,j)$ of $x(i,j)$ given the observations $y(k,\ell)$ for $k \leq i$ and all ℓ, i.e.

$$\hat{x}(i,j) = E\left\{x(i,j) \,\middle|\, y(k,\ell),\ k \leq i \text{ and } \ell = 0, \pm 1, \ldots\right\} \tag{5.36}$$

This is "line-by-line" filtering since the entire i^{th} line $\{x(i,j)\}_{j=0,\pm 1,\ldots}$ is estimated from the observations of all the lines to the left of line i. An estimate of the image $z(i,j)$ is then simply

$$\hat{z}(i,j) = H\, \hat{x}(i,j) \tag{5.37}$$

The filtered estimate $\hat{x}(i,j)$ is obtained by Attasi from a two step procedure analogous to discrete-time Kalman filtering. First, the predicted value $\tilde{x}(i,j)$ is defined as

$$\tilde{x}(i,j) = E\left\{x(i,j) \,\middle|\, y(k,\ell),\ k < i \text{ and } \ell = 0, \pm 1, \ldots\right\} \tag{5.38}$$

and may be computed from

$$\tilde{x}(i,j) = F_1 \hat{x}(i-1,j) \tag{5.39}$$

At this point, it is unclear why $\tilde{x}(i,j)$ is given by (5.39); this will become obvious shortly. The second step is the estimation of the error $e(i,j)$ in the predicted value,

$$e(i,j) = x(i,j) - \tilde{x}(i,j) \tag{5.40}$$

from the observations of line i. Once the estimate $\hat{e}(i,j)$ is available, the filtered estimate of $x(i,j)$ is just

$$\hat{x}(i,j) = \tilde{x}(i,j) + \hat{e}(i,j) \tag{5.41}$$

The remaining problem, therefore, is computing the estimate of the error from the observations $\{y(i,j)\}_{j=0,\pm1,\ldots}$ along line i. Define the innovations on line i as

$$I(i,j) = y(i,j) - H \tilde{x}(i,j) \tag{5.42}$$

Then it is immediate that

$$I(i,j) = H e(i,j) + v(i,j) \tag{5.43}$$

i.e. the innovations are noisy measurements of the errors. Thus the innovations along line i can be used to produce $\hat{e}(i,j)$ as the solution of a smoothing problem along the line. This smoothing problem is easily solved once the autocorrelation function $E\{e(i,\ell)\ e'(i,k)\}$ of the pre-

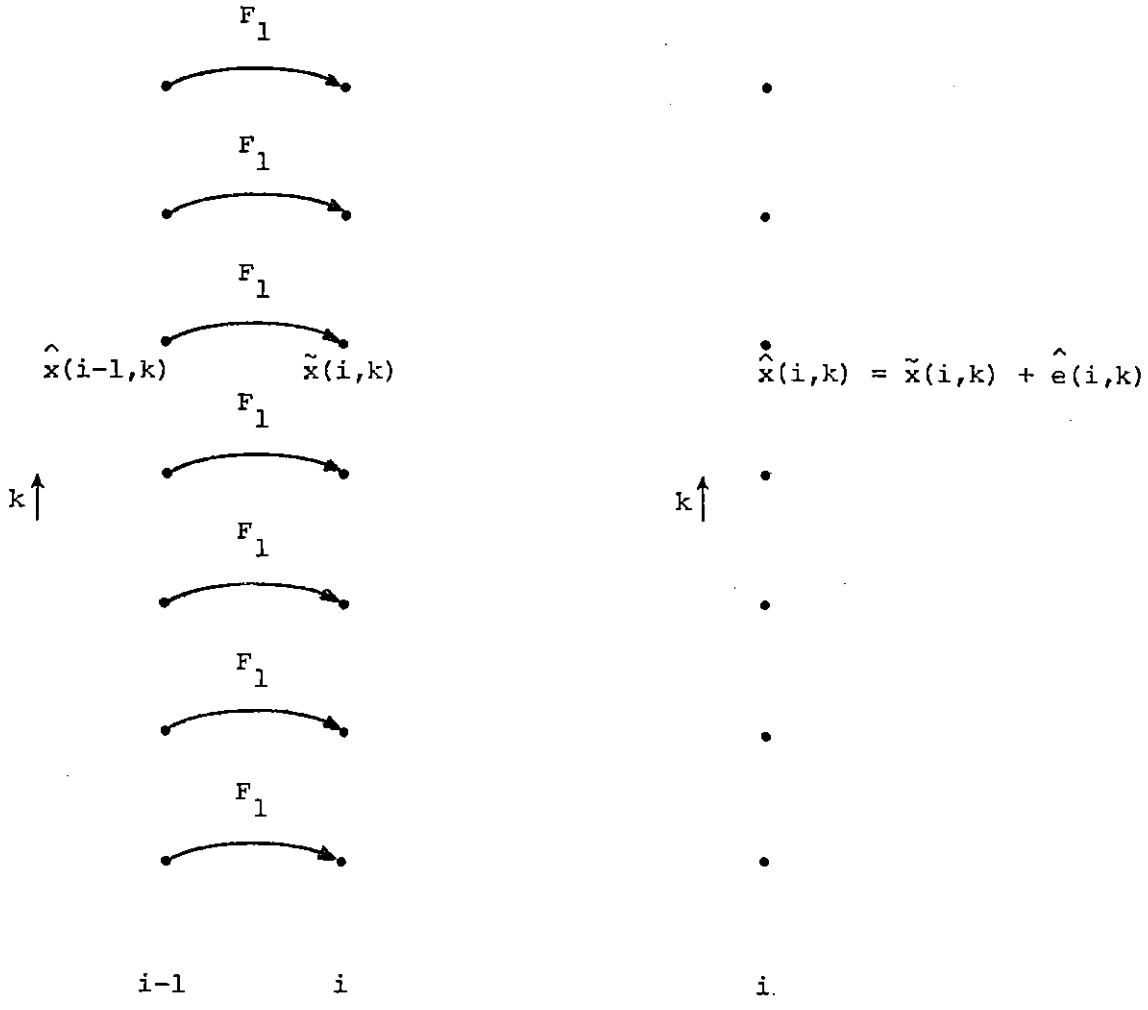
dicted error of line i has been determined. Attasi is able to specify the error autocorrelation function by finding a recursive equation for the spectrum of the error [28]. The details of using the spectrum to perform smoothing and thereby obtain $\hat{e}(i,j)$ are given in Section 5.3. The filtering algorithm is illustrated in Figure 5.1.

This two step filtering procedure, consisting of prediction by (5.39) and then smoothing, was developed by Attasi for the two-parameter model (5.33)-(5.35). The nature of filtering procedure, however, is suggestive of an infinite-dimensional one-parameter model -- infinite-dimensional because each line has infinite extent, one-parameter because the lines are handled one at a time. It will now be shown that (5.33)-(5.35) are equivalent to a Toeplitz system. In Section 5.3, Attasi's two step filtering procedure will be extended to a much wider class of Toeplitz systems.

Define the infinite-dimensional state and driving noise vectors as

$$x(i) = \begin{pmatrix} \vdots \\ x(i,-1) \\ x(i,0) \\ x(i,1) \\ \vdots \end{pmatrix} \quad , \quad w(i) = \begin{pmatrix} \vdots \\ w(i,-1) \\ w(i,0) \\ w(i,1) \\ \vdots \end{pmatrix} \tag{5.44}$$

Also, two Toeplitz matrices are defined,



a.) uncoupled prediction

b.) smoothing along line i yields $\hat{e}(i,k)$

Figure 5.1 Attasi's two step filtering procedure consists of predicting ahead from line i-1 to line i and then smoothing along line i.

$$T = \begin{pmatrix} \ddots & I & & \\ \ddots & & & \\ & -F_2 & I & \\ & & & \\ & F_2 & I & \ddots \\ & & \ddots & \end{pmatrix} \tag{5.45}$$

$$S = \begin{pmatrix} \ddots & F_1 & & \\ & & & \\ & & F_1 & \\ & & & \\ & & F_1 & \ddots \end{pmatrix} \tag{5.46}$$

Then it is claimed that the dynamics of Attasi's model may be written as

$$T\, x(i+1) = T\, S\, x(i) + w(i) \tag{5.47}$$

This is easily proved by writing out the j^{th} component of (5.47),

$$-F_2 x(i+1,j-1) + x(i+1,j) = -F_2 F_1\, x(i,j-1) + F_1 x(i,j) + w(i,j) \tag{5.48}$$

Recalling that F_1 and F_2 commute, it is easily seen that (5.48) is

identical to (5.35). Pre-Multiplying (5.47) by T^{-1} yields the Toeplitz system

$$x(i+1) = S\,x(i) + T^{-1}\,w(i) \tag{5.49}$$

where

$$T^{-1} = \begin{pmatrix} \ddots & & & \\ & I & & \\ & \ddots & \ddots & \\ & & F_2 & I \\ & & \ddots & \ddots \\ & & F_2^2 & F_2 \\ & & \ddots & \ddots \end{pmatrix} \tag{5.50}$$

Note that this requires that all the eigenvalues of F_2 lie inside the unit circle. Componentwise, the dynamics (5.49) are

$$x_k(i+1) = F_1 x_k(i) + \sum_{\ell=0}^{+\infty} F_2^{\ell}\, w_{k-\ell}(i) \tag{5.51}$$

Thus the two-parameter model used by Attasi can be considered to be a Toeplitz system. Also, from (5.49) it is clear why the prediction step is just (5.39) -- the system matrix is diagonal. Once the dynamics are written as (5.49), it is possible to use Attasi's filtering algorithm even if F_1 and F_2 do not commute.

Those Toeplitz systems which correspond to Attasi's two-parameter model are a very special class within the set of all Toeplitz systems. One might ask whether this class could be enlarged and still permit

filtering by prediction and smoothing. This is the subject of the next section.

5.3 The Estimation Problem

5.3.1 Formulation and Solution

Motivated by the work of Attasi on estimation for two-parameter models, the problem of obtaining minimum variance estimates for Toeplitz systems is addressed. The objective is to use z-transforms to determine a filter which has an efficient on-line implementation.

A stochastic discrete-time Toeplitz system is one of the form

$$x_k(i+1) = \sum_{\ell=-\infty}^{+\infty} A_{k-\ell} x_\ell(i) + D_{k-\ell} w_\ell(i) \tag{5.52}$$

$$y_k(i) = \sum_{\ell=-\infty}^{+\infty} C_{k-\ell} x_\ell(i) + v_k(i) \tag{5.53}$$

where $k = 0, \pm 1, \pm 2, \ldots$. Here the noises $w_\ell(i)$ and $v_\ell(i)$ are independent zero-mean Gaussian white noise processes,

$$E\left\{w_k(i) w_\ell'(j)\right\} = Q_o \, \delta_{i,j} \, \delta_{k,\ell} \tag{5.54}$$

$$E\left\{v_k(i) v_\ell'(j)\right\} = R_o \, \delta_{i,j} \, \delta_{k,\ell} \tag{5.55}$$

The initial state $x_k(0)$ is also assumed to be zero-mean and Gaussian and is independent of the driving and observation noises. In terms of the infinite-dimensional state vector x and observation vector y, (5.51) and (5.52) are written as

$$x(i+1) = A\,x(i) + D\,w(i) \tag{5.56}$$

$$y(i) = C\,x(i) + v(i) \tag{5.57}$$

where A, C, D are Toeplitz matrices. The interpretation here is that $x_k(i)$ is the state of subsystem k at time i.

The filtering problem is to obtain the minimum variance estimate $\hat{x}_k(i|i)$ of $x_k(i)$ given all the observations up to and including time i, $\{y_\ell(j) \mid 0 \leq j \leq i \text{ and all } \ell\}$. Under the condition of detectability for the system (5.56) and (5.57), Hager and Horowitz [82] have shown that the solution to this problem is a Kalman filter,

$$\hat{x}(i|i) = \hat{x}(i|i-1) + K(i)[y(i) - C\,\hat{x}(i|i-1)] \tag{5.58}$$

$$\hat{x}(i|i-1) = A\,\hat{x}(i-1|i-1) \tag{5.59}$$

$$K(i) = P(i|i-1)C'[C\,P(i|i-1)C' + R]^{-1} \tag{5.60}$$

$$P(i|i) = [I - K(i)C]P(i|i-1) \tag{5.61}$$

$$P(i|i-1) = A\,P(i-1|i-1)A' + DQD' \tag{5.62}$$

where the matrices $P(i|i)$ and $P(i|i-1)$ are covariance matrices of the filtered error $\hat{x}(i|i) - x(i)$ and the predicted error $\hat{x}(i|i-1) - x(i)$, respectively. Equations (5.58), (5.60), (5.61) define the measurement update step; the prediction step is given by (5.59) and (5.62)

If the initial state covariance matrix $P(0|-1)$ is Toeplitz, then

$K(i)$, $P(i|i)$, and $P(i|i-1)$ will be Toeplitz matrices for all i. Thus the Kalman filter can be written in the z-transform domain,

$$\hat{x}(i,z|i) = \hat{x}(i,z|i-1) + K(i,z)[y(i,z) - C(z)\hat{x}(i,z|i-1)] \quad (5.63)$$

$$x(i,z|i-1) = A(z)\ x(i-1,z|i-1) \quad (5.64)$$

$$K(i,z) = P(i,z|i-1)C^*(z)[C(z)P(i,z|i-1)C^*(z) + R(z)]^{-1} \quad (5.65)$$

$$P(i,z|i) = [I - K(i,z)C(z)]P(i,z|i-1) \quad (5.66)$$

$$P(i,z|i-1) = A(z)\ P(i-1,z|i-1)\ A^*(z) + D(z)Q(z)D^*(z) \quad (5.67)$$

The z-transforms $P(i,z|i)$ and $P(i,z|i-1)$ are the spectra of the filtered and predicted errors, respectively.

