

**THE SPECTRAL PROPERTIES OF MARKOV CHAIN
RANDOM PROCESSES WITH APPLICATION
TO SPEECH MODELING AND SPREAD
SPECTRUM COMMUNICATIONS**

By

WILLIAM SCOTT KING

**Bachelor of Science
Oklahoma State University
Stillwater, Oklahoma
1987**

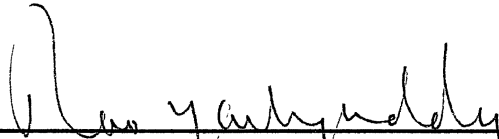
**Master of Science
Oklahoma State University
Stillwater, Oklahoma
1989**

**Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 1991**

Thesis
1991D
K54A

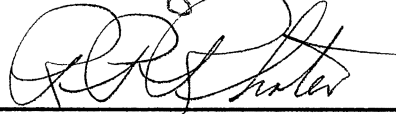
THE SPECTRAL PROPERTIES OF MARKOV CHAIN
RANDOM PROCESSES WITH APPLICATION
TO SPEECH MODELING AND SPREAD
SPECTRUM COMMUNICATIONS

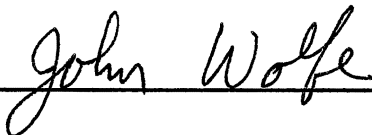
Thesis Approved:




Thesis Adviser









Dean of the Graduate College

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Rao Yarlagadda, for his guidance and counseling throughout my graduate studies at Oklahoma State University. His comments and suggestions were always relevant and appreciated. I am also grateful to Dr. John Hershey for his ideas concerning this research. Dr. Ron Rhoten, Dr. George Scheets, and Dr. John Wolfe deserve thanks as well for their time and effort serving on my thesis committee. Additionally, I am indebted to the Department of Defense under contract number MCA 904-88-6017 and Sandia National Laboratories under contract number 63-0906 for their funding of this research.

To the students, faculty, and staff at Oklahoma State University go my deepest appreciation for their friendship. Last but certainly not least, I would like to thank my parents and family for their constant and continuing encouragement. Without such wonderful people it would not have been possible or worthwhile.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
Markov Processes	1
Literature Survey	4
Overview	6
II. MARKOV CHAIN RANDOM PROCESSES	9
Markov Chains	9
Markov Chain Notation	10
Stationary Markov Chains	15
Markov Chain Random Processes	19
Implementation Aspects of MCRPs	20
First Order Statistics of MCRPs	22
Autocorrelation of MCRPs	25
Power Spectrum of MCRPs: General Expression	31
Power Spectrum of MCRPs: Diagonable Transition Probability Matrix	37
Power Spectrum of MCRPs: Non-Diagonable Transition Probability Matrix	39
Hidden Markov Models	41
III. SYNTHESIS OF MARKOV CHAIN RANDOM PROCESSES	43
Probability Distribution Synthesis	43
Power Spectral Density Synthesis	47
A Stochastic Eigenvalue Problem	49
A Circulant Structure	52
A Summed Structure	57
A Ring Structure	59
Joint Probability Distribution - Power Spectrual Density Synthesis ..	73
IV. STOCHASTIC MODELING WITH MARKOV CHAIN RANDOM PROCESSES	78
Stationary Processes	78
ARMA Processes	79
ARMA Processes and MCRPs	80
Parameter Solutions for ARMA Processes	82
Parameter Solutions for MCRPs	84
AR Processes and Ring-Structured MCRPs	86
Nonstationary Processes	94

Chapter	Page
V. SPEECH MODELING WITH MARKOV CHAIN RANDOM PROCESSES	96
Speech and Standard Speech Models	97
MCRP Model of Speech	100
Examples	102
VI. SPREAD SPECTRUM COMMUNICATIONS	110
Direct Sequence Spread Spectrum	111
MCRP Spread Spectrum	115
Complementary MCRPs	118
Effects of Partial-Band Jamming	125
Probability of Bit Error for DS-SS	126
Probability of Bit Error for MCRP-SS	131
Comparative Results	135
VII. MARKOV RANDOM FIELDSS	141
VII. CONCLUSIONS	149
Summary	149
Considerations for Future Research	153
REFERENCES	156
APPENDIX - DERIVATION OF SPECTRAL PARAMETERS FOR RING-STRUCTURED MARKOV CHAIN RANDOM PROCESSES	160

LIST OF TABLES

Table		Page
4.1.	Comparison of ARMA Processes and MCRPs	82
5.1.	Comparison of LPC and MCRP Models	102
7.1.	Estimate of Transition Probabilities	147
7.2.	Estimate of Second Pass Transition Probabilities	148

LIST OF FIGURES

Figure	Page
1.1. Example Outcomes for ARMA Process and MCRP Examples	3
1.2. Power Spectra of ARMA Process and MCRP	4
2.1. 3-State Markov Chain Flow Graph	11
2.2. An MCRP with 3 States	19
2.3. State Decision Function	21
2.4. Autocorrelation of Example MCRP	30
2.5. Power Spectrum of Example MCRP	36
2.6. Common Hidden Markov Model Structures	42
3.1. Waveforms for Laplacian Process and MCRP	46
3.2. Variance and Skew of MCRP as a Function of N	47
3.3. Block Diagram of Power Spectrum Synthesis	48
3.4. Domain of Eigenvalues for Stochastic $N \times N$ Matrices	50
3.5. Domain of Eigenvalues for Stochastic and Circulant Stochastic $N \times N$ Matrices	57
3.6. MCRP Ring Structure	60
3.7. Power Spectrum for Example Ring-Structured MCRP	66
3.8. Power Spectrum of Synthesis Example	73
3.9. MCRP Power Spectrum from Joint Synthesis	77
4.1. MCRP Modeling Strategy	86
4.2. Power Spectra for AR Process and MCRP	88
4.3. AR and MCRP Power Spectra	90

Figure	Page
4.4. Difference in AR and MCRP Peak Frequencies. Pole Magnitudes: 0.2, 0.4, and 0.8	92
4.5. AR Power Spectrum and MCRP Power Spectrum for Two Pole-Pair Example	94
5.1. Speech Waveforms: Unvoiced /sh/ Sound from the Word 'Show,' and Voiced /o/ Sound from the Word 'Show'	99
5.2. Block Diagram of MCRP Model of Speech	101
5.3. Power Spectra of AR and MCRP for Various Phonemes	104
5.4. Original and Synthesized Waveforms for 'Off'	106
5.5. Original and Synthesized Waveforms for 'Ash'	107
5.6. Original and Synthesized Waveforms for 'Us'	108
5.7. Original and Synthesized Waveforms for 'Eat'	109
6.1. Block Diagram of DS-SS System	112
6.2. Power Spectrum of Signal without Spreading and Signal with Spreading, for $G_p = 4$	115
6.3. Block Diagram of MCRP-SS System	117
6.4. Power Spectra of Discrete-Time Complementary Processes	119
6.5. Resulting Signal Spectrum and Added-Noise Spectrum	121
6.6. Power Spectra of Complementary MCRPs: $\Phi_1(\omega)$, $\Phi_2(\omega)$, and $\Phi_{\text{sum}}(\omega)$	124
6.7. DS-SS and MCRP-SS Power Spectra	125
6.8. Spectral Magnitude Variation Versus $\Phi_1(0)$	137
6.9. P_b for DS-SS and MCRP-SS with $E_b/N_o = 6\text{dB}$	138
6.10. P_b for DS-SS and MCRP-SS with $E_b/N_o = 9\text{dB}$	139
6.11. P_b for DS-SS and MCRP-SS with $E_b/N_o = 12\text{dB}$	140
7.1. Neighborhoods $\eta_{i,j}^1$ and $\eta_{i,j}^2$	143
7.2. Original and First Pass MRF Images	147
7.3. First and Second Pass MRF Images	148

NOMENCLATURE

a_i	Output map associated with state i
\mathbf{a}	Vector of a_i 's
\bar{a}_i	$a_i - \mu$
$\bar{\mathbf{a}}$	Vector of \bar{a}_i 's
β_k	Spectral coefficient of eigenvalue λ_k
\mathbf{B}_N	Domain of eigenvalues for circulant stochastic matrices
\mathbf{C}	Circulant matrix
D_N	Region of complex, N -dimensional space containing vectors of eigenvalues
$d(t)$	Continuous-time, ± 1 -valued data signal
λ_k	Eigenvalue
Λ	Diagonal matrix of eigenvalues
\mathbf{M}_N	Domain of eigenvalues for stochastic matrices
μ	Mean of the process
$\underline{\mu}$	Vector where every element is μ
N	Number of states
$\underline{\mathbf{1}}_N$	Vector of length N where every element is 1
P_b	Probability of bit error
\mathbf{P}_i	Polygonal region in complex space
$\phi(n)$	Autocorrelation of discrete-time signal
$\Phi(\omega)$	Power spectrum of discrete-time signal
$\Phi(z)$	Z-transform of autocorrelation of discrete-time signal

$s_i(n)$	State probability for i^{th} state at time n
$\mathbf{s}(n)$	Vector of $s_i(n)$'s
$s(t)$	Continuous-time signal
s_i	Stationary state probability for i th state
\mathbf{s}	Vector of stationary state probabilities
\mathbf{S}	Diagonal matrix of stationary state probabilities
$S(f)$	Power spectrum of continuous-time signal
T	Stochastic matrix
T^∞	$\lim_{n \rightarrow \infty} T^n$ for an irreducible stochastic matrix
$t_{i,j}$	$[T]_{i,j}$
\mathbf{u}	Left eigenvector: $\mathbf{u}^T T = \lambda \mathbf{u}^T$
\mathbf{v}	Right eigenvector: $T \mathbf{v} = \lambda \mathbf{v}$
V	Modal matrix

CHAPTER I

INTRODUCTION

Markov Processes

One of the most useful classes of random processes is the class of *Markov processes*. In a Markov process, the current state of the process is only influenced by the most recent behavior of the process. In short, the process has a finite memory. This *Markov property* is not an overwhelming restriction on a random process. For a great number of random processes found in nature, the assumption of finite memory makes good sense.

As an example, consider the following difference equation representing a discrete-time random process:

$$y(n) = 0.8y(n - 1) + u(n) \quad (\text{I.1})$$

Here, the random process $\{y(n)\}$ is driven by $\{u(n)\}$, which is also a random process.

For this example, let us assume that $\{u(n)\}$ is a white noise process with unity variance and Gaussian density. From the equation, it is clear that the process remembers only the previous value of y : $y(n)$ is not directly influenced by $\{y(n - 2), y(n - 3), \dots\}$. Hence, the random process $\{y(n)\}$ is a Markov process. Actually, this is an example of an autoregressive, moving-average (ARMA) process. ARMA processes have some very nice properties, such as linearity, which have proved valuable in modeling a wide range of processes.

Another example of a Markov process is a *Markov chain*. A Markov chain is essentially a state system in which the state transitions are defined by a set of probabilities. As a simple example, suppose there are two states, numbered 1 and 2. Let the probability that the process does not change state be 0.9; thus, there is a probability of 0.1 that the process will change states. Now, if the process is in state 1 at a given time, consider the next state of the process. Clearly, this only depends upon the transition probabilities, and not upon any of the previous states which the chain occupied. Hence, this is also a Markov process.

Although the two examples given above are both Markov processes, there are clear differences between them. In the first example, the random process assumes real numbers; e.g., $y(2) = 0.34$. However, the second example is a state system. At a given time, the process occupies a particular state. But this state can be practically anything. For example, a state could represent a quality such as excellent, good or poor; a condition such as on or off; or even a number such as π . It all depends upon the application.

The focus of this dissertation is Markov chains for which the states represent real numbers. This type of process has been called a Markov chain random process (MCRP). For the two-state Markov chain described above, one could create an MCRP by assigning the states real numbers. As an example, state 1 could represent the number 1.667, and state 2 the number -1.667.

To provide some insight, let us compare the MCRP and ARMA examples. It has been noted that both processes are Markov, but that there are differences between them. To illustrate, Figure 1.1 shows sample outcomes for both processes. The most notable difference is that the ARMA process assumes a wide range of values, whereas the MCRP only assumes ± 1.667 . In general, ARMA processes are continuous-valued whereas MCRPs are discrete-valued.