It is to be noted that the Kalman filtering equations (5.63)-(5.67) can be formally thought of as defining the optimal filter for the system

$$x(i+1,z) = A(z)\ x(i,z) + D(z)\ w(i,z) \quad (5.68)$$

$$y(i,z) = C(z)\ x(i,z) + v(i,z) \quad (5.69)$$

where the noises w and v are white noises in both i and z, i.e.

$$E[w(i,z_1)w'(j,z_2)] = Q_o\ \delta(z_1-z_2)\delta_{i,j} \quad (5.70)$$

$$E[v(i,z_1)v'(j,z_2)] = R_o\ \delta(z_1-z_2)\delta_{i,j} \quad (5.71)$$

These equations, of course, are purely formal. But in the transform domain, the dynamics and observations are decoupled and the noises are

independent. Thus the filtering problems for different values of z are independent of one another, e.g. the estimate $\hat{x}(i,z_o|i)$ depends only on the observations $y(j,z_o)$, $j \leq i$. The solution to the filtering problem, therefore, is just a set of low-order Kalman filters, (5.63)-(5.67), indexed by the continuous transform variable z.

In order to determine how the Kalman filter (5.58)-(5.62) can be efficiently implemented, the prediction and update cycles of the discrete-time filter will be addressed separately. First, the measurement update procedure will be considered. The predicted error of substate k at time i is defined as

$$e_k(i) = x_k(i) - \hat{x}_k(i|i-1) \tag{5.72}$$

It is immediately seen that the innovations process $I_k(i)$ is related to the predicted error by

$$\begin{aligned} I_k(i) &= y_k(i) - \sum_{\ell=-\infty}^{+\infty} C_{k-\ell}\hat{x}_\ell(i|i-1) \\ &= \sum_{\ell=-\infty}^{+\infty} C_{k-\ell}e_\ell(i) + v_k(i) \end{aligned} \tag{5.73}$$

By the Orthogonal Projection Theorem of Hilbert Space Theory [83], the estimate $\hat{x}_k(i|i)$ is simply the sum of $\hat{x}_k(i|i-1)$ and $\hat{e}_k(i)$, the optimal estimate of $e_k(i)$. Thus the problem of interest is to determine the estimate $\hat{e}_k(i)$. By (5.58), the optimal estimate of $e_k(i)$ is just the convolution of the filter gains $\{K_k(i)\}$ with the innovations process $\{I_k(i)\}$,

$$\hat{e}_k(i) = \sum_{\ell=-\infty}^{+\infty} K_{k-\ell}(i)\ I_\ell(i) \tag{5.74}$$

Schoute, et al. [76] have given a recursive realization of a Toeplitz operator such as the filter gain K(i) in (5.74). The context in which they worked was recursive estimation of images. They modeled an image as the output of a stochastic, discrete-time Toeplitz system (5.52) and (5.53). The Toeplitz systems considered, however, were of a very special type -- the substates were scalars, the output $y_k(i)$ was just $x_k(i) + v_k(i)$ (i.e. $C_k = \delta_{k,0}$), and the system matrix A was just ρ (a scalar between 0 and 1) times the identity. The image enhancement problem was then solved by performing filtering for the Toeplitz model.

Schoute, et al. observed that if the linear dynamical system S has impulse response $\{K_k(i)\}_{k=0,\pm1,\ldots}$ (i is fixed), then by (5.74) the estimates $\hat{e}_k(i)$ are the output of S when the input is $I_k(i)$. The system S, however, is not a causal system since $K_k(i)$ is not zero for all negative k. It is possible to express the impulse response of S as the sum of a causal and an anticausal part,

$$\{K_k(i)\} = \{h_k^+\} + \{h_k^-\} \tag{5.75}$$

where

$$\{h_k^+\} = \{\ldots,0,0,\alpha K_o(i),\ K_1(i),\ K_2(i),\ldots\}$$

$$\{h_k^-\} = \{\ldots,K_{-2}(i),\ K_{-1}(i),\ (1-\alpha)K_o(i),0,0,\ldots\}$$

and α is an arbitrary scalar. Let S^+ (S^-) be a causal (anticausal) dynam-

ical system with impulse response $\{h_k^+\}$ ($\{h_k^-\}$). Then

$$\hat{e}_k(i) = \varepsilon_k^+ + \varepsilon_k^- \tag{5.76}$$

where ε_k^+ (ε_k^-) is the output of S^+ (S^-) when the input is $I_k(i)$. The system S^- is anticausal for increasing k, but causal for decreasing k. That is, if S^- is viewed as a linear system running backward over k, then it is causal. The estimate $\hat{e}_k(i)$ is the sum of the outputs of two linear dynamical systems -- one (S^+) running forward over k and one (S^-) running backward. This provides a recursive method for implementing (5.74). Moreover, under Assumption 5.1 that the transforms A(z), C(z), D(z), and P(,z|-1) are all rational, the transformed filter gain K(i,z) is also rational for all i. This implies that the sequences $\{h_k^+\}$ and $\{h_k^-\}$ have rational z-transforms and, therefore, the linear systems S^+ and S^- have finite-dimensional realizations. It is to be noted that Assumption 5.1 is only needed to insure that there exist finite-dimensional realizations of S^+ and S^-.

For the special type of Toeplitz systems considered by Schoute, et al., the filter gain matrix is symmetric, K = K'. This means that

$$h_k^+ = h_{-k}^- \tag{5.76}$$

for all non-zero k. Equation (5.76) holds for all k including zero if α is taken to be $\frac{1}{2}$. In this case, the forward system S^+ and the backward system S^- are equal. Thus the same finite-dimensional realization can be

used to compute ε_k^+ and ε_k^-.

Summarizing the development this far, the update cycle of the filter consists of the following steps:

(1.) propagate the error spectrum by (5.65)-(5.67)

(2.) obtain realizations for S^+ and S^-

(3.) run the forward and backward systems S^+ and S^- with input $I_k(i)$

(4.) combine the two outputs by (5.76) to get $\hat{e}_k(i)$

(5.) the updated estimate is $\hat{x}_k(i|i) = \hat{x}_k(i|i-1) + \hat{e}_k(i)$.

The spectrum propagation and realization problem can be done off-line.

From (5.73), it is clear that the innovations are nothing but linear observations of $e_k(i)$ corrupted by additive noise. In general, each $I_k(i)$ depends on all the $e_\ell(i)$. In the special case when $C(z) = C_o$, the innovation $I_k(i)$ is simply a noisy observation of $e_k(i)$. Then determining $\{e_k(i)\}$ from $\{I_k(i)\}$ is a smoothing problem. It is interesting to note that the models of both Attasi [28] and Schoute, et al. [76] satisfy this condition on $C(z)$. The physical interpretation of this assumption is that each subsystem observes a suboutput which is a function of only the local substate.

Assumption 5.2 $C(z) = C_o$, i.e. $C_k = C_o \delta_{k,o}$

Even with assumption 5.2, the optimal estimate $\hat{e}_k(i)$ cannot be made on the basis of $I_k(i)$ alone since the predicted errors $e_k(i)$ and $e_\ell(i)$ are correlated. The cross-correlation is given by

$$P_{k-\ell}(i|i-1) = E[e_\ell(i)e_k'(i)] \tag{5.77}$$

and so $P(i,z|i-1)$ is the spectrum of the predicted error process. For all i, $P(i,z|i-1)$ will be a rational function of z if Assumption 5.1 is valid. In other words, the stationary discrete process $e_k(i)$, as a function of k, has a rational spectrum. Therefore, by the spectral factorization theorem [84], there exists a finite-dimensional linear time-invariant system

$$\xi_{k+1}(i) = \Phi(i)\ \xi_k(i) + \Gamma(i)\mu_k(i) \tag{5.78}$$

$$\zeta_k(i) = \Theta(i)\ \xi_k(i) \tag{5.79}$$

driven by the vector white noise process $\mu_k(i)$ having identity covariance, such that the spectrum of $\zeta_k(i)$ equals $P(i,z|i-1)$. That is to say, the sequence $e_k(i)$ can be identified with the sequence $\zeta_k(i)$ and viewed as the output of the above system. Then the innovations $I_k(i)$ are simply noisy linear observations of the state $\xi_k(i)$,

$$I_k(i) = C_o\Theta(i)\ \xi_k(i) + v_k(i) \tag{5.80}$$

and can be used to estimate it. The optimal estimate $\hat{e}_k(i)$ is now $\Theta(i)\ \hat{\xi}_k(i)$ where $\hat{\xi}_k(i)$ is the smoothed estimate of $\xi_k(i)$.

There are many ways to obtain the smoothed estimate $\hat{\xi}_j(i)$; the discrete-time two-filter smoother described in Appendix B is one alternative. Using this smoother, the estimate is

$$\hat{\xi}_k(i) = P_s(i)\left[P_f^{-1}(i)\ \hat{\xi}^f_{k|k}(i) + P_r^{-1}(i)\ \hat{\xi}^r_{k|k+1}(i)\right] \tag{5.81}$$

$$P_s(i) = \left[P_f^{-1}(i) + P_r^{-1}(i) - \sigma^{-1}(i)\right]^{-1} \tag{5.82}$$

where

$\hat{\xi}^f_{k|k}(i)$ = the forward Kalman filter estimate of $\xi_k(i)$

$\hat{\xi}^r_{k|k+1}(i)$ = the reversed-time Kalman filter one-step-ahead predicted estimate of $\xi_k(i)$

$P_f(i)$ = the steady-state covariance of the forward Kalman filter estimate

$P_r(i)$ = the steady-state covariance of the reversed-time Kalman filter predicted estimate

$P_s(i)$ = the smoothed error covariance

$\sigma(i)$ = the steady-state *a priori* system covariance of (5.78)

The forward Kalman filter operates in the positive k direction; the reversed-time Kalman filter operates in the negative k direction. Both filters are in the steady-state since the k index extends to plus and minus infinity. This smoother may be interpreted as realizing the filter gain matrix by two causal dynamical systems, as proposed by Schoute, et al., by choosing the parameter α in (5.75) equal to one.

Under Assumptions 5.1 and 5.2, the update cycle of the filter can be

summarized as

(1.) propagate the error spectrum by (5.65)-(5.67)

(2.) obtain a realization $(\Phi(i),\Gamma(i),\Theta(i))$ of $P(i,z|i-1)$

(3.) perform forward and backward Kalman filtering for this system

(4.) combine the filtered estimates according to (5.81) to get the smoothed estimate

(5.) the updated estimate is $\hat{x}_j(i|i) = \hat{x}_j(i|i-1) + \Theta(i)\hat{\xi}_j(i)$

Implementation considerations for this algorithm will be discussed in Section 5.3.3. For now it suffices to note that the first two steps of this procedure can be performed off-line.

The remaining step in the filtering algorithm is prediction. In the spatial domain, the prediction step will be a convolution sum,

$$\hat{x}_k(i|i-1) = \sum_{\ell=-\infty}^{+\infty} A_{k-\ell}\hat{x}_\ell(i-1|i-1) \tag{5.83}$$

One noteworthy feature of this expression is that if the dynamics of each sybsystem depend directly on only a finite number of other subsystems, then the convolution sum (5.83) will be a finite sum. That is, suppose A_ℓ is non-zero for only a finite number of values, say $\ell \in J$. Then the predicted estimate $\hat{x}_k(i|i-1)$ can be computed from only $\{\hat{x}_{k-\ell}(i-1|i-1) \,|\, \ell \in J\}$, a finite set. In the general case, the Toeplitz operator A can be realized by the procedure introduced earlier for the operator K(i). Namely, consider the sequence $\{A_k\}$ to be the impulse response of a linear system, and then express this response as the sum of a causal

and an anti-causal part. In this way the predicted estimate $\hat{x}_k(i|i-1)$ is obtained as the sum of the outputs of a forward and a backward linear system. These systems have inputs $\hat{x}_k(i-1|i-1)$ and have a finite-dimensional realization if A(z) is rational.

In summary, the prediction and update steps of the Toeplitz Kalman filter can both be realized by forward and backward linear dynamical systems. Under Assumption 5.1, these systems are finite-dimensional for all i. The prediction step can be obtained as a finite sum whenever there are only a finite number of subsystem interactions. Under Assumption 5.2, the update step is equivalent to a smoothing problem and can be realized by forward and backward Kalman filters.

The models used by Attasi [28] and Schoute, et al. [76] for image processing fit into this framework very nicely. In both cases (see (5.49)), the Toeplitz system matrices are block diagonal and so prediction is particularly simple. Also, both classes of models employ a block diagonal C matrix; hence Assumption 5.2 is satisfied, and smoothing can be done to update the estimates. The substates in the model of Schoute, et al. were constrained to be scalars because they were considering enhancement of monochromatic images. For color pictures, it is necessary to have vector substates. The development of Section 5.3.1 can, therefore, be viewed as extending the algorithm of Schoute, et al. to color pictures. This was left as an open problem in [76].

5.3.2 Implications for Filtering in Large-Scale Systems

The Kalman filter for a large-scale system requires, in general, the totally centralized processing of all the subsystem outputs. In the case of a large-scale system modelled as a Toeplitz system by (5.52) and (5.53), the filter has a special structure. In particular, the preceding development will be applied to show how the Kalman filter can be implemented with only very limited inter-subsystem communication.

Consider first the measurement update step for a Toeplitz system satisfying Assumption 5.2. In this case, the update step is equivalent to a smoothing problem and is accomplished by two steady-state Kalman filters. In order to better understand the update procedure, consider subsystem k as it estimates $e_k(i)$. Subsystem k-1 computes the forward estimate $\hat{\xi}^f_{k-1|k-1}(i)$ and communicates it to subsystem k. The estimate $\hat{\xi}^f_{k|k}(i)$ can now be computed as

$$\hat{\xi}^f_{k|k}(i) = \Phi(i)\hat{\xi}^f_{k-1|k-1}(i) + K_f(i)\left[I_k(i) - C_o\Theta(i)\ \Phi(i)\hat{\xi}^f_{k-1|k-1}(i)\right] \tag{5.84}$$

where $K_f(i)$ is the steady-state gain of the forward filter. This estimate is then furnished to subsystem k+1, and the forward filter continues up the line. Meanwhile, the reversed-time filter is operating down the line independently of the forward filter. At some point, the reversed-time filter reaches subsystem k+1. The one-step-ahead predicted estimate for the reversed-time filter, $\hat{\xi}^r_{k|k+1}(i)$, is then furnished to subsystem k from

subsystem k+1. The next estimate is computed from

$$\hat{\xi}^{r}_{k-1|k}(i) = \Phi_r(i)\left\{\hat{\xi}^{r}_{k|k+1}(i) + K_r(i)\left[I_k(i) - C_o\Theta(i)\ \hat{\xi}_{k|k+1}(i)\right]\right\} \tag{5.85}$$

where $K_r(i)$ is the steady-state gain of the reversed-time filter and the reversed-time system matrix is $\Phi_r(i) = \sigma(i)\ \Phi'(i)\ \sigma^{-1}(i)$. The reversed-time filter then continues in the negative k direction. Subsystem k can now use (5.81) and (5.82) to obtain the smoothed estimate $\hat{\xi}_k(i)$ and then $\hat{e}_k(i) = \Theta(i)\ \hat{\xi}_k(i)$.