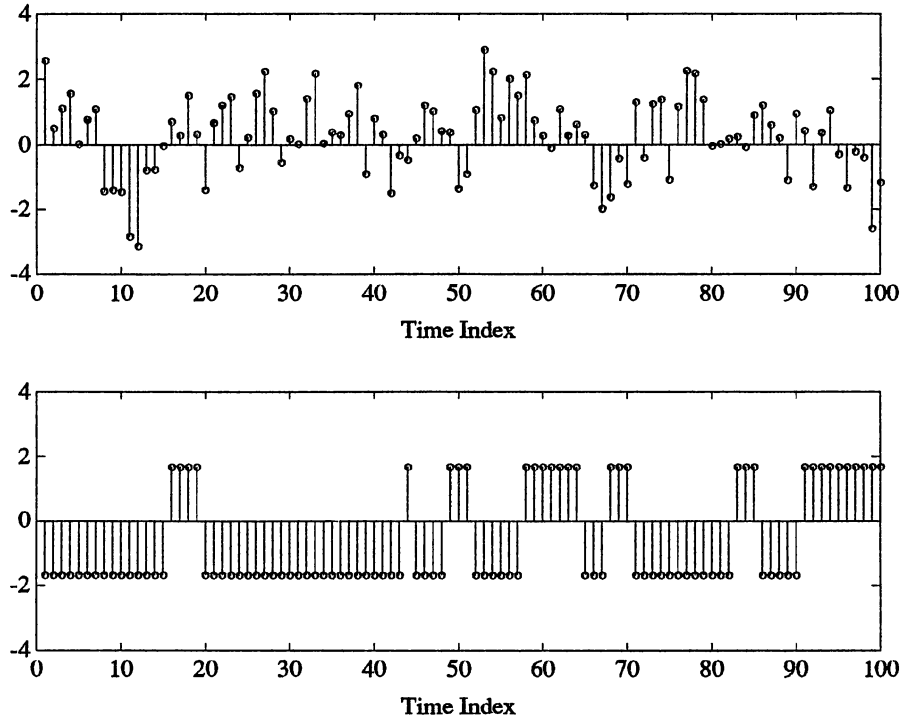


Figure 1.1. Example outcomes for ARMA process and MCRP examples.

Let us now look at the power spectra of these processes. The details of the power spectra shall be examined in Chapter II. However, not worrying about the details of this operation for now, the results are plotted in Figure 1.2. As can be seen, the two processes have identical power spectra. This is perhaps surprising because in the time-domain, as was illustrated in Figure 1.1, the processes look nothing alike. However, the fact that their power spectra are identical implies that, at least on some level, the two processes are similar. More importantly, it gives us insight into the statistical properties of MCRPs.

The goal of this dissertation is to explore MCRPs. The power spectra of these processes is a main concern both from an analysis and synthesis viewpoint. The relationship between ARMA processes and MCRPs is examined in light of the previous example.

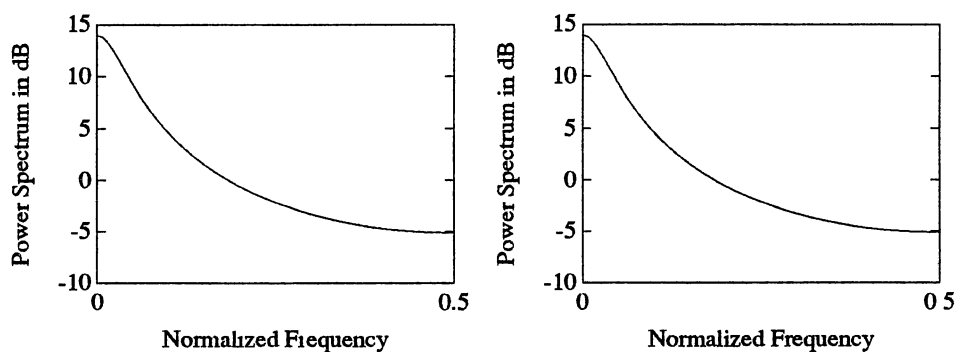


Figure 1.2. Power Spectra of ARMA process and MCRP.

Additionally, several applications of MCRPs shall be considered. Overall, the objective of this research is to examine the distinguishing characteristics of MCRPs and to utilize these processes in several applications.

Literature Survey

Markov chains have been extensively studied and applied to a wide number of fields, including operations research, queueing networks, biology (life/death chains), gambling and many others [Kem76, Bha60]. The success of Markov chains is a testament to the fact that many processes are essentially state systems. However, while the literature on Markov chains is extensive, most have little bearing on this study of MCRPs.

More relevant is the work of Sittler, who introduced Markov chain random processes in the 1950's [Sit54, Sit56]. Sittler's work was based on flow graphs and the relation of MCRPs to linear discrete-time systems and circuits. He found that the tools used for analyzing circuits could be used to analyze MCRPs. Proceeding in this manner, he

determined the power spectral density for MCRPs. In fact, although relatively little work has been done with Markov chain random processes, most of it has been concerned with the power spectrum.

Since Sittler's initial papers, several others have analyzed the spectral properties of MCRPs. These include Huggins [Hug57], and Zadeh [Zad57], who generalized Sittler's approach to processes with time-varying parameters. Additionally, some researchers, including Barnard [Bar64], have analyzed MCRPs which generate continuous-time signals by mapping each state to a waveform instead of a real number. To date, the most general treatment of MCRPs appears to be by Yoshida [Yos73].

After the spectral properties of MCRPs had been thoroughly analyzed, a few researchers began to investigate the spectral synthesis problem for MCRPs. That is, they wished to determine the MCRP parameters which result in a given power spectrum. However, this is a very difficult problem and a general solution has not been found. Nevertheless, some researchers have had limited success by considering special classes of MCRPs. One of the earlier works in this area was performed by Mullis and Steiglitz [Mul72], their approach is based on MCRPs with circulant transition probability matrices. Others have tried similar approaches, most notable Steinhardt, who considered MCRPs with normal transition probability matrices. However, these approaches offer only a partial solution to the synthesis problem.