The striking aspect of the update step of this Kalman filter is the very limited communication between adjacent subsystems. All that is required is that each subsystem furnish its two nearest neighbors with estimates of the process $\xi_k(i)$. When Assumption 5.2 is not valid, the update step can be realized by a causal and an anticausal dynamical system. These two systems do not have the interpretation of Kalman filters, but they can be implemented by the same inter-subsystem communication pattern described above. Therefore, the update step requires only limited communication between adjacent subsystems, regardless of whether Assumption 5.2 holds.

The prediction cycle of the Kalman filter is implemented by the convolution sum (5.83). Consider the important case of a Toeplitz system with only nearest neighbor interactions, i.e. $A_k=0$, $k\neq -1,0,1$. The predicted estimate of substate k is then simply

$$\hat{x}_k(i|i-1) = A_{-1}\hat{x}_{k+1}(i-1|i-1) + A_o\hat{x}_k(i-1|i-1) + A_1\hat{x}_{k-1}(i-1|i-1) \tag{5.86}$$

This means that subsystem k can form the predicted estimate of its own state from the state estimates of its two nearest neighbors. That is, each sybsystem can predict optimally with only nearest neighbor communication of local estimates. Of course, whenever the subsystem interactions are localized spatially, a similar result holds. Indeed for systems of the type considered by Attasi or Schoute, each subsystem is able to predict independently of all others. In the general case, the operator A can be realized by two linear dynamical systems. The prediction step, then, requires only communication between adjacent subsystems, the same inter-subsystem communication pattern as for the update step.

5.3.3 Filter Implementation Issues

The optimal centralized Kalman filter for Toeplitz systems satisfying Assumption 5.1 has prediction and update cycles which involve a very interesting communication pattern among the subsystems. The purpose of section 5.3.3 is to examine in depth some of the issues that are involved in implementing this Kalman filter.

The optimal filter, of course, will be time-varying. For the update step, this implies that for each time i, a realization problem must be solved for S^+ and S^-, the two finite-dimensional linear systems which are used to implement the operator K(i). The computational problem of determining these two realizations can be done off-line. The transformed

filter gain K(i,z) is recursively computed from the discrete-time Riccati equation (5.65)-(5.67), and hence the degree of K(i,z) as a rational function of z grows rapidly with increasing i. In other words, the complicating fact in this filtering procedure is that the order of the realizations S^+ and S^- is quickly increasing with i. One is lead, therefore, to consider a time-invariant suboptimal filter.

The time-invariant steady-state Kalman filter is specified by the steady-state predicted error covariance P. The corresponding spectrum P(z) is given by the discrete-time algebraic Riccati equation in the transform domain,

$$P(z) = A(z)\left\{P(z) - P(z)C^*(z)[C(z)P(z)C^*(z) + R]^{-1}C(z)P(z)\right\}A^*(z) + \\ + D(z)Q\,D^*(z) \qquad (5.87)$$

The difficulty here is that even though $P(i,z|i-1)$ is a rational function of z for all i, in general, the limiting value P(z) will <u>not</u> be a rational function. There are cases when P(z) is rational, as the following example demonstrates.

<u>Example 5.1</u>. The Toeplitz system in this example has scalar subsystems and dynamics given by

$$x_k(i+1) = \alpha x_{k-1}(i) + x_k(i) + w_{k-1}(i) + w_k(i)$$

The output y_k is just the substate $x_k(i)$ plus noise, and the driving and observation noises have covariances equal to one. Thus

Example 5.1. (contd.)

$$A(z) = z^{-1} + 1$$

$$C(z) = 1$$

$$D(z) = z^{-1} + 1$$

$$Q = 1$$

$$R = 1$$

In the scalar case, (5.75) reduces to

$$[C(z)C^*(z)]P^2(z) + [R-A(z)A^*(z)R - C(z)C^*(z)D(z)D^*(z)Q]P(z)$$

$$- D(z)D^*(z)QR = 0$$

Substituting,

$$P^2(z) + [1 - (\alpha z^{-1}+1)(\alpha z+1) - (z^{-1}+1)(z+1)]P(z) - (z^{-1}+1)(z+1) = 0$$

$$P^2(z) - [(\alpha+1)z + (\alpha^2+2) + (\alpha+1)z^{-1}]P(z) - (z+2+z^{-1}) = 0$$

From the quadratic formula,

$$P(z) = \frac{1}{2}\left\{\left[(\alpha+1)z+(\alpha^2+2)+(\alpha+1)z^{-1}\right] + \sqrt{\left[(\alpha+1)z+(\alpha^2+2)+(\alpha+1)z^{-1}\right]^2 + 4(z+2+z^{-1})}\right\}$$

$$= \frac{1}{2}\left\{\left[(\alpha+1)z+(\alpha^2+2)+(\alpha+1)z^{-1}\right] + \right.$$

$$\left. + \sqrt{(\alpha^2+2\alpha+1)z^2 + (2\alpha^3+2\alpha^2+4\alpha+8)z + (\alpha^4+6\alpha^2+4\alpha+14) + (2\alpha^3+2\alpha^2+4\alpha+8)z^{-1} + (\alpha^2+2\alpha+1)z^{-2}}\right\}$$

Example 5.1. (contd.)

For α between zero and one satisfying $\alpha^3-\alpha^2-2\alpha+1 = 0$, the expression under the radical is $\left[(\alpha+1)z+(-\alpha^2+4\alpha+2)+(\alpha+1)z^{-1}\right]^2$. Then P(z) is

$$P(z) = \frac{1}{2}\left\{\left[(\alpha+1)z+(\alpha^2+2)+(\alpha+1)z^{-1}\right]+\left[(\alpha+1)z+(-\alpha^2+4\alpha+2)+(\alpha+1)z^{-1}\right]\right\}$$

$$= (\alpha+1)z + 2(\alpha+1) + (\alpha+1)z^{-1}$$

In this case, P(z) is a rational function of z. One realization of this spectrum is

$$\begin{bmatrix}\xi_1(k+1)\\ \xi_2(k+1)\end{bmatrix} = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}\xi_1(k)\\ \xi_2(k)\end{bmatrix} + \mu(k)$$

$$\zeta(k) = \begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}\xi_1(k)\\ \xi_2(k)\end{bmatrix}$$

where $E\,\mu(k)\mu'(k) = \begin{bmatrix}0 & 0\\ 0 & \alpha+1\end{bmatrix}$. The steady-state covariance σ of this system satisfies

$$\begin{bmatrix}\sigma_1 & \sigma_2\\ \sigma_2 & \sigma_3\end{bmatrix} = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}\sigma_1 & \sigma_2\\ \sigma_2 & \sigma_3\end{bmatrix}\begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix} + \begin{bmatrix}0 & 0\\ 0 & \alpha+1\end{bmatrix}$$

$$= \begin{bmatrix}\alpha_3 & 0\\ 0 & \alpha+1\end{bmatrix}$$

Example 5.1 (contd.)

Hence $\sigma = \begin{bmatrix} \alpha+1 & 0 \\ 0 & \alpha+1 \end{bmatrix}$. From Appendix B, the reversed-time system matrix is

$$\phi_r = \sigma\phi'\sigma^{-1} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

and the reversed-time driving noise process μ_r has covariance

$$Q_r = \sigma - \phi_r\phi\sigma$$

$$= (\alpha+1)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - (\alpha+1)\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \alpha+1 & 0 \\ 0 & 0 \end{bmatrix}$$

The predicted error covariance of the forward steady-state Kalman is obviously of the form

$$P_f(-) = \begin{bmatrix} \rho & 0 \\ 0 & \alpha+1 \end{bmatrix}$$

where ρ is to be determined. The gain matrix is then

$$K_f = P_f(-)\Theta'[\Theta P_f(-)\Theta' + R]^{-1}$$

$$= \begin{bmatrix} \rho & 0 \\ 0 & \alpha+1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix}\left\{\begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} \rho & 0 \\ 0 & \alpha+1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1\right\}^{-1}$$

Example 5.1. (contd.)

$$K_f = \left(\frac{1}{2+\rho+\alpha}\right) \begin{bmatrix} \rho \\ \alpha+1 \end{bmatrix}$$

The updated error covariance $P_f(+)$ is given by

$$P_f(+) = [I - K_f\Theta]P_f(-)$$

$$= \left\{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \left(\frac{1}{2+\rho+\alpha}\right)\begin{bmatrix} \rho \\ \alpha+1 \end{bmatrix} [1 \quad 1]\right\}\begin{bmatrix} \rho & 0 \\ 0 & \alpha+1 \end{bmatrix}$$

$$= \left(\frac{1}{2+\rho+\alpha}\right)\begin{bmatrix} \rho(\alpha+2) & -\rho(\alpha+1) \\ -\rho(\alpha+1) & (\rho+1)(\alpha+1) \end{bmatrix}$$

Then $P_f(-)$ must obey

$$P_f(-) = \phi\, P_f(+)\phi' + Q$$

$$= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\left(\frac{1}{2+\rho+\alpha}\right)\begin{bmatrix} \rho(\alpha+2) & -\rho(\alpha+1) \\ -\rho(\alpha+1) & (\rho+1)(\alpha+1) \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & 0 \\ 0 & \alpha+1 \end{bmatrix} = \begin{bmatrix} \frac{(1+\alpha)(1+\rho)}{(2+\rho+\alpha)} & 0 \\ 0 & \alpha+1 \end{bmatrix}$$

Therefore, ρ is given by

$$\rho = \frac{(1+\alpha)(1+\rho)}{(2+\rho+\alpha)}$$

$$\therefore \quad \rho = -\frac{1}{2} + \frac{1}{2}\sqrt{5+4\alpha}$$

Example 5.1. (contd.)

Similarly, the covariance and gain matrices for the reversed-time filter are formed to be

$$P_r(-) = \begin{bmatrix} \alpha+1 & 0 \\ 0 & \rho \end{bmatrix}$$

$$P_r(t) = \left(\frac{1}{2+\rho+\alpha}\right) \begin{bmatrix} (\rho+1)(\alpha+1) & -\rho(\alpha+1) \\ -\rho(\alpha+1) & \rho(\alpha+2) \end{bmatrix}$$

$$K_r = \left(\frac{1}{2+\rho+\alpha}\right) \begin{bmatrix} \alpha+1 \\ \rho \end{bmatrix}$$

The estimates of these two filters are then combined according to (5.81) and (5.82).

Example 5.1 shows that is is possible for the steady-state error spectrum $P(z)$, and hence the filter gain $K(z)$, to be a rational function of z. Characterizing precisely when this is the case is still an open problem. It is expected, however, that the occurrence of a rational $P(z)$ and, therefore, a finite-dimensional realization of the update step will be quite rare.

When $P(z)$ is irrational, some approximation must be made in order to obtain a finite-dimensional realization of $K(z)$. Let $P_a(z)$ be a rational approximation of $P(z)$. The corresponding approximate filter gain $K_a(z)$ is given by

$$K_a(z) = P_a(z)C^*(z)[C(z)P_a(z)C^*(z) + R(z)]^{-1} \qquad (5.88)$$

Alternatively, one could directly approximate $K(z)$ by some rational $K_a(z)$ without the intermediate step of obtaining $P_a(z)$. In either case, Schoute, et al. [76] suggest defining the relative approximation error ε_K by

$$\varepsilon_K = \frac{\| K - K_a \|}{\| K \|} \tag{5.89}$$

where the norms here are the induced operator norms introduced in Section 5.1.1. By (5.9), the relative approximation error can be computed from the z-transforms of K and K_a,

$$\varepsilon_K = \frac{\underset{z \in U}{\text{ess sup}} \| K(z) - K_a(z) \|_2}{\underset{z \in U}{\text{ess sup}} \| K(z) \|_2} \tag{5.90}$$

Of more interest than the relative approximation error, however, is the steady-state covariance that results when $K_a(z)$ is used as the filter gain. The predicted error is given by

$$\begin{aligned} e(i+1) &= x(i+1) - \hat{x}(i+1|i) \qquad (5.91) \\ &= \{A\, x(i) + Dw(i)\} - A\{\hat{x}(i|i-1) + K_a[y(i) - C\hat{x}(i|i-1)]\} \\ &= A[I - K_a C]e(i) + Dw(i) - AK_a v(i) \end{aligned}$$

If the filter system matrix $A[I - K_a C]$ is stable, then the predicted error covariance reaches a steady-state value. The resulting predicted error spectrum $P_p(z)$ is given by

$$\begin{aligned} P_p(z) = {} & A(z)[I - K_a(z)C(z)]P_p(z)[I - K_a(z)C(z)]^* A^*(z) + D(z)Q(z)D^*(z) + \\ & + A(z)K_a(z)R(z)K_a^*(z)A^*(z) \end{aligned} \tag{5.92}$$

Equation (5.92) is just the steady-state Lyapunov equation for the system described by (5.91). In terms of $P_p(z)$, the filtered error spectrum $P_f(z)$ is simply

$$P_f(z) = [I - K_a(z)C(z)]\, P_p(z)\, [I - K_a(z)C(z)]^* + K_a(z)R(z)K_a^*(z) \quad (5.93)$$

The suboptimality of the approximate filter gain $K_a(z)$ can be evaluated, therefore, from (5.92) and (5.93).

The update step when Assumption 5.2 holds, i.e., $C(z) = C_o$, will now be examined further. Recall that in this case the update step is equivalent to a smoothing problem and can be realized by two Kalman filters. Let $P_a(z)$ be a rational function of z which very closely approximates $P(z)$. The corresponding finite-dimensional linear system $(\Phi_a, \Gamma_a, \Theta_a)$ might have a very high dimension. Assume that this is the case. The idea here is not to use $(\Phi_a, \Gamma_a, \Theta_a)$ to implement the smoothing of the predicted errors, but rather to use it as a benchmark against which reduced order smoothers can be compared. Consider the reduced-order model $(\Phi_r, \Gamma_r, \Theta_r)$. Then the suboptimality of smoothing with this reduced order model instead of the higher-order model can be evaluated from the results in Appendix B--assuming that the spectrum $P_a(z)$ is the actual predicted error spectrum. Of course, $P_a(z)$ only approximates the actual spectrum $P(z)$, and so the above procedure yields only the approximate suboptimality of using the reduced-order model. Nevertheless, if $P_a(z)$ is chosen to be an accurate representation of $P(z)$, this approach will provide a useful measure of the suboptimality of the reduced-order smoother.