There have been several applications of MCRPs in the literature. These include analysis of signals with timing jitter [Hug57, Yos73] and random telegraph messages [Hug57]. The natural sciences have been represented by papers on modeling of wind [Jon86] and solar radiation [Agu88]. Two papers have discussed MCRPs in the context of speech modeling [Mul72, Sie76], which is one of the applications considered in this research. Additionally, binary MCRPs have been employed in lossless data compression

and have achieved excellent results [Cor87]. Of these applications, only the wind, radiation, and speech papers are directly related to this research. More information on these topics is discussed in Chapters IV and V.

Overview

Examination of Markov chain random processes begins in Chapter II with a review of Markov chains. As noted previously, the literature on Markov chains is extensive, and much has little bearing on this research of MCRPs. Consequently, the material in the second chapter is restricted to those topics which relate directly to this study.

Following this, Chapter II defines and examines MCRPs. Several statistics are derived including the power spectrum. The derivations in this chapter are somewhat different from any approach previously documented. Sittler's derivation involved flow graphs and analysis techniques of electrical circuits. However, this procedure is difficult for large systems. The results from other papers are too general, involving time-varying parameters and arbitrary output waveforms. Consequently, it was necessary to reformulate the results for the purposes of this research. In particular, the derivations and results of Chapter II are given in terms of matrices and vectors, with emphasis on eigen decompositions. In addition, Chapter II also discusses Hidden Markov Models in relation to Markov chains and Markov chain random processes.

The subject of Chapter III is synthesis. The key goal of this chapter is to locate MCRPs which achieve desired power spectra. After carefully discussing the key obstacles to a solution of this problem, several approaches are considered. In particular, a ring-structured MCRP is introduced and examined in detail. Although there are some limitations of this structure, it appears to be the best solution to the spectral synthesis problem to date. Additionally, Chapter II considers synthesis of MCRP with constraints of distribution function and joint constraints of distribution function and power spectrum.

Chapter IV considers MCRPs as models of random processes. The discussion leads to an exploration of the relationship between ARMA processes and MCRPs. It is shown that, under certain conditions, the power spectrum expressions for MCRPs and ARMA processes driven by white noise are similar. Using this relationship and the ring-structured MCRP introduced in Chapter III, a procedure is formulated whereby a colored noise process can be approximated by an MCRP.

In Chapters V through VII, several applications for MCRPs are considered. In Chapter V, MCRPs are studied as models for speech signals. This provides an excellent opportunity to test the results of Chapter IV. The motivation for using MCRPs for speech is partially based on the success of Hidden Markov Models (HMMs) for speech. This chapter begins by examining conventional speech models. An MCRP model for speech is then presented and discussed. Preliminary results are offered which confirm the validity of this approach.

Chapter VI discusses the use of MCRPs in direct-sequence spread spectrum (DS-SS) communications. Conventionally, m-sequences are used as the spreading sequence for DS-SS. In Chapter VI, the use of MCRPs as spreading sequences is explored. The main advantage of this approach is that, as binary random processes, MCRPs are much more flexible than m-sequences. This extra flexibility can prove useful in a jamming environment.

The topic of Chapter VII is Markov random fields. In essence, this is an extension of MCRPs to two-dimensional processes, and to sampled images in particular. A thorough examination of such processes is beyond the scope of this document. However, Chapter VII introduces MCRPs in the context of Markov random fields and discusses their feasibility. This approach seems especially well-suited to image processing applications in which there are a relatively few number of pixel bits. A simple example is given which demonstrates the potential of MCRPs in this area.

Finally, Chapter VIII provides a brief summary of this manuscript, noting the contributions and results of this research. Additionally, ideas for future research involving MCRPs are discussed.

CHAPTER II

MARKOV CHAIN RANDOM PROCESSES

This chapter is concerned with Markov chains and related processes. In the first section of this chapter, Markov chains are discussed. The main purpose of this section is to develop an understanding of Markov chains which shall be essential to the study of Markov chain random processes. The literature on Markov chains is extensive, and much has little bearing on MCRPs. Consequently, the material in this section shall be restricted to those topics which are pertinent to this study. In particular, this section shall concentrate on stationary Markov chains. The second section of this chapter examines Markov chain random processes. Several important statistics of stationary MCRPs are discussed, including mean, variance, autocorrelation, and power spectral density. The final section of this chapter briefly reviews hidden Markov models (HMMs) and discusses their relation to Markov chains and MCRPs.

Markov Chains

A Markov chain is essentially a probabilistic, discrete-time, state system. In general, a Markov chain can have an infinite number of states. However, this study is restricted to Markov chains having a finite number of states which for convenience shall be numbered from 1 to N . Markov chains are distinguished from other state systems because they obey a first-order Markov property [Jos80]. In words, this property requires that the state transitions depend only on the present state of the system; previous states of the system do not influence state transitions. The Markov property is perhaps better

explained by the conditional probability equation

$$P[x(n+1) = j \mid x(n) = i, x(n-1) = h, \dots] = P[x(n+1) = j \mid x(n) = i] \quad (\text{II.1})$$

where $x(n)$ is an integer-valued random process which assumes the value of the state which the system occupies at time n , and h , i , and j are integers which correspond to state numbers. The equation implies that $x(n)$ is statistically dependent upon $x(n-1)$ and independent of $x(n-2)$, $x(n-3)$, etc.

It would appear that the Markov property severely limits the usefulness of Markov chains since most real-world processes have a higher degree of dependence upon the past behavior of the system. However, it is possible to represent such processes in terms of Markov chains [Kaz82]. To see how this is done, consider a process which has N states and for which the state transitions depend on the present state as well as the previous $(k-1)$ states. Such a process is termed an N -state, k^{th} order system. By letting each ordered k -tuple of states in this system comprise a state in a new system, a 1st order process with N^k states is created which behaves like a Markov chain. The main obstacle in dealing with higher order systems directly is that, unlike Markov chains, convenient representation of parameters and statistics in terms of matrices is cumbersome.