The above discussion has centered on approximating the irrational function of z, K(z), by a rational function $K_a(z)$. The rational filter gain can then be implemented by two finite-dimensional linear systems. This approximation necessarily introduces a degree of suboptimality into the filter. There is a second possible source of suboptimality which will now be discussed.

The estimate $\hat{e}_k(i)$ of the error at subsystem k depends, in general, on the innovations $I_\ell(i)$ all along the line, i.e., $\forall\, \ell$. This centralized estimate is obtained from two Kalman filters, both starting infinitely far away from subsystem k. Clearly, this implementation has some undesirable properties, e.g., an infinite delay is required to compute $\hat{e}_k(i)$. Consider, therefore, estimating $e_k(i)$ from only a finite number of the innovations $I_\ell(i)$ at neighboring subsystems. This corresponds to using a filter gain K_a which has only a finite number of nonzero elements. Any such gain K_a has a transform $K_a(z)$ which is a rational function of z, and hence (5.92) and (5.93) can be used to evaluate the suboptimality of using only a finite number of measurements to update each substate estimate $\hat{x}_k(i|i)$.

In general, it is not particularly clear how such a K_a, with only a finite number of nonzero elements, should be chosen. But under Assumption 5.2, it is quite obvious how to handle this situation. Suppose the estimate $\hat{e}_k(i)$ is restricted to being a function of $I_\ell(i)$ for $k-N_1 \leq \ell \leq k+N_2$. This is nothing but a finite interval smoothing problem. Thus, the estimate $\hat{e}_k(i)$ can be obtained as the output of two Kalman filters -- the forward one starting at $k-N_1$ and the reversed-time one starting at $k+N_2$. The two filters are not in the steady state, but are time(space)-varying.

The covariance of the estimate $\hat{e}_k(i)$ can be obtained from the reduced order smoother results of Appendix B assuming that $P_a(z)$ is the spectrum of the process $e_k(i)$.

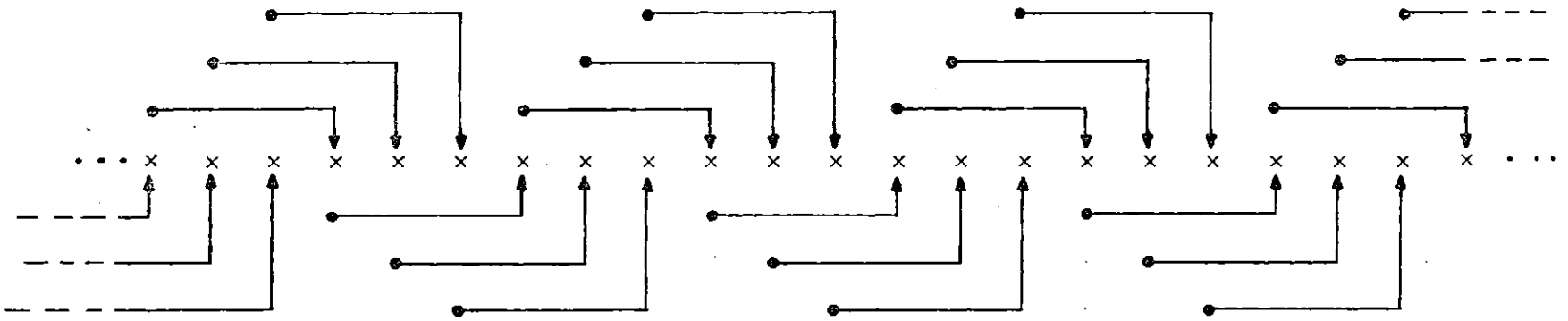
The implementation of the finite-interval smoother will now be considered. Since the estimate $\hat{e}_k(i)$ is obtained from two Kalman filters, only the forward filter will be explicitly considered. Similar comments apply to the reversed-time filter.

The forward estimate of $e_k(i)$ is based on $I_\ell(i)$ for $k-N_1 \leq \ell \leq k$. This estimate is obtained from a forward Kalman filter starting at subsystem $k-N_1$. If the forward estimate of $e_{k+1}(i)$ is based on the same number of innovations, i.e., $I_\ell(i)$ for $k-N_1+1 \leq \ell \leq k+1$, then another Kalman filter starting at subsystem $k-N_1+1$ is required. Continuing this argument, it is clear that if each estimate uses exactly N_1+1 of the innovations, then it is necessary to start a forward Kalman filter at each subsystem. Also, each innovation $I_\ell(i)$ is used to update N_1+1 different forward Kalman filters. Figure 5.2 illustrates the use of a separate filter to obtain each estimate. The number of updates required to compute the forward estimate of $e_k(i)$, of course, is also N_1+1.

Since each estimate is computed from a separate forward Kalman filter, this is a totally parallel computational scheme. Consider now the processing that occurs at each subsystem as this parallel filtering is performed. The innovations $I_k(i)$ at subsystem k are used N_1+1 times to update N_1+1 different filters, as previously noted. Subsystem k transmits N_1 of these estimates to subsystem $k+1$; the other estimate yields $\hat{e}_k(i)$. The

83702AW074

$N_1 = 3$



× Subsystem

⟶ An estimate is provided

Figure 5.2. Each forward estimate uses exactly 4 innovations and is obtained from a separate Kalman filter.

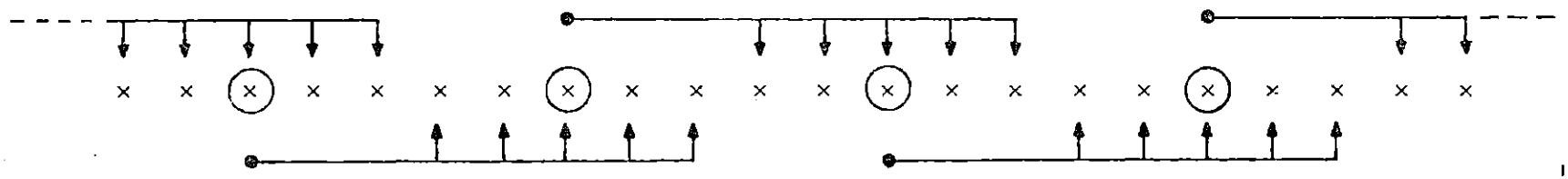
computational burden at each sybsystem under this parallel filtering procedure, therefore, is very much different from the computational burden when all the forward estimates are computed from one filter starting at minus infinite -- essentially a totally serial computational scheme. In the latter case, the innovation $I_k(i)$ is used to update only one filter, and only one estimate is transmitted from subsystem k to subsystem k+1. The conclusion is that the parallel scheme greatly increases both the number of times each subsystem must update a filter, and the number of transmissions between subsystems.

By altering the requirement that each forward estimate must be based on exactly N_1+1 of the innovations, it is possible to obtain a tradeoff between: (i) the computations and transmissions required at each subsystem, and (ii) the delay in computing all the forward estimates. Delay here refers to the maximum number of updates and transmissions needed to compute any particular estimate. This will become more clear shortly. Suppose the requirement is that <u>at least</u> N_1+1 innovations are used for every forward estimate and a new Kalman filter is started every M_1 subsystems. This implementation is shown in Figure 5.3. Under the totally parallel scheme, M_1 was equal to one.

If a new filter happens to be started at subsystem k_o, this filter provides its first estimate (based on N_1+1 innovations) at subsystem $k_o + N_1$. The next filter starts at subsystem $k_o+ M_1$ and provides its first estimate at subsystem $k_o+ M_1 + N_1$. The filter that started at subsystem k_o, therefore, must provide estimates at subsystems k_o+M_1, k_o+M_1+1, ..., $k_o+M_1+N_1-1$ -- a total of M_1 estimates. This last estimate (the one

$N_1 = 3$

$M_1 = 5$



× Subsystems

(×) Subsystems at which a new filter is begun

⟶ An estimate is provided

Figure 5.3 A new forward Kalman filter is begun every 5 subsystems, and at least 4 innovations are used in every estimate.

at subsystem $k_o+M_1+N_1-1)$ is based on N_1+M_1 innovations. This means that the delay in obtaining this estimate is due to updates at N_1+M_1 subsystems and transmissions between N_1+M_1-1 subsystems. When a new Kalman filter is begun at every subsystem, i.e., $M_1 = 1$, its last (in fact only) estimate requires just N_1+1 updates and N_1 transmissions. Choosing M_1 greater than one, therefore, increases the delay in obtaining the forward estimates.

However, what is the effect of increasing M_1 on the computational burden of each subsystem? Consider the choice $N_1=3$ and $M_1=5$ as depicted in Figure 5.3. If a new filter is begun at subsystem k_o, then subsystems k_o, k_o+1, and k_o+2 must update two different filters, and subsystems k_o and k_o+1 must transmit two estimates apiece. The remaining subsystems update only one filter and transmit only one estimate. For general N_1 and M_1, simple counting arguments may be used to determine how many updates and transmissions must be performed by the various subsystems. The point here is that the computational burden at the individual subsystems is reduced by increasing M_1, the separation between adjacent Kalman filters. Increasing M_1, however, was previously observed to result in an increased delay before the last estimate of a Kalman filter was available. Therefore the parameter M_1 can be used to perform a tradeoff between (i) computations and transmissions, and (ii) delay, as was desired.

The above discussion has dealt with a suboptimal update step that uses only a finite number of the innovations to update each subsystem. How many innovations are required to yield a good estimate of $e_k(i)$?

Recall that the process $\{e_k(i)\}$ is viewed as the output of the linear system $[\Phi(i),\Gamma(i), \Theta(i)]$. If the innovations used to estimate $e_k(i)$ extend about subsystem k for several of the slowest time (space) constants of $\Phi(i)$, then the resulting estimate should be quite good. There are, therefore, two time scales of interest here. First, of course, is the actual time index of the system. For example, one may ask how large i must be before the system can be considered to be in the steady state. The second time scale is actually a spatial scale. The question here is how many neighboring subsystems must be used to provide an accurate estimate of a substate.

It is important to realize that as more neighboring subsystems are used to compute a substate estimate, i.e., as the number of space constants is increased, the time required to perform the update step is also increased. This means the space and time indices of the system directly interact in the filtering process. There is an interesting and important tradeoff, therefore, between (i) the actual time period at which the observations are sampled, and (ii) the number of neighboring subsystems used to update a substate estimate. Work in the area of filtering for systems with multiple time scales, such as [17] might be useful here in making this tradeoff.

5.4 The Dual Control Problem

The Toeplitz optimal control problem which is the dual of the estimation problem in Section 5.3 will be solved in this section. The dynamics of the deterministic discrete-time Toeplitz system under consideration are given in terms of the infinite-dimensional state vector x and input vector u as

$$x(i+1) = A'x(i) + C'u(i) \tag{5.94}$$

For subsystem x_k, the dynamics are written

$$x(i+1) = \sum_{\ell=-\infty}^{+\infty} A'_{-k+\ell}\, x(i) + C'_{-k+\ell} u(i) \tag{5.95}$$

where the subsystem matrices $A'_{-k+\ell}$ and input matrices $C'_{-k+\ell}$ are reversed as well as transposed. The quadratic cost functional to be minimized is

$$J = \frac{1}{2} x'(T) S x(T) + \frac{1}{2} \sum_{i=0}^{T-1} [x'(i) DQD' x(i) + u'(i) R u(i)] \tag{5.96}$$

where $S = S'$ is a positive definite Toeplitz operator.

The optimal control problem consists of determining the input time function $u(i)$ which minimizes the cost functional J. If the system (5.94) is stabilizable, then Hager and Horowitz [82] have shown that the optimal control at time i is given by

$$u(i) = - G(i) A' x(i) \tag{5.97}$$

where the gain G(i) is

$$G(i) = [R + CL(i+1)C']^{-1} CL(i+1) \tag{5.98}$$

and L(i) is given by the Riccati equation

$$L(i) = A \left\{ L(i+1) - L(i+1)C' [R + CL(i+1)C']^{-1} CL(i+1) \right\} A' + DQD' \tag{5.99}$$

$$L(T) = S$$

The operator L(i) yields the cost-to-go since

$$\frac{1}{2} x'(i) L(i) x(i) = \min_{\{u(j)\}_i^{T-1}} \left\{ \frac{1}{2} x'(T) S x(T) + \frac{1}{2} \sum_{j=1}^{T-1} [x'(j) DQD' x(j) + u'(j) R u'j)] \right\} \tag{5.100}$$

Suppose the cost operator S is chosen equal to $P(0|-1)$, the initial state covariance in the estimation problem. Then it is easily shown from (5.60)-(5.62) and (5.99) that

$$L(i) = P(T-i|T-i-1) \tag{5.101}$$

for all i between 0 and T. Moreover, comparing (5.98) with (5.60) yields

$$G(i) = K'(t-i) \tag{5.102}$$

i.e., the control gain at time i equals the transposed filter gain at time T-i. Because of the equalities in (5.101) and (5.102), the optimal control problem defined by (5.95) and (5.96) is called the dual of the estimation problem covered in Section 5.3.

For notational convenience, let $x^p(i)$ equal $A'x(i)$. Then the optimal control $u_k(i)$ is the convolution of the control gains $\{G_k(i)\}$ with the projected process $\{x_k^p(i)\}$,

$$u_k(i) = \sum_{\ell=-\infty}^{+\infty} G_{k-\ell}(i)x_\ell^p(i) \tag{5.103}$$

The Toeplitz operator G(i), of course, can be recursively realized by a causal and an anticausal linear system, as explained in Section 5.3.1. Denote these two linear systems by S^+ and S^-, respectively, and let μ_k^+ (μ_k^-) be the output of S^+ (S^-) when the input is $x_k^p(i)$. Then

$$u_k(i) = \mu_k^+ + \mu_k^- \tag{5.104}$$

The various comments made in Section 5.3.1 concerning this realization and the discussion in Sections 5.3.2 and 5.3.3 are equally applicable here and will not be repeated.

Systems which satisfy Assumption 5.2, however, will be investigated further. Recall that Assumption 5.2 was that $C_k = C_0 \delta_{k,0}$; in the context of the optimal control problem, this means that the input matrix is block diagonal, i.e., the input directly affects only the local subsystem. Under this assumption, the smoothed estimate of $e_k(i)$ was found to be a weighted sum of two filtered estimates [see (5.81)]. It will now be shown that the optimal control $u_k(i)$ is just the sum of two optimal controls — one control is for a system operating in the positive k direction, the other control is for a system operating in the negative k direction. This proof will proceed by first reconsidering the dual filtering problem at time T-i and then proposing the forward and backward control problems which yield $u_k(i)$.

Recall that because $P(T-i,z|T-i-1)$, the spectrum of the predicted error process, was a rational function of z, it was possible to view $e_k(T-i)$ as the output of the linear system defined by the triple $[\Phi(T-i),\Gamma(T-i),\Theta(T-i)]$. In order to obtain the smoothed estimate of $e_k(T-i)$, the forward estimate $\hat{\xi}^f_{k|k}(T-i)$ and reversed-time estimate $\hat{\xi}^r_{k|k+1}(T-i)$ were both needed. The forward estimate is computed by the usual Kalman filter (explicit dependence on T-i now is suppressed),

$$\hat{\xi}^f_{k+1|k+1} = \Phi \hat{\xi}^f_{k|k} + K_f \left[I_{k+1} - (C_o\Theta) \Phi \hat{\xi}^f_{k|k} \right] \tag{5.105A}$$

$$K_f = P_f^- (C_o\Theta)' \left[(C_o\Theta) P_f^- (C_o\Theta)' + R_o \right]^{-1} \tag{5.105B}$$

$$P_f^+ = [I - K_f(C_o\Theta)]P_f^- \tag{5.105C}$$

$$P_f^- = \Phi P_f^+ \Phi' + \Gamma\Gamma' \tag{5.105D}$$

Here P_f^+ and P_f^- denote the steady-state filter and one-step-ahead predictor error covariances, respectively. In Section 5.3, only the filter error covariance was needed, and so P_f^+ was simply denoted as P_f. Using the variation of constants formula together with (5.105A), the filter estimate $\hat{\xi}^f_{k|k}$ may be written as

$$\hat{\xi}^f_{k|k} = \sum_{\ell=0}^{+\infty} [(I - K_f C_o\Theta)]^\ell K_f I_{k-\ell} \tag{5.106}$$

Similarly, the reversed-time estimate $\hat{\xi}^r_{k|k+1}$ is computed by a reversed-time one-step-ahead predictor,

$$\hat{\xi}^r_{k|k+1} = \Phi_r\left\{\hat{\xi}^r_{k+1|k+2} + K_r\left[I_{k+1} - (C_o\Theta)\hat{\xi}^r_{k+1|k+2}\right]\right\} \tag{5.107A}$$

$$K_r = P_r^-(C_o\Theta)'\left[(C_o\Theta)P_r^-(C_o\Theta)' + R_o\right]^{-1} \tag{5.107B}$$

$$P_r^+ = [I - K_r(C_oR)]P_r^- \tag{5.107C}$$

$$P_r^- = \Phi_r P_r^+ + \Gamma_r\Gamma_r' \tag{5.107D}$$

where the matrices Φ_r and Γ_r are computed from the formulas (B.3) and (B.4) in Appendix B. Once again, P_r^+ and P_r^- denote the steady-state filter and one-step-ahead predictor error covariances, respectively. In Section 5.3, $P_r = P_r^-$. Alternatively,

$$\hat{\xi}^{r}_{k|k+1} = \sum_{\ell=0}^{+\infty} [\Phi_r (I - K_r C_o \Theta)]^{\ell} \Phi_r K_r I_{k+\ell+1} \tag{5.108}$$

Therefore, using (5.81),

$$\hat{e}_k(i) = \Theta P_s \left[(P_f^+)^{-1} \hat{\xi}^{f}_{k|k} + (P_r^-)^{-1} \hat{\xi}^{r}_{k|k+1} \right] \tag{5.109}$$

$$= \sum_{\ell=0}^{+\infty} \Theta P_s [(I - K_f C_o \Theta)\Phi]^{\ell} K_f I_{k-\ell} +$$

$$+ \sum_{\ell=0}^{+\infty} \Theta P_s [\Phi_r (I - K_r C_o \Theta)]^{\ell} \Phi_r K_r I_{k+\ell+1}$$

Since $\hat{e}_k(T-i)$ is known from (5.74) to be the convolution of the filter gains $\{K_k(T-i)\}$ with the innovations $\{I_k(i)\}$, (5.109) implies that the filter gains are

$$K_k(T-i) = \begin{cases} \Theta P_s P_f^{-1} [(I - K_f C_o \Theta)\Phi]^{k} K_f & , k \geq 0 \\ \Theta P_s P_r^{-1} [\Phi_r (I - K_r C_o \Theta)]^{-k+1} \Phi_r K_r & , k < 0 \end{cases} \tag{5.110}$$

A forward-time and a reversed-time optimal control problem will now be posed, solved, and related to the two filters used in the previous paragraph. First, consider the following forward-time system:

$$\eta^{r}_{k} = \Phi_r' \eta^{r}_{k-1} + (C_o \Theta)' u^{r}_{k-1} \tag{5.111}$$

with the infinite-horizon quadratic cost functional

$$J_r = \sum_{k=0}^{+\infty} \eta_k^{r'} (\Gamma_r \Gamma_r') \eta_k^r + u_k^{r'} R_o u_k^r \tag{5.112}$$

The superscript r was used in the dual filtering problem to denote reversed-time quantities. Hence in the control problem, the superscript r denotes the forward-time state η_k^r and control u_k^r. The realization in (5.111) may be viewed as coming from a factorization of the cost-to-go matrix L(i). The well known solution to this control problem is that the optimal control u_k^r is given by

$$u_k^r = - G_r \Phi_r' \eta_k^r \tag{5.113}$$

where the gain G_r is

$$G_r = [R_o + (C_o\Theta) L_r (C_o\Theta)']^{-1} (C_o\Theta) L_r \tag{5.114}$$

and the cost-to-go L_r satisfies the algebraic Riccati equation

$$L_r = \Phi_r L_r \Phi_r' + \Gamma_r \Gamma_r' - \Phi_r L_r (C_o\Theta)' [R_o + (C_o\Theta) L_r (C_o\Theta)'] (C_o\Theta) L_r \Phi_r' \tag{5.115}$$

The closed-loop system, therefore, is

$$\eta_k^r = [(I - \Theta' C_o' G_r) \Phi_r'] \eta_{k-1}^r \tag{5.116}$$

Next, consider the reversed-time system

$$\eta_k^f = \Phi' \eta_{k+1}^f + (C_o\Theta\Phi)' u_{k+1}^f \tag{5.117}$$

with the infinite-horizon cost functional

$$J_f = \sum_{k=0}^{-\infty} \left\{ [\eta_k^f + (C_o\Theta)' u_k^f]' (\Gamma\Gamma') [\eta_k^f + (C_o\Theta) u_k^f] + u_k^{f'} R_o u_k^f \right\} \tag{5.118}$$

The solution to this problem is

$$u_k^f = - G_f \eta_k^f \tag{5.119}$$

where the control gain G_f is

$$G_f = \left\{ (C_o\Theta)[\Phi L_f \Phi' + \Gamma\Gamma'](C_o\Theta)' + R_o \right\}^{-1} (C_o\Theta)[\Phi L_f \Phi' + \Gamma\Gamma'] \tag{5.120}$$

and the cost-to-go L_f is given by

$$L_f = [I - G_f'(C_o\Theta)][\Phi L_f \Phi' + \Gamma\Gamma'][I - (C_o\Theta)' G_f] + G_f' R_o G_f \tag{5.121}$$

The resulting closed-loop system is

$$\eta_k^f = [\Phi'(I - \Theta' C_o' G_f)] \eta_{k+1}^f \tag{5.122}$$

The forward- and reversed-time optimal control systems are now related to the two filters which yield the smoothed estimate $\hat{e}_k$. First, eliminating P_r^+ from (5.107B)- (5.107D) yields

$$K_r = P_r^- (C_o\Theta)' [(C_o\Theta) P_r^- (C_o\Theta)' + R_o]^{-1} \tag{5.123}$$

$$P_r^- = \Phi_r P_r^- \Phi_r' + \Gamma\Gamma' - \Phi_r P_r^- (C_o\Theta)' [(C_o\Theta) P_r^- (C_o\Theta)' + R_o]^{-1} (C_o\Theta) P_r^- \Phi_r' \tag{5.124}$$

A comparison of (5.115) with (5.124) shows that

$$L_r = P_r^- \tag{5.125}$$

Therefore, from (5.114) and (5.123), the control gain equals the transposed filter gain,

$$G_r = K_r' \tag{5.126}$$

Likewise, eliminating P_f^- from (5.105B) - (5.105D) yields

$$K_f = [\Phi P_f^+ \Phi' + \Gamma\Gamma'](C_o\Theta)' \left\{(C_o\Theta)[\Phi P_f^+ \Phi' + \Gamma\Gamma'](C_o\Theta)' + R_o\right\}^{-1} \tag{5.127}$$

$$P_f^+ = [I-K_f(C_o\Theta)][\Phi P_f^+ \Phi' + \Gamma\Gamma'] \tag{5.128}$$

It is not obvious that P_f^+ equals L_f. However, from (5.121),

$$\begin{aligned} L_f &= [I-G_f'(C_o\Theta)][\Phi L_f\Phi' + \Gamma\Gamma'][I-(C_o\Theta)'G_f] + G_f' R_o G_f \\ &= [\Phi L_f\Phi' + \Gamma\Gamma'] - G_f'(C_o\Theta)[\Phi L_f\Phi' + \Gamma\Gamma'] - [\Phi L_f\Phi' + \Gamma\Gamma'](C_o\Theta)' G_f + \\ &\quad + G_f' \left\{(C_o\Theta)[\Phi L_f\Phi + \Gamma\Gamma'](C_o\Theta)' + R_o\right\} G_f \\ &= [\Phi L_f\Phi' + \Gamma\Gamma'] - G_f'(C_o\Theta)[\Phi L_f\Phi' + \Gamma\Gamma'] \\ &= [I-G_f'(C_o\Theta)][\Phi L_f\Phi' + \Gamma\Gamma'] \end{aligned} \tag{5.129}$$

Now, comparing (5.128) with (5.129) and (5.127) with (5.120), one sees that

$$L_f = P_f^+ \tag{5.130}$$

$$G_f = K_f' \tag{5.131}$$

Consider the closed-loop systems (5.116) and (5.122) where now x_k^p enters these systems as a disturbance. That is,

$$\eta_k^r = [(I-\Theta' C_o' K_r')\Phi_r']\eta_{k-1}^r + (P_r^{-1} P_s \Theta')\, x_{k-1}^p \tag{5.132}$$

$$\eta_k^f = [\Phi'(I-\Theta' C_o' K_f')]\eta_{k+1}^f + (P_f^{-1} P_s \Theta')\, x_k^p \tag{5.133}$$

where K_r' and K_f' have been substituted for G_r and G_f, respectively. From the variation of constants formula,

$$\eta_k^r = \sum_{\ell=0}^{+\infty} [(I - \Theta' C_o' K_r')\Phi_r']^{\ell-1} (P_r^{-1} P_s \Theta')\, x_{k-\ell-1}^p \tag{5.134}$$

$$\eta_k^f = \sum_{\ell=0}^{+\infty} [\Phi'(I - \Theta' C_o' K_f')]^{\ell} (P_f^{-1} P_s \Theta')\, x_{k+\ell}^p \tag{5.135}$$

From (5.103), the optimal input u_k is the negative of the convolution of the control gains $\{G_k\}$ with the projected process $\{x_k^p\}$. Furthermore, from (5.102) $G_k = K_{-k}'$ and so, using (5.110),

$$G_k = \begin{cases} K_f' [\Phi'(I-\Theta' C_o' K_f')] P_f^{-1} P_s \Theta' & , k \leq 0 \\ K_r' \Phi_r' [(I-\Theta' C_o' K_r')\Phi_r'] P_r^{-1} P_s \Theta' & , k > 0 \end{cases} \tag{5.136}$$

But, from (5.113), (5.119), (5.134) and (5.135), the optimal controls u_k^r and u_k^f are

$$u_k^r = - \sum_{\ell=0}^{+\infty} K_r' \Phi_r' [(I-\Theta' C_o' K_r')\Phi_r']^{\ell-1} (P_r^{-1} P_s \Theta') x_{k-\ell-1}^p \tag{5.137}$$

$$u_k^f = - \sum_{\ell=0}^{+\infty} K_f' [\Phi'(I-\Theta' C_o' K_f')]^{\ell} (P_f^{-1} P_s \Theta') x_{k+\ell}^p \tag{5.138}$$

A comparison of (5.137) and (5.138) with the two parts of (5.136) shows that

$$u_k = u_k^r + u_k^f \tag{5.139}$$

That is to say, the optimal input $u_k(i)$ to subsystem k at time i is simply the sum of two inputs -- the one u_k^r, the input for a system propagating in positive k; the other u_k^f, the input for a system propagating in negative k. Interestingly enough, this is <u>not</u> a weighted sum of the forward and backward inputs, but simply their algebraic sum. The weights which are used to combine the two estimates $\hat{\xi}^f_{k|k}$ and $\hat{\xi}^r_{k|k+1}$ in the dual smoothing problem are found on the process x_k^p as it enters as a disturbance in (5.132) and (5.133).

In summary, this section has considered the dual optimal control problem to the Toeplitz estimation problem covered in Section 5.3. The realization of a Toeplitz operator given in Section 5.3.1 and the discussion in Sections 5.3.2 and 5.3.3 are, by duality, also applicable to the dual control problem. The most interesting result in this section, however, concerns the optimal control problem for systems satisfying Assumption 5.2. The update step of the dual filtering problem in this case was shown in Section 5.3 to be essentially a smoothing problem. By dynamic programming, one step of the optimal control problem consists of the following minimization:

$$\min_{\{u_k\}} \sum_{k=-\infty}^{+\infty} u_k' R_o u_k + \sum_{\ell=-\infty}^{+\infty} \left| (x_k^p + C_o' u_k)' L_{k-\ell} (x_\ell^p + C_o' u_\ell) \right| \qquad (5.140)$$

The minimizing u_k has been expressed as the sum of the optimal control for a forward-time system plus the optimal control for a reversed-time system. This appears to be a novel solution to the dual of the standard smoothing problem.