Markov Chain Notation

Markov chains are easily illustrated by means of flow graphs. As an example, Figure 2.1 shows a flow graph for a particular 3-state Markov chain. In the figure, nodes indicate states of the system while arrows between nodes represent state transitions. A number, $t_{i,j}$, is associated with each arrow indicating the probability of transition from state i to state j . In general, the *transition probabilities* $\{t_{i,j}\}$ could be functions of time. In fact, the only constraints on these probabilities are as follows:

$$0 \leq t_{i,j} \leq 1 \quad \forall i, j \quad (\text{II.2a})$$

$$\text{and } \sum_{j=1}^N t_{i,j} = 1 \quad \forall i. \quad (\text{II.2b})$$

For the purposes of this paper, we shall assume that $\{t_{ij}\}$ are not functions of time. In this case, the Markov chain is said to be *homogeneous*.

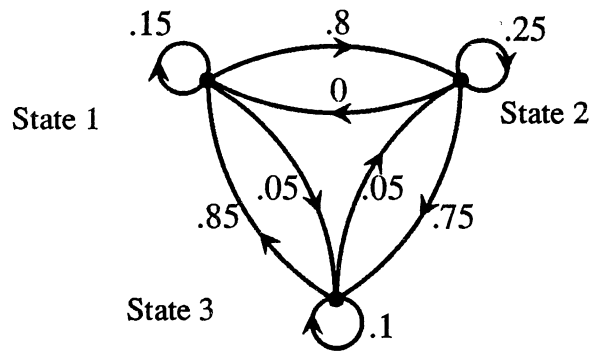


Figure 2.1. 3-State Markov Chain Flow Graph.

Related to transition probabilities are *state probabilities*: let $s_i(n)$ represent the probability that the chain is in state i at time n . The state probabilities are related to the transition probabilities by the equation

$$s_j(n+1) = \sum_{i=1}^N s_i(n)t_{i,j} \quad (\text{II.3})$$

for any j . Similarly, one could write

$$s_j(n+2) = \sum_{i=1}^N s_i(n)t_{i,j}(2),$$

where $t_{i,j}(2)$ is defined to be the two-step transition probability $P[x(n+2) = j | x(n) = i]$.

In general, let $t_{i,j}(m)$ be the m -step transition probability $P[x(n+m) = j | x(n) = i]$ for $m \geq 0$. Note in particular that $t_{i,i}(0) = 1$ and $t_{i,j}(0) = 0$ for $i \neq j$. Also, $t_{i,j}(1)$ is just $t_{i,j}$. To find the relationship between the set of m -step transition probabilities, $\{t_{i,j}(m)\}$, and the set of one-step transition probabilities, $\{t_{i,j}\}$, it is more convenient to utilize matrix notation.

Towards this end, T shall be defined as the $N \times N$ matrix with elements $[T]_{i,j} = t_{i,j}$. That is, the element in the i^{th} row and j^{th} column of T is $t_{i,j}$. Since T is composed of transition probabilities, it is referred to as the *transition probability matrix*. For the Markov chain in Figure 2.1,

$$T = \begin{pmatrix} .15 & .8 & .05 \\ 0 & .25 & .75 \\ .85 & .05 & .1 \end{pmatrix}.$$

Similarly, the state probabilities can be represented in vector notation as

$$s(n) = (s_1(n) \quad s_2(n) \quad \cdots \quad s_N(n))^T. \quad (\text{II.4})$$

Since $s_j(n+1) = \sum_{i=1}^N s_i(n)t_{i,j}$, the set $\{s_j(n+1)\}$ may be more compactly represented by

$$s^T(n+1) = s^T(n)T. \quad \text{Through a simple substitution, } s^T(n+2) = s^T(n+1)T = s^T(n)T^2.$$

Repeating this argument, the following becomes apparent:

$$s^T(n+m) = s^T(n)T^m \quad \text{for } m \geq 0. \quad (\text{II.5})$$

Also, since $s_j(n+m) = \sum_{i=1}^N s_i(n)t_{i,j}(m)$, it follows that the m -step transition probabilities

are

$$t_{i,j}(m) = [T^m]_{i,j}. \quad (\text{II.6})$$

Transition probability matrices belong to a class of matrices called *stochastic* or *probability* matrices. These matrices are characterized by having real, nonnegative elements and by the fact that each row of the matrix sums to unity. One can see that T is a stochastic matrix since $[T]_{i,j} = t_{i,j} \geq 0$ and $\sum_{j=1}^N t_{i,j} = 1$. This matrix plays an important part in the later parts of this chapter and in subsequent chapters. Properties of stochastic matrices are discussed in texts on Markov chains, such as [Kem76], and in many texts on matrices, such as [Lan85]. Some of these properties shall be specified later in this chapter. However, there are several properties of stochastic matrices which are worth noting at this time. First, T always has an eigenvalue equal to 1 with associated right eigenvector $(1 \ 1 \ \cdots \ 1)^T$:

$$\begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1N} \\ t_{21} & t_{22} & \cdots & t_{2N} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ t_{N1} & t_{N2} & \cdots & t_{NN} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ 1 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ 1 \end{pmatrix}$$

since $\sum_{j=1}^N t_{i,j} = 1$ for all i . This eigenvalue shall be denoted throughout this manuscript as λ_0 , and the remaining eigenvalues as $\lambda_1, \lambda_2, \dots, \lambda_{N-1}$. It is also possible to show that any eigenvalue of T is less than or equal to 1 in magnitude: letting λ be any eigenvalue of T with associated left eigenvector $\mathbf{u} = (u_1 \ u_2 \ \cdots \ u_N)^T$,

$$\lambda \mathbf{u}^T = \mathbf{u}^T T$$

$$\lambda u_j = \sum_i u_i t_{i,j}$$

$$\text{so } \sum_j |\lambda u_j| = \sum_j \left| \sum_i u_i t_{i,j} \right|$$

$$|\lambda| \sum_j |u_j| \leq \sum_j \sum_i |u_i| t_{i,j}$$

$$|\lambda| \sum_j |u_j| \leq \sum_i |u_i| \sum_j t_{i,j} = \sum_i |u_i|$$

$$\therefore |\lambda| \leq 1.$$

(II.7)

Based on the transition probabilities, it is possible to categorize Markov chains into several classes. This classification is based on the concept of communication between states: the states i and j are said to *communicate* if it is possible for the Markov chain to transition from state i to j and from state j to i . Note that communication does not have to be direct; for the Markov chain in Figure 2.1, even though $t_{2,1} = 0$, state 2 communicates with state 1 since it can reach state 1 by first passing through state 3.