CHAPTER 6

CONTRIBUTIONS AND SUGGESTIONS

Throughout this dissertation, spatial transformations have been used to provide a deep, coherent treatment of spatially invariant linear systems and the associated control and estimation problems. Both finite-dimensional circulant systems and infinite-dimensional Toeplitz systems were investigated. It is strongly felt that the results in this dissertation demonstrate the utility of viewing control and estimation problems from the spatial frequency domain. The major contributions of this work follow:

(1) The complete treatment of circulant systems, their system theoretic properties, and the associated Lyapunov and Riccati equations by using the spatial DFT to perform decomposition.

(2) The procedure for imbedding in circulant systems both symmetric and antisymmetric tridiagonal systems, such as those arising from some discretized partial differential equations or systems having longitudinal structure.

(3) The development of design procedures for suboptimal decentralized control gains from the optimal centralized control gains by analogy with digital filter design.

(4) The study of the two-filter smoother and the resulting increased understanding of the role and nature of the backwards filter in computing the smoothed estimate.

(5) The development of a sensitivity analysis and a reduced-order smoother analysis for the two-filter smoother and change of initial conditions formulas for smoothed estimates.

(6) The application of the spatial transform together with the two-filter smoother to solve the filtering problem for certain Toeplitz systems.

(7) The formulation and solution of the optimal control problem which is the dual of the fixed-interval smoothing problem.

Some of the possible topics for future research related to the results obtained and approaches developed in this disseration are the following:

(1) The imbedding of other classes of linear systems within circulant systems, e.g., can more general tridiagonal systems be imbedded in circulant systems by combining the results for symmetric and anti-symmetric tridiagonal systems?

(2) The characterization of when a circulant decentralized controller is stabilizing and the development of bounds on the resulting suboptimality.

(3) The development of procedures for determining the appropriate circulant approximation to a given large-scale system for purposes of control and estimation; the characterization of the stability of the resulting closed-loop system; and the bounding of the suboptimality of the circulant controller or estimator.

(4) A study of the numerical properties of the new two-filter smoothing formulas and the equations for sensitivity analysis and reduced-order smoother analysis.

(5) A theoretical study of the use of reversed-time realizations and the reversed-time innovations process.

(6) An examination of the space-time interplay that occurs in filtering for Toeplitz systems, e.g., the tradeoff between the number of innovations used to estimate the state of each subsystem and the time interval at which observations are made.

(7) The development of the finite-dimensional version of the filtering results obtained for Toeplitz systems.

APPENDIX A. CIRCULANT MATRICES

Let A be an $N \times N$ circulant matrix

$$A = \begin{pmatrix} a_0 & a_{N-1} & a_{N-2} & \cdots & a_1 \\ a_1 & a_0 & a_{N-1} & \cdots & a_2 \\ a_2 & a_1 & a_0 & \cdots & a_3 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ a_{N-1} & a_{N-2} & a_{N-3} & \cdots & a_0 \end{pmatrix} \tag{A.1}$$

The eigenvalues and eigenvectors of A will now be found following the proof given by Bellman [44]. If $W_N = \exp\left(\frac{j2\pi}{N}\right)$, then W_N^{-k} is one of the N^{th} roots of unity, i.e., $\left(W_N^{-k}\right)^N = 1$. Let

$$\lambda_k = \sum_{i=0}^{N-1} a_i \left(W_N^{-k}\right)^i \tag{A.2}$$

From (A.2) and the fact that W_N^{-k} is an N^{th} root of unity, it is obvious that λ_k satisfies the following set of equations:

$$\lambda_k = a_0 + a_{N-1}\left(W_N^k\right) + a_{N-2}\left(W_N^k\right)^2 + \ldots + a_1\left(W_N^k\right)^{N-1} \tag{A.3}$$

$$\lambda_k\left(W_N^k\right) = a_1 + a_0\left(W_N^k\right) + a_{N-1}\left(W_N^k\right)^2 + \ldots + a_2\left(W_N^k\right)^{N-1}$$

$$\vdots$$

$$\lambda_k\left(W_N^k\right)^{N-1} = a_{N-1} + a_{N-2}\left(W_N^k\right) + a_{N-3}\left(W_N^k\right)^2 + \ldots + a_0\left(W_N^k\right)^{N-1}$$

Writing this set of equations in matrix form yields

$$\lambda_k \begin{pmatrix} 1 \\ W_N^k \\ \cdot \\ \cdot \\ \cdot \\ W_N^{(N-1)k} \end{pmatrix} = \begin{pmatrix} a_0 & a_{N-1} & \cdots & a_1 \\ a_1 & a_0 & \cdots & a_2 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{N-1} & a_{N-2} & \cdots & a_0 \end{pmatrix} \begin{pmatrix} 1 \\ W_N^k \\ \\ \\ \\ W_N^{(N-1)k} \end{pmatrix} \tag{A.4}$$

or

$$\lambda_k \phi_k = A\phi_k \tag{A.5}$$

where the eigenvector ϕ_k is simply

$$\phi_k = \begin{pmatrix} 1 \\ W_N^k \\ W_N^{2k} \\ \cdot \\ \cdot \\ \cdot \\ W_N^{(N-1)k} \end{pmatrix} \tag{A.6}$$

The eigenvalues λ_k of the circulant matrix A are given by (A.2) -- the elements of the discrete Fourier transform of the top row of A. The eigenvectors ϕ_k are observed to depend only on N and not on the elements of A.

Let Φ be the matrix of eigenvectors,

$$\Phi = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W_N^1 & W_N^2 & \cdots & W_N^{(N-1)} \\ 1 & W_N^2 & W_N^4 & \cdots & W_N^{(N-2)} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1 & W_N^{(N-1)} & W_N^{(N-2)} & \cdots & W_N^{(N-1)(N-1)} \end{pmatrix} \qquad \text{(A.7)}$$

This matrix is invertible, since its inverse can be explicitly written as

$$\Phi^{-1} = \frac{1}{N} \begin{pmatrix} 1 & 1 & 1 & & 1 \\ 1 & W_N^{(N-1)} & W_N^{(N-2)} & & W_N^1 \\ 1 & W_N^{(N-2)} & W_N^{(N-4)} & & W_N^2 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1 & W_N^1 & W_N^2 & & W_N^{(N-1)} \end{pmatrix} \qquad \text{(A.8)}$$

Hence, the eigenvectors of a circulant matrix form a linearly independent set, and so any circulant matrix can be diagonalized,

$$\Phi^{-1} A \Phi = \begin{pmatrix} \lambda_0 & & & \\ & \lambda_1 & & \\ & & \ddots & \\ & & & \lambda_{N-1} \end{pmatrix} \qquad \text{(A.9)}$$

Consider now the block circulant matrix A where the blocks A_k have dimensions $r \times s$,

$$A = \begin{pmatrix} A_0 & A_{N-1} & A_{N-2} & \cdots & A_1 \\ A_1 & A_0 & A_{N-1} & \cdots & A_2 \\ A_2 & A_1 & A_0 & \cdots & A_3 \\ \vdots & \vdots & \vdots & & \vdots \\ A_{N-1} & A_{N-2} & A_{N-3} & \cdots & A_0 \end{pmatrix} \tag{A.10}$$

The partitioned $rN \times rN$ matrix Φ_r is defined as

$$\Phi_r = \begin{pmatrix} I_r & I_r & I_r & \cdots & I_r \\ I_r & W_N^1 I_r & W_N^2 I_r & \cdots & W_N^{N-1} I_r \\ I_r & W_N^2 I_r & W_N^4 I_r & \cdots & W_N^{2(N-1)} I_r \\ \vdots & \vdots & \vdots & & \vdots \\ I_r & W_N^{N-1} I_r & W_N^{2(N-1)} I_r & \cdots & W_N^{(N-1)(N-1)} I_r \end{pmatrix} \tag{A.11}$$

where I_r is the $r \times r$ identity matrix. By analogy with (A.8), the inverse of Φ_r can be explicitly written as

$$\Phi_r^{-1} = \frac{1}{N} \begin{pmatrix} I_r & I_r & I_r & \cdots & I_r \\ I_r & W_N^{N-1} I_r & W_N^{N-2} I_r & \cdots & W_N^1 I_r \\ I_r & W_N^{N-2} I_r & W_N^{N-4} I_r & \cdots & W_N^2 I_r \\ \vdots & \vdots & \vdots & & \vdots \\ I_r & W_N^1 I_r & W_N^2 I_r & \cdots & W_N^{N-1} I_r \end{pmatrix} \tag{A.12}$$

In the case of 1×1 blocks, the matrix Φ or Φ_1 has the interpretation of the matrix of eigenvectors. The matrices Φ_r or Φ_s, of course, do not have this interpretation. Nevertheless, Φ_r and Φ_s can be used to block diagonalize the matrix A.

Proposition A.1. If A is a block circulant matrix with $r \times s$ blocks, then $\Phi_r^{-1} A \Phi_s$ is block diagonal,

$$\Lambda = \Phi_r^{-1} A \Phi_s$$

$$\begin{pmatrix} \bar{A}_0 & & & \\ & \bar{A}_1 & & \\ & & \bar{A}_2 & \\ & & & \ddots \\ & & & & \bar{A}_{N-1} \end{pmatrix}$$

where $\bar{A}_k = \sum_{i=0}^{N-1} A_i W_N^{-ki}$.

Proof: The (k,ℓ) block of $A\Phi_s$ is

$$(A\Phi_s)_{k,\ell} = \sum_{i=0}^{N-1} (A)_{k,i} (\Phi_s)_{i,\ell}$$

$$= \sum_{i=0}^{N-1} A_{k-i} W_N^{i\ell}$$

where $k,\ell = 0,1, \ldots, N-1$ and the indices are modulo N. The (m, ℓ) block of $\Phi_r^{-1} A \Phi_s$ can now be written,

$$(\Phi_r^{-1} A \Phi_s)_{m,\ell} = \sum_{k=0}^{N-1} (\Phi_r^{-1})_{m,k} (A\Phi_s)_{k,\ell}$$

$$= \sum_{k=0}^{N-1} \left(\frac{1}{N} W_N^{-mk}\right)\left(\sum_{i=0}^{N-1} A_{k-i} W_N^{i\ell}\right)$$

$$= \frac{1}{N}\sum_{k=0}^{N-1}\sum_{i=0}^{N-1} A_{k-i} W_N^{i\ell-mk}$$

$$= \frac{1}{N}\sum_{k=0}^{N-1}\sum_{n=0}^{N-1} A_n W_N^{(k-n)\ell-mk} \qquad , \quad n = k-i$$

$$= \frac{1}{N}\sum_{n=0}^{N-1} A_n W_N^{-n\ell}\left(\sum_{k=0}^{N-1} W_N^{k(\ell-m)}\right)$$

$$= \begin{cases} \sum_{n=0}^{N-1} A_n W_N^{-n} & , \ \ell = m \\ 0 & , \ \ell \neq m \end{cases}$$

Q.E.D.

The converse of Proposition A.1 is also true. Namely, if Λ is a block diagonal matrix, then $\Phi_r \Lambda \Phi_s^{-1}$ is block circulant. The details are omitted. Using this converse, it is easily shown that the product of two block circulant matrices A and B is also block circulant,

$$AB = \Phi_r\left[(\Phi_r^{-1} A \Phi_s)(\Phi_s^{-1} B \Phi_r)\right]\Phi_r^{-1} \tag{A.13}$$

$$= \Phi_r \Lambda_A \Lambda_B \Phi_r^{-1}$$

$$= \Phi_r \Lambda_{AB} \Phi_r^{-1}$$

where it is assumed that A has $r \times s$ blocks and B has $s \times r$ blocks. The matrix Λ_{AB} is block diagonal, since both Λ_A and Λ_B are, and so AB is block circulant. Similarly, the inverse of a block circulant matrix is still block circulant,

$$A^{-1} = \Phi_s(\Phi_s^{-1} A^{-1} \Phi_r)\, \Phi_r^{-1} \tag{A.14}$$

$$= \Phi_s(\Phi_r^{-1} A\, \Phi_s)^{-1}\, \Phi_r^{-1}$$

$$= \Phi_s\, \Lambda_A^{-1} \Phi_r^{-1}$$

Let C_r be the set of all block circulant matrices of order N having $r \times r$ blocks. Then, since the sum of two block circulant matrices is obviously a block circulant, C_r is an associative ring with respect to the operations of matrix addition and multiplication. The unit element is just the $rN \times rN$ identity matrix. C_1 is also a commutative ring, but none of the other C_r is.

APPENDIX B. DISCRETE-TIME SMOOTHING FORMULAS

Consider the discrete-time linear system

$$x(k+1) = \phi(k+1,k)x(k) + w(k) \tag{B.1}$$

$$y(k) = C(k)x(k) + v(k) \tag{B.2}$$

where $x(0)$, $w(k)$, and $v(k)$ are all independent, zero-mean, Gaussian random variables and

$$E\, x(0)x'(0) = \Sigma(0), \quad E\, w(k)w'(i) = Q(k)\delta_{i,k} \quad E\, v(k)v'(i) = R(k)\delta_{i,k}$$

The discrete-time FI smoothing problem is to compute the conditional expectation of $x(k)$ given the observations $\{y(i) \mid 0 \leq i \leq T\}$. The estimate is denoted $\hat{x}_s(k)$ and the error covariance is $P_s(k)$. The system covariance $Ex(k)x'(k)$ at time k is denoted by $\Sigma(k)$.

The reversed-time system corresponding to (B.1) is

$$\begin{aligned} x_r(k) &= [\,\Sigma(k)\phi'(k+1,k)\Sigma^{-1}(k+1)\,]\, x_r(k+1) + \xi(k+1) \\ &= \phi_r(k,k+1)x_r(k+1) + \xi(k+1) \end{aligned} \tag{B.3}$$

where the covariance of the reversed-time white noise driving process is

$$\begin{aligned} Q_r(k+1) &= E\, \xi_r(k+1)\xi'_r(k+1) \\ &= \Sigma(k) - \Sigma(k)\phi'(k+1,k)\Sigma^{-1}(k+1)\phi(k+1,k)\,\Sigma(k) \end{aligned} \tag{B.4}$$

If the state transition matrix $\phi(k+1,k)$ is invertible, this covariance may be written as

$$Q_r(k+1) = \Sigma(k)\phi'(k+1,k)\Sigma^{-1}(k+1)Q(k)\phi^{-1}(k+1,k)' \qquad (B.5)$$

$$= \phi_r(k,k+1)Q(k)\phi^{-1}(k+1,k)'$$

The random variables $x_r(T)$ and $\xi(k)$ are independent, zero-mean, Gaussian and $E\, x_r(T)x_r'(T) = \Sigma(T)$. Under these conditions, the processes $x(k)$ of (B.1) and $x_r(k)$ of (B.3) have the same covariance function and, therefore, the same joint probability density functions. It should be noted that Friedlander, Kailath, Ljung [85] and Sidhu, Desai [74] have previously given reversed-time realizations of the discrete-time process $x(k)$. The reversed-time systems in both these papers, however, are <u>incorrect</u>, and it is believed that (B.3) and (B.4) are the first correct equations for a discrete-time reversed-time realization.

For the smoothing problem, there are four disjoint sets of information about $x(k)$,

1) past observations: $\{y(i) \mid 0 \leq i < k\}$

2) future observations: $\{y(i) \mid k < i \leq T\}$

3) present observation: $\{y(k)\}$

4) *a priori* data: mean 0 and covariance $\Sigma(k)$

By grouping these sets of information in various ways, one arrives at the variety of estimated quantities below:

Kalman filter estimate	$\hat{x}(k\|k) = 1) + 3) + 4)$
Kalman filter one-step predictor	$\hat{x}(k\|k-1) = 1) + 4)$
reversed-time Kalman filter estimate	$\hat{x}_r(k\|k) = 2) + 3) + 4)$

reversed-time Kalman filter one-step predictor	$\hat{x}_r(k\|k+1) = 2) + 4)$
backwards estimate	$\hat{x}_b(k\|k) = 2) + 3)$
present conditional expectation	$\hat{x}_{p.a.p}(k\|k) = 3) + 4)$
a priori estimate	$\hat{x}_{a.p.}(k) = 4)$

and others. The reversed-time Kalman filter estimate and one-step predicted estimate are obtained from applying the Kalman filter equations to the reversed-time realization (B.3). The estimate $\hat{x}_{p.a.p.}(k)$ is just $E[x(k)|y(k)]$. The backwards estimate $\hat{x}_b(k|k)$ is used in the discrete-time Mayne-Fraser smoother.

The smoothed estimate, of course, must incorporate all four sets of information exactly once. There obviously exists a plethora of ways to combine these various estimates to obtain $\hat{x}_s(k)$,

$$\hat{x}_s(k) = \Sigma_s(k)[P^{-1}(k|k)\hat{x}(k|k) + P_b^{-1}(k|k+1)\hat{x}_b(k|k+1)] \tag{B.6A}$$

$$= \Sigma_s(k)[P^{-1}(k|k)\hat{x}(k|k) + P_r^{-1}(k|k+1)\hat{x}_r(k|k+1)] \tag{B.7A}$$

$$= \Sigma_s(k)[P^{-1}(k|k)\hat{x}(k|k) + P_r^{-1}(k|k)\hat{x}_r(k|k) - P_{p.a.p.}^{-1}(k|k)\hat{x}_{p.a.p.}(k|k)] \tag{B.8A}$$

$$= \Sigma_s(k)[P^{-1}(k|k-1)\hat{x}(k|k-1) + P_r^{-1}(k|k+1)\hat{x}_r(k|k+1) + P_{p.a.p.}^{-1}(k|k)\hat{x}_{p.a.p.}(k|k)] \tag{B.9A}$$

$$\Sigma_s(k) = [P^{-1}(k|k) + P_b^{-1}(k|k+1)]^{-1} \tag{B.6B}$$

$$= [P^{-1}(k|k) + P_r^{-1}(k|k+1) - \Sigma^{-1}(k)]^{-1} \tag{B.7B}$$

$$= [P^{-1}(k|k) + P_r^{-1}(k|k) - P_{p.a.p.}^{-1}(k|k)]^{-1} \tag{B.8B}$$

$$= [P^{-1}(k|k-1) + P_r^{-1}(k|k+1) + P_{p.a.p}^{-1}(k|k) - 2\Sigma^{-1}(k)]^{-1} \tag{B.9B}$$

The proof of these results is analogous to the proof of Theorem 4.3 and is omitted. Equation (B.6) is just the Mayne-Fraser two-filter smoother. Equation (B.7) expresses the smoothed estimate as a combination of two Kalman filter estimates. The other two formulas, (B.8) and (B.9), are included to show that the smoothed estimate can be written in terms of an expression that is symmetric with respect to forward- and reversed-time. There are, of course, many other possibilities besides (B.6)-(B.9) for giving the smoothed estimate.

For the reduced-order smoother analysis, the formula (B.7) will be used. Notice that this expression is asymmetric with respect to forward- and reversed-time, and so it is to be anticipated that the resulting reduced-order covariance expressions will also have some asymmetry. It is assumed that the model used in reduced-order smoothing is

$$x^*(k+1) = \phi^*(k+1,k)x^*(k) + w^*(k) \tag{B.10}$$

$$y^*(k) = C^*(k)x^*(k) + v^*(k) \tag{B.11}$$

$$z^*(k) = H^*(k)x^*(k) \tag{B.12}$$

The actual process $x(k)$ and observations $y(k)$ are generated by (B.1) and (B.2), and the actual output $z(k)$ is given by

$$z(k) = H(k)x(k) \tag{B.13}$$

The approach and assumptions are the same as in Section 4.4. Also, explicit dependence on time will be suppressed.

Forward-Time System and Filter - A reduced-order Kalman filter is designed on the basis of the model (B.10) and (B.11). Let K^* be the gain of this filter. Then consider the augmented state vector consisting of x and $\hat{x}^*$,

$$\begin{bmatrix} x(k+1) \\ \hat{x}^*(k+1|k+1) \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ K^*C\phi & (I-K^*C^*)\phi^* \end{bmatrix} \begin{bmatrix} x(k) \\ \hat{x}^*(k|k) \end{bmatrix} + \begin{bmatrix} I & 0 \\ K^*C & K^* \end{bmatrix} \begin{bmatrix} u(k) \\ v(k+1) \end{bmatrix} \tag{B.14}$$

Let

$$\begin{bmatrix} \Sigma & M \\ M' & N \end{bmatrix} = E\left\{ \begin{bmatrix} x \\ \hat{x}^* \end{bmatrix} \begin{bmatrix} x' & \hat{x}^{*\prime} \end{bmatrix} \right\} \tag{B.15}$$

This augmented state covariance is given by the discrete-time Lyapunov equation.

Reversed-Time System and Filter – Let K_r^* be the filter gain of the reduced-order, reversed-time Kalman filter. Then

$$\begin{bmatrix} x_r(k) \\ \hat{x}_r^*(k|k+1) \end{bmatrix} = \begin{bmatrix} \phi_r & 0 \\ \phi_r^* K_r^* C & \phi_r^* (I-K_r^* C^*) \end{bmatrix} \begin{bmatrix} x_r(k+1) \\ \hat{x}_r^*(k+1|k+2) \end{bmatrix} + \begin{bmatrix} I & 0 \\ 0 & \phi_r^* K_r^* \end{bmatrix} \begin{bmatrix} \xi_r(k+1) \\ v(k+1) \end{bmatrix} \tag{B.16}$$

Let the corresponding system covariance be

$$\begin{bmatrix} \Sigma & M_r \\ M_r' & N_r \end{bmatrix} = E\left\{ \begin{bmatrix} x_r \\ x_r^* \end{bmatrix} \begin{bmatrix} x_r' & x_r^{*\prime} \end{bmatrix} \right\} \tag{B.17}$$

Cross-Correlation of $x^*(k|k)$ and $x_r^*(k|k+1)$ – Using the discrete-time versions of the arguments given in Section 4.4 for continuous time yields

$$E\left[\hat{x}^*(k|k) x_r^*(k|k+1)\right] = \alpha^*(k)\Sigma(k)\beta^*(k) \tag{B.18}$$

where α^* and β^* are recursively computed from

$$\alpha^*(k) = K^*(k)C(k) + \{[I-K^*(k)C^*(k)]\phi^*(k,k-1)\}\alpha^*(k-1)\{\Sigma(k-1)\phi'(k,k-1)\Sigma^{-1}(k)\} \tag{B.19}$$

and

$$\beta^*(k) = \phi'(k+1,k)C(k)K_r^*(k+1)\Sigma^{*-1}(k+1)\phi^*(k+1,k)\Sigma^*(k) +$$

$$+ \{\phi'(k+1,k)\}\beta^*(k+1)\{\Sigma^*(k)\phi^{*\prime}(k+1,k)\Sigma^{*-1}(k+1)[I-K_r^*(k+1)C^*(k+1)]\}' \tag{B.20}$$

with initial conditions $\alpha^*(-1) = \beta^*(T) = 0$.

The smoothed output error covariance is therefore

$$\text{cov}[z(k) - z^*(k)] = [H - H^*] \; E\left\{\begin{bmatrix} x(k) \\ \hat{x}_s^*(k) \end{bmatrix}[x'(k) \quad x_s^{*\prime}(k)]\right\}\begin{bmatrix} H' \\ H^{*\prime} \end{bmatrix} \tag{B.21}$$

$$= H\Sigma H' - H^*E\{\hat{x}_s^* x'\}H' - H\,E\{x\hat{x}_s^{*\prime}\}\,H^{*\prime} + H^*E\{\hat{x}_s^*\,\hat{x}_s^{*\prime}\}\,H^{*\prime}$$

Using (B.7) for the smoothed estimate,

$$E\{x\,\hat{x}_s^{*\prime}\} = E\{x[\hat{x}^{*\prime}P^{*-1} + \hat{x}_r^{*\prime}\,P_r^{*-1}]\,\Sigma_s^*\} \tag{B.22}$$

$$= [MP^{*-1} + M_rP_r^{*-1}]\,\Sigma_s^*$$

and

$$E\{\hat{x}_s^*\,\hat{x}_s^{*\prime}\} = E\{\Sigma_s^*[P^{*-1}\hat{x}^* + P_r^{*-1}\,\hat{x}_r^*]\,[\hat{x}^{*\prime}P^{*-1} + \hat{x}_r^{*\prime}P_r^{*-1}]\,\Sigma_s^*\} \tag{B.23}$$

$$= \Sigma_s^*[P^{*-1}NP^{*-1} + P_r^{*-1}\beta^{*\prime}\Sigma\alpha^{*\prime}P^{*-1} + P^{*-1}\alpha^*\Sigma\beta^*P_r^{*-1} +$$

$$+ P_r^{*-1}N_rP_r^{*-1}]\Sigma_s^*$$

Substituting (B.22) and (B.23) into (B.21) yields

$$\mathrm{cov}[z(k)-z^*(k)] = H\Sigma H' - H^*\Sigma_s^*[P^{*-1}M' + P_r^{*-1}M_r']H' - H[MP^{*-1}+M_rP_r^{*-1}]\Sigma_s^*H^{*'} +$$

$$+ H^*\Sigma_s^*[P^{*-1}NP^{*-1} + P_r^{*-1}\beta^{*'}\Sigma\alpha^{*'}P^{*-1} + P^{*-1}\alpha^*\Sigma\beta^*P_r^{*-1} + P_r^{*-1}N_rP_r^{*-1}]\Sigma_s^*H^{*'} \qquad (B.24)$$

The sensitivity analysis problem is solved by taking H and H* equal to the identity matrix.

REFERENCES

[1] Athans, M. (Editor), Special Issue on Large-Scale Systems and Decentralized Control, in: IEEE Trans. Automat. Contr., AC-23, #2, April 1978.

[2] Guardabassi, G., and Locatelli, A. (Editors), Large-Scale Systems Theory and Applications, Udine, Italy, 16-20 June 1976, (Pittsburgh, PA, USA: Instrument Soc. of America, 1976).

[3] Proceedings of the Sixth Triennial World Congress of the International Federation of Automatic Control, Boston and Cambridge, MA , USA, 24-30 August 1975, (Pittsburgh, PA, USA: Instrument Society of America, 1975).

[4] Davison, E.J. and Tripathi, N.K., "The optimal decentralized control of a large power system: lead and frequency control", IEEE Trans. Automat. Contr., AC-23, #2, pp. 312-324, April 1978.

[5] Jocic, Lj.B., Ribbens-Pavella, M., Siljak, D.D., "Multimachine power systems: stability, decomposition, and aggregation", IEEE Trans. Automat. Contr ., AC-23, #2, pp. 325-332, April 1978.

[6] Arafeh, S.A., "Hierarchical control of power distribution systems", IEEE Trans. Automat. Contr., AC-23, #2, pp. 333-343, April 1978.

[7] Houpt, P.K. Decentralized Stochastic Control of Finite State Systems with Application to Vehicular Traffic Flow, Ph.D. dissertation, Mass. Inst. Tech., Cambridge, MA, November 1974.

[8] Looze, D.P., Decentralized Control of a Freeway Traffic Corridor, M.S. thesis, Mass. Inst. Tech., Cambridge, MA , August 1974.

[9] Looze, D.P., Houpt, P.K. Sandell, N.R., Athans, M., "On decentralized estimation and control with application to freeway ramp metering", IEEE Trans. Automat. Contr., AC-23, #2, pp. 268-275, April 1978.

[10] Chow, G.C., and Megdal, S.B., "The control of a large-scale nonlinear econometric system", IEEE Trans. Automat. Contr., AC-23, #2, pp. 344-349, April 1978.

[11] Aoki, M., Optimal Control and Systems Theory in Dynamic Economic Analysis, North Holland/American Elsevier, New York, N.Y. (1976).

[12] Schoute, F.C., "Decentralized control in packet switched satellite communication", IEEE Trans. Automat. Contr., AC-23, #2, pp. 362-371, April 1978.

References (continued)

[13] Sandell, N.R., and Athans, M., "Relativistic control theory and the dynamic control of communication networks", Proc. 1975 Decision and Control Conference, Houston, TX, December 1975.

[14] Segall, A., "New analytical models for dynamic routing in computer networks", National Telecommunication Conference, New Orleans, LA , December 1975.

[15] Haddad, A.H., and Kokotovic, P.V., "On singular perturbations in linear filtering and smoothing", Proc. Fifth Symp. Nonlinear Estimation and Applications, September 1974.