Markov chains can be classified as either *reducible* or *irreducible* [Ios80]. In an irreducible Markov chain, every pair of states communicate; conversely, if a Markov chain has any pair of noncommunicating states, it is said to be reducible. Irreducible chains are further classified as either *regular* or *cyclic*. A Markov chain is regular if all of the elements of T^n are strictly positive for some $n > 0$; a Markov chain is cyclic if it is not regular. For the example in Figure 2.1, every pair of states communicates, so this Markov chain is irreducible. Further, since every element of

$$T^2 = \begin{pmatrix} .065 & .3225 & .6125 \\ .6375 & .1 & .2625 \\ .2125 & .6975 & .09 \end{pmatrix}$$

is strictly positive, the Markov chain is regular.

It should be noted that there is a relationship between Markov chain and matrix terminologies. For example, there is a definition of reducibility for both matrices and Markov chains. Fortunately, the two are related: a Markov chain is reducible if and only if its transition probability matrix is reducible. A similar statement is true for the terms irreducible, regular, cyclic, primitive, and imprimitive¹ (see, for example [Lan85, Sen81, Kem76, and Ios80]).

¹ The terms primitive and imprimitive will be introduced in the next section.

Stationary Markov Chains

A *random* or *stochastic process* is a (usually infinite) set of samples or outcomes each of which is a function of time. In a Markov chain, the outcomes take the form of infinite sequences which denumerate the state transitions as a function of time. For example, in the 3-state Markov chain in Figure 2.1, a possible outcome is $\{2, 3, 1, 1, 3, 2, 2, \dots\}$, indicating that the Markov chain began in state 2, transitioned to state 3, then to state 1, and so on.

For the purposes of this research, the main interest is *stationary* random processes. For a random process to be stationary, the statistics of the process must be time-invariant. For homogeneous Markov chains, stationarity is insured if the state probabilities are constant [Pap84]; that is, if $s(n)$ does not depend on n .

To consider the stationarity of a Markov chain, recall that the state probabilities were expressed in matrix-vector notation as $s^T(n+m) = s^T(n)T^m$ for $m \geq 0$. For the Markov chain to be stationary, $s(n)$ must be such that $s^T(n) = s^T(n)T^m$ for all n and for all $m \geq 0$. In particular, this means that $s^T(n) = s^T(n)T$. Since it has already been shown that T has the eigenvalue $\lambda_0 = 1$, the previous equation implies that $s(n)$ must be a left eigenvector of T associated with this eigenvalue. Hence, a Markov chain will be stationary if $s(n)$ is a left eigenvector of T associated with the eigenvalue $\lambda_0 = 1$. A state distribution $s(n)$ which makes the Markov chain stationary will be called a *stationary distribution* of the Markov chain and shall be denoted by s .

Now, in general, the eigenvalue $\lambda_0 = 1$ could have multiplicity greater than one. In this case, the stationary distribution would not be unique. However, suppose the Markov chain is irreducible. If so, then the stationary distribution is unique. This result follows from the Perron-Frobenius theorem [Lan85]:

Theorem (Perron-Frobenius): For a nonnegative square irreducible matrix,
 (i) The matrix has a real eigenvalue equal to its *spectral radius*,
 (ii) There is a real right eigenvector associated with this eigenvalue; and
 (iii) The algebraic multiplicity of this eigenvalue is 1.

The spectral radius is defined as the magnitude of the largest eigenvalue of the matrix. For T , the spectral radius is 1 since $|\lambda_k| \leq 1 = \lambda_0$. It follows from the Perron-Frobenius Theorem that the eigenvalue $\lambda_0 = 1$ has algebraic multiplicity of 1 for irreducible Markov chains. Consequently, the stationary distribution of irreducible Markov chains must be unique.

If the Markov chain is further limited to being regular, then an even stronger statement can be made: the state probabilities converge to unique values independent of the initial state distribution $s(0)$. To see why, first recall from the previous section that a regular Markov chain has the property that every element of T^n is positive for some $n > 0$. An equivalent definition is that T can have only one eigenvalue with unity magnitude [Lan85]. This property is called *primitivity*, and some texts use the term primitive instead of regular to describe these Markov chains (for example, [Sen81]).

The benefits of regular Markov chains arise from the fact that every eigenvalue of T , except λ_0 , has magnitude strictly less than one. To see why this is such a nice property, express T in its eigen decomposition as

$$T = V \begin{pmatrix} 1 & & & \\ & J_1 & & \\ & & \ddots & \\ & & & J_M \end{pmatrix} V^{-1} \quad (\text{II.8})$$

where V is the modal matrix of T and where J_i is the Jordan block for the i^{th} eigenvalue (λ_i) of T . Now consider $\lim_{n \rightarrow \infty} T^n$. Since all of the eigenvalues except for $\lambda_0 = 1$ are strictly less than unity in magnitude, it follows that $\lim_{n \rightarrow \infty} J_i^n$ converges to a matrix of all zeros. Thus,

$$\begin{aligned} \lim_{n \rightarrow \infty} T^n &= V \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & 0 \end{pmatrix} V^{-1} \\ &= v_0 u_0^T \end{aligned}$$

where $v_0 = (1 \ 1 \ \cdots \ 1)^T$ is the right eigenvector and u_0 is the left eigenvector of T associated with $\lambda_0 = 1$ and $u_0 v_0^T = 1$. Then, letting $u_0 = (u_{0,1} \ u_{0,2} \ \cdots \ u_{0,N})^T$, and denoting $\lim_{n \rightarrow \infty} T^n$ as T^∞ , one can see that

$$\begin{aligned} T^\infty &= v_0 u_0^T \\ &= \begin{pmatrix} u_{0,1} & u_{0,2} & \cdots & u_{0,N} \\ u_{0,1} & u_{0,2} & \cdots & u_{0,N} \\ \cdot & \cdot & \cdots & \cdot \\ u_{0,1} & u_{0,2} & \cdots & u_{0,N} \end{pmatrix} \end{aligned} \quad (\text{II.9})$$