[16] Haddad, A.H., "Linear filtering of singularly perturbed systems", IEEE Trans. Automat. Contr., AC-21, #4, pp. 515-519, August 1976.

[17] Teneketzis, D., and Sandell, N.R., "Linear regulator design for stochastic systems by a multiple time-scales method", IEEE Trans. Automat. Contr., AC-22, #4, pp. 615-621, August 1977.

[18] Haddad, A.H., and Kokotvic, P.V., "Stochastic control of linear singularly perturbed systems", IEEE Trans. Automat. Contr., AC-22, #5, pp. 815-821, October 1976.

[19] Ramapriyan, H.K., Study of Coupling in Interconnected Systems, Ph.D. dissertation, Univ. of Minnesota, Minneapolis, MN , March 1970.

[20] Bailey, F.N., and Ramapriyan, H.K., "Bounds on suboptimality in the control of linear dynamic systems", IEEE Trans. Automat. Contr., AC-18, #5, pp. 532-534, October 1973.

[21] Teneketzis, D., Perturbation Methods in Decentralized Stochastic Control, M.S. thesis, Mass. Inst. Tech., Cambridge, MA, June 1976.

[22] Teneketzis, D., and Sandell, N.R., "Perturbation methods in decentralized stochastic control", 1st Int. Conf. on Information Sciences and Systems, Patras, Greece, August 1976.

[23] Sandell, N.R., Varaiya, P., Athans, M., Safonov, M.G., "Survey of decentralized control methods for large scale systems", IEEE Trans. Automat. Contr., AC-23, #2, pp. 108-128, April 1978.

[24] Sandell, N.R., "Decomposition vs. decentralization in large scale system theory", 1976 IEEE Conf. on Decision and Control, Clearwater Beach, FL , December 1976.

[25] Habibi, A., "Two-dimensional Bayesian estimate of images", Proc. IEEE, 60, #7, pp. 878-883, July 1972.

References (continued)

[26] Jain, A.K., and Angel, E., "Image restoration, modelling and reduction of dimensionality", IEEE Trans. Comput., C-23, #5, pp. 470-476, May 1974.

[27] Jain, A.K., "Image coding via a nearest neighbor image model", IEEE. Trans. Commun., COM-23, #3, pp. 318-331, March 1975.

[28] Attasi, S., "Modeling and recursive estimation for double indexed sequences", IRIA Rept. IA/129, Domaine de Voluceau, Rocquencourt, 78150 Le Chesnay, B.P. 5, France, July 1975.

[29] Willsky, A.S., "Digital signal processing and control and estimation theory -- points of tangency, area of intersection, and parallel directions", M.I.T. Electronic Systems Lab. Report ESL-R-712, Cambridge, MA, January 1976.

[30] Melzer, S.M., and Kuo, B.C., "Optimal regulation of systems described by a countably infinite number of objects", Automatica, 7, #3, pp. 359-366, May 1971.

[31] Chu, K.C, "Optimal decentralized regulation for a string of coupled systems", IEEE. Trans. Automat. Contr., AC-19, #3, pp. 243-246, June 1974.

[32] Dickerson, J.R., and Erickson, P., "Stability of cyclic dynamical systems", J. Dyn. Syst., Meas. and Contr., Trans. ASME, Series G, 96, #1, pp. 393-398, December 1975.

[33] Biglieri, E., and Elia, M., "Cyclic-group codes for the Gaussian channel", IEEE Trans. Info. Th., IT-22, #5, pp. 624-629, September 1976.

[34] Karlin, M., "New binary coding results by circulants", IEEE Trans. Info. Th., IT-15, #1, pp. 81-92, January 1969.

[35] Mullis, C.T., and Steiglitz, K., "Circulant Markov chains as digital signal sources", IEEE Trans. Audio and Electroacous., AU-20, #4, pp. 246-248, October 1972.

[36] Willems, J.L., "Generalized Clarke components for polyphase networks", IEEE Trans. Educ., E-12, #1, pp. 69-71, March 1969.

[37] Berlin, T.H., and Kac, M., "The spherical model of a ferromagnet", Phys. Rev., 86, #6, pp. 821-835, June 15, 1952.

References (continued)

[38] Ekstrom, M.P., "An iterative-improvement approach to the numerical solution of vector Toeplitz systems", IEEE Trans. Comput., C-23, #3, pp. 320-325, March 1974.

[39] Hunt, B.R., "The application of constrained least-square estimation to image restoration by digital computer", IEEE Trans. Comput., C-22, #9, pp. 805-812, September 1973.

[40] Hunt, B.R., "Digital image processing", Proc. IEEE, 63, #4, pp. 693-708, April 1975.

[41] Mayne, D.Q., "A solution of the smoothing problem for linear dynamic systems", Automatica, 4, #2, pp. 73-92, November 1966.

[42] Fraser, D.C., A New Technique for the Optimal Smoothing of Data, Sc.D. dissertation, Mass. Inst. Tech., Cambridge, MA, January 1967.

[43] Mehra, R.K., Studies in Smoothing and in Conjugate Gradient Methods Applied to Optimal Control Problems, Ph.D. dissertation, Harvard University, Cambridge, MA, May 1967.

[44] Bellman, R., Introduction to Matrix Analysis, McGraw-Hill, Inc., New York , N.Y., (1970).

[45] Brockett, R.W., and Willems, J.L., "Least-square optimization for stationary linear partial differential equations", Proc. 1971 IFAC Symp. on Cont. of Dist. Paramet. Syst., Banff, Alta., Canada, June 21-23, 1971.

[46] Sneddon, I.N., Elements of Partial Differential Equations, McGraw-Hill, Inc., New York, N.Y. (1957).

[47] Dickerson, J.R., and Erickson, P., "Stability of a chain of linear differential equations", J. of Math. Anal. and Appl., 47, #3, pp. 513-519, September 1974.

[48] Michel, A.N., "Stability analysis of interconnected systems", SIAM J. Contr., 12, #3, pp. 554-579, August 1974.

[49] Arcidiacono, V., Ferrari, E., and Saccomanno, F., "Studies on damping of electromechanical oscillations in multimachine systems with longitudinal structure", IEEE Trans. Power App. and Syst., PAS-95, #2, pp. 450-460, March 1976.

[50] Elgerd, O. I., Electric Energy Systems Theory: An Introduction, McGraw-Hill, Inc., New York, N.Y. (1971).

References (continued)

[51] Forsythe, G.E. and Wasow, W., Finite Difference Methods for Partial Differential Equations, John Wiley & Sons, New York, N.Y. (1960).

[52] Kalman, R.E., Falb, P.L., and Arbib, M.A., Topics in Mathematical System Theory, McGraw-Hill, Inc., New York, N.Y. (1969).

[53] Brockett, R.W., Finite Dimensional Linear Systems, John Wiley & Sons, New York, N.Y. (1970).

[54] Kleinman, D.L., and Athans, M., "The design of suboptimal linear time-varying systems", IEEE Trans. Automat. Contr., AC-13, #2, pp. 150-158, April 1968.

[55] Levine, W.S., and Athans, M., "On the determination of the optimal constant output-feedback gains for linear multivariable systems", IEEE Trans. Automat. Contr., AC-15, #1, pp. 44-48, February 1970.

[56] Kosut, R.L., "Suboptimal control of linear time-invariant systems subject to control structure constraints", IEEE Trans. Automat. Contr., AC-15, #5, pp. 557-563, October 1970.

[57] Chong, C.Y., and Athans, M., "On the stochastic control of linear systems with different information sets", IEEE Trans. Automat. Contr., AC-16, #5, pp. 423-430, October 1971.

[58] Levine, W.S., Johnson, T.L., and Athans, M., "Optimal limited state feedback controllers for linear systems", IEEE Trans. Automat. Contr., AC-16, #6, pp. 785-793, December 1971.

[59] Wang, F.C.H., Suboptimization of Decentralized Control Systems, Ph.D. dissertation, Univ. of Minnesota, Minneapolis, MN, 1972.

[60] Cohen, G., "Commande par contre réaction, sous constrainte de structure du gain, d'un systèm stochastique linéaire", R.A.I.R.O., J-3, November 1975.

[61] Looze, D.P., Hierarchical Control and Decomposition of Decentralized Linear Stochastic Systems, Ph.D. dissertation, Mass. Inst. Tech., Cambridge, MA, August 1978.

[62] Oppenheim, A.V., and Schafer, R.W., Digital Signal Processing, Prentice-Hall, Englewood Cliffs, N.J. (1975).

[63] Creedon, J.F., and Lindgren, A.G., "Control of the optical surface of a thin deformable primary mirror with application to an orbiting astronomical observatory", Proc. 3rd IFAC Symp. Automat. Contr. in Space, Toulouse, France, March 2, 1970.

References (continued)

[64] McClamroch, N.H., "Evaluation of suboptimality and sensitivity in control and filtering processes", IEEE Trans. Automat. Contr., AC-14, #3, pp. 282-285, June 1969.

[65] Ljung, L., and Kailath, T., "A unified approach to smoothing formulas", Automatica, 12, #2, pp. 147-157, March 1976.

[66] Meditch, J.S., "A survey of data smoothing for linear and nonlinear dynamic systems", Automatica, 9, #2, pp. 151-162, March 1973.

[67] Schweppe, F.C., Uncertain Dynamic Systems, Prentice-Hall, Inc., Englewood Cliffs, N.J. (1973).

[68] Fraser, D.C., and Potter, J.E., "The optimum linear smoother as a combination of two optimum linear filters", IEEE Trans. Automat. Contr., AC-14, #4, pp. 387-390, August 1969.

[69] Rauch, H.E., Tung, F., and Striebel, C.T., "On the maximum likelihood estimates for linear dynamic systems", AIAA Journal, 3, #8, pp. 1445-1450, August 1965.

[70] Kailath, T., and Frost, P., "An innovations approach to least-squares estimation — Part II: linear smoothing in additive white noise", IEEE Trans. Automat. Contr., AC-13, #6, pp. 646-660, December 1968.

[71] Ljung, L., and Kailath, T., "Backward Markovian models for second-order stochastic processes", IEEE Trans. Info. Th., IT-22, #4, pp. 488-491, July 1976.

[72] Nishimura, R., "Error bounds of continuous Kalman filters and the application to orbit determination problems", IEEE Trans. Automat. Contr., AC-12, #3, pp. 268-275, June 1967.

[73] Fitzgerald, R.J., "Error divergence in optimal filtering problems", Proc. 2nd IFAC Symp. on Automat. Contr. in Space, Vienna, Austria, September 1967.

[74] Sidhu, G.S., and Desai, U.B., "New smoothing algorithms based on reversed-time lumped models", IEEE Trans. Automat. Contr., AC-21, pp. 538-541, August 1976.

[75] Lainiotis, D.G., "General backward Markov models", IEEE Trans. Automat. Contr., AC-21, #4, pp. 595-598, August 1976.

[76] Schoute, F.C., Ter Horst, M.F., and Willems, J.L., "Hierarchic recursive image enhancement", IEEE Trans. Circ. and Syst., CAS-24, #2, pp. 67-78, February 1977.

[77] Kailath, T., Lectures on Linear Least-Squares Estimation, unpublished notes.

References (continued)

[78] Ljung, L., and Kailath, T., "Efficient change of initial conditions, dual Chandrasekhar equations, and some applications", IEEE Trans. Automat. Contr., AC-22, #3, pp. 443-447, June 1977.

[79] Griffin, R.E., and Sage, A.P., "Sensitivity analysis of discrete filtering and smoothing algorithms", AIAA Journal, 7, #10, pp. 1890 1897, October 1969.

[80] Jury, E.I., Theory and Application of the z-transform Method, John Wiley & Sons, New York, N.Y. (1964).

[81] Widom, H., Toeplitz Matrices, in: Studies in Real and Complex Analysis, I. I. Hirschman, Jr., Ed., Prentice-Hall, Inc., Englewood Cliffs, N.J., (1965).

[82] Hager, W.W., and Horowitz, L.L., "Convergence and stability properties of the discrete Riccati operator equation and the associated optimal control and filtering problems", SIAM J. Contr. and Opt., 14, #2, pp. 295-312, February 1976.

[83] Luenberger, D.G., Optimization by Vector Space Methods, John Wiley & Sons, New York, N.Y. (1969).

[84] Anderson, B.D.O., "An algebric solution to the spectral factorization problem, IEEE Trans. Automat. Contr., AC-12, #4, pp. 410-414, August 1967.

[85] Friedlander, B., Kailath, T., and Ljung, L., "Scattering theory and linear least-squares estimation — II. Discrete-time problems", J. Franklin Inst., 301, #1 & 2, pp. 71-82, January 1976.

BIOGRAPHICAL NOTE

Joseph Edward Wall, Jr., was born on April 5, 1951, in New Orleans, Louisiana. He graduated as salutatorian from Chalmette High School in May, 1969, and entered the Tulane University of Louisiana in September, 1969, as a National Merit Scholar. In May, 1973, he received (from the Department of Electrical Engineering) the degree of Bachelor of Science in Engineering, with honors for his undergraduate research on the electrical, magnetic, and spectroscopic properties of $CsNiI_3$. At this same time, he received the C. W. Ricker Award as the outstanding senior in the Department. While an undergraduate, Mr. Wall was elected to Tau Beta Pi and Phi Eta Sigma and served as an officer in Tau Beta Pi (president, 1973) and the IEEE student branch (vice chairman, 1973 and secretary, 1972).

Mr. Wall entered the Massachusetts Institute of Technology in the Department of Electrical Engineering as a full-time graduate student in September, 1973. He was awarded the degrees of Master of Science in Electrical Engineering and Electrical Engineer jointly in February, 1976. His S.M. thesis was entitled "Determination of Neonatal State by Computer". His graduate program at M.I.T. has been supported by a National Science Foundation Fellowship (1973-1976), a teaching assistantship (1976), a research assistantship (1977-1978) and a Vinton Hayes Fellowship (1978).

His professional experience includes summer positions with New Orleans Public Service, Inc. (1971 and 1972) modeling the underground electrical network, the Atlantic Richfield Company (1973 and 1974) performing digital signal processing of seismic signals, and The Analytic Sciences Corporation (1977) developing stochastic guidance laws for air-to-air missiles.

Mr. Wall is married to the former Barbara Louise Bienvenu of Chalmette, Louisiana, and they have two children, a son Joey and a daughter Susan.