Next, consider $\lim_{n \rightarrow \infty} s(n)$ for an arbitrary initial state distribution $s(0)$:

$$\begin{aligned} \lim_{n \rightarrow \infty} s^T(n) &= \lim_{n \rightarrow \infty} s^T(0) T^n \\ &= s^T(0) \lim_{n \rightarrow \infty} T^n \\ &= s^T(0) T^\infty \\ &= s^T(0) v_0 u_0^T. \end{aligned}$$

Now, $s^T(0) v_0 = (s_1(n) \ s_2(n) \ \cdots \ s_N(n)) (1 \ 1 \ \cdots \ 1)^T = \sum_i s_i(n) = 1$ regardless of $s(0)$. Thus,

$$\begin{aligned}\lim_{n \rightarrow \infty} s^T(n) &= s^T(0) v_0 u_0^T \\ &= u_0^T\end{aligned}$$

regardless of $s(0)$. Also, because $\lambda_0 = 1$ has unit multiplicity for regular Markov chains, the stationary probability s is unique, and so, $s = u_0$. Hence,

$$\lim_{n \rightarrow \infty} s^T(n) = s. \quad (\text{II.10})$$

The conclusion is that, for regular Markov chains, $s(n)$ converges to the stationary distribution independent of the initial distribution. Because of this, regular Markov chains are always asymptotically stationary [Pap84].

Regular Markov chains are also *ergodic* random processes [Ios80]. That is, with probability one, any statistic of the process can be determined from a single outcome [Pap84]. It should be noted that, in many texts, irreducible Markov chains are called ergodic Markov chains (for example, [Kem76, Ios80]). However, this is a misnomer as the ergodic property holds only for regular Markov chains [Ios80]. This special case when the Markov chain is regular is the most important case and is assumed here unless otherwise mentioned.

The example in Figure 2.1 is a regular Markov chain and therefore ergodic and asymptotically stationary. For this example,

$$T^\infty = \begin{pmatrix} .31915 & .3617 & .31915 \\ .31915 & .3617 & .31915 \\ .31915 & .3617 & .31915 \end{pmatrix}$$

and therefore the stationary state distribution, as well as the limiting state distribution, is $s = (.31915 \quad .3617 \quad .31915)^T$. This indicates that the Markov chain spends slightly more time in state 2 on the average than in either state 1 or 3.

Markov Chain Random Processes

Markov chain random processes (MCRPs) were first introduced by Sittler [Sit54, Sit56]. Simply stated, an MCRP is created by associating a real number to each state of a Markov chain. To explain further, let a_i denote the real number associated with state i . Then, if the Markov chain is in state j at time n , the output of the MCRP is equal to a_j at time n . For convenience, let $\mathbf{a} = (a_1 \ a_2 \ \cdots \ a_N)^T$, which shall be referred to as the *output map*.

To fix ideas, reconsider the Markov chain from Figure 2.1. This Markov chain has three states, so an MCRP based on this Markov chain requires three real numbers to make up the output map. Although any three real numbers would suffice, suppose that $a_1 = -2.0$, $a_2 = 1.5$, and $a_3 = 0.3$. This MCRP, which is illustrated graphically below in Figure 2.2, shall be used to illustrate the concepts of this section. Recall from a previous section that a typical outcome of the example Markov chain was $\{2, 3, 1, 1, 3, 2, 2, \dots\}$, indicating that the Markov chain began in state 2, transitioned to state 3, then to state 1, and so on. The corresponding output for the Markov chain random process would then be $\{a_2, a_3, a_1, a_1, a_3, a_2, a_2, \dots\}$, or $\{1.5, 0.3, -2.0, -2.0, 0.3, 1.5, 1.5, \dots\}$.

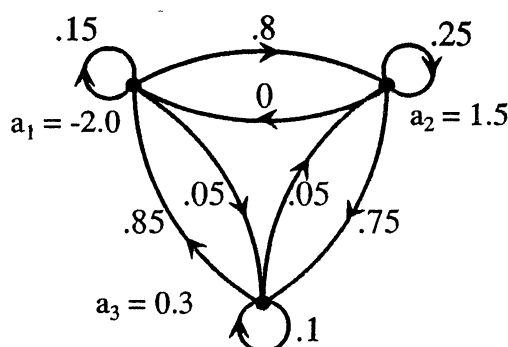


Figure 2.2. An MCRP with 3 States.

Now, in general, one could use any Markov chain and create an MCRP in the fashion described above. However, there are types of Markov chains which are better suited for this than others. This is especially true since the statistics of MCRPs are dependent upon the underlying Markov chain. Thus, while it is possible to discuss the statistics of MCRPs based on an arbitrary Markov chain, it is beneficial to restrict the attention to classes of Markov chains which have desirable properties. Consequently, this section is focused on the class of regular Markov chains. As enumerated in a previous section of this chapter, these Markov chains (and therefore the MCRPs derived from them) have many desirable properties: they can be made stationary; they are always asymptotically stationary independent of the initial state distribution; and they are ergodic.

Most of the material presented below is not new; the statistics of MCRPs have been discussed by several authors, including Sittler, Huggins, Zadeh, Barnard and Yoshida [Sit54, Sit56, Hug57, Zad57, Bar64, and Yos73]. However, most of these treatments are either too general or too specialized. Consequently, the author has found it necessary to rederive several pertinent statistics. In particular, the focus has been on eigen decomposition of the transition probability matrix. By deriving statistics in this manner, a few statistical parameters can be determined which give greater insight into the nature of these processes.

Implementation Aspects of MCRPs ✓

Perhaps the first question one might ask about MCRPs concerns how they are implemented. This is a good query and provides a natural introduction to Markov chain random processes. Since MCRPs are so closely related to Markov chains, it suffices to show how Markov chains can be implemented. To begin, suppose that the process is in state i at time n and consider how the next state should be chosen. To facilitate determi-

nation of the state transitions, let $u(n)$ be a random process with a uniform $[0,1]$ distribution. Let us determine the state transitions as follows: the state at time $n+1$ will be state j if and only if

$$\sum_{k=1}^{j-1} t_{i,k} < u(n) < \sum_{k=1}^j t_{i,k} \quad (\text{II.11})$$

Note that it is possible to interpret the above equation as a transfer function which has $u(n)$ as its input. As an example, suppose that the MCRP has only 3 states, and that the current state is $x(n) = i$. Then, the next state, $x(n+1)$, is the output of the state decision function illustrated in Figure 2.3. That is, $x(n+1) = 1$ if $u(n) < t_{i,1}$; $x(n+1) = 2$ if $t_{i,1} < u(n) < t_{i,1} + t_{i,2}$; and $x(n+1) = 3$ otherwise.

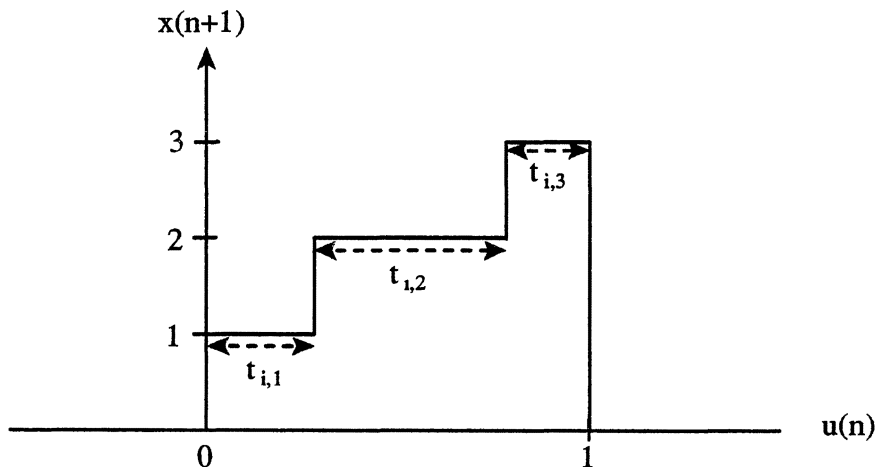


Figure 2.3. State Decision Function.

To see why this works, consider that, according to Equation (II.11),

$$P[x(n+1) = j | x(n) = i] = P\left[\sum_{k=1}^{j-1} t_{i,k} < u(n) < \sum_{k=1}^j t_{i,k}\right].$$

But, $u(n)$ has a uniform $[0,1]$ distribution, so

$$\begin{aligned} P[x(n+1) = j | x(n) = i] &= \sum_{k=1}^j t_{i,k} - \sum_{k=1}^{j-1} t_{i,k} \\ &= t_{i,j}. \end{aligned}$$

But this is exactly as desired since $t_{i,j}$ is defined as $P[x(n+1) = j | x(n) = i]$. Hence, Equation (II.11) accurately reflects how the state transitions should be determined based on the transition probabilities and $u(n)$.

Because of the way MCRPs are defined, an MCRP implementation is practically identical to a Markov chain implementation. However, for Markov chain random processes, the state decision would need to be followed by a look-up to determine the real number associated with the new state. Now, this implementation procedure provides an interesting interpretation of MCRPs. As noted above, if the state transitions are determined by Equation (II.11), then $u(n)$ can be thought of as the input to the process, which, according to Figure 2.3 is nonlinear. Thus, Markov chain random processes can be interpreted as nonlinear stochastic processes.

Summarizing this section, MCRPs are clearly easy to implement and require little computing power since the multiplication and addition of real numbers is not required. In fact, the only requirements are memory for the output map and transition probabilities; the generation of $u(n)$, and logic to perform the state decision function.

First Order Statistics of MCRPs

Suppose a regular Markov chain has transition probability matrix T with stationary probability s , and output map a . Then the triplet (T,s,a) defines a stationary Markov chain random process. The first step in the analysis of a stationary process is usually a determination of the process' probability density function or probability distribution,

depending upon whether the process is of a continuous or discrete nature, respectively. Clearly, an MCRP can only assume values from the output map. Hence, MCRPs are discrete processes, and the probability distribution function must be found.

To begin, suppose $s = (s_1 \ s_2 \ \cdots \ s_N)^T$ is the stationary probability vector and $a = (a_1 \ a_2 \ \cdots \ a_N)^T$ is the output map for a given MCRP. As in a previous section, let $x(n)$ be an integer-valued random process which assumes the value of the state which the Markov chain occupies at time n . Similarly, let $y(n)$ be a real-valued random process which assumes the value of the output map associated with the state which the Markov chain occupies at time n . For example, if the process is in state 2 at time n , then $x(n) = 2$ and $y(n) = a_2$.

Now consider the probability distribution for an MCRP. Note that $P[y(n) = a_i] = P[x(n) = i]$ since the statement $y(n) = a_i$ is equivalent to $x(n) = i$. But, $P[x(n) = i] = s_i(n)$, the probability that the Markov chain is in state i at time n . Further, since the process is stationary, $s_i(n) = s_i$. Therefore, the probability distribution for the MCRP is

$$P[y(n) = a_i] = s_i \quad \text{for } 1 \leq i \leq N. \quad (\text{II.12})$$

Having easily determined the probability distribution, the mean and variance of the process may be found. The mean is given by

$$\begin{aligned} \mu_y &= E\{y(n)\} \\ &= \sum_{i=1}^N P[y(n) = a_i] a_i \\ &= \sum_{i=1}^N s_i a_i \end{aligned} \quad (\text{II.13})$$

The mean may be more compactly represented in vector notation as

$$\mu_y = s^T a \quad (\text{II.14})$$

The variance is found in a similar fashion:

$$\begin{aligned}
\sigma_y^2 &= E\{(y(n) - \mu_y)^2\} \\
&= \sum_{i=1}^N P[y(n) = a_i] (a_i - \mu_y)^2 \\
&= \sum_{i=1}^N s_i (a_i^2 - 2a_i \mu_y + \mu_y^2) \\
&= \sum_{i=1}^N s_i a_i^2 - 2\mu_y \sum_{i=1}^N s_i a_i + \mu_y^2 \sum_{i=1}^N s_i \\
&= \sum_{i=1}^N s_i a_i^2 - \mu_y^2
\end{aligned}$$

To facilitate the expression of the variance in matrix-vector form, let S be a diagonal matrix of stationary probabilities. That is, let $[S]_{i,i} = s_i$. For the example from Figure 2.2, $s = (.31915 \quad .3617 \quad .31915)^T$, so

$$S = \begin{pmatrix} 0.31915 & 0 & 0 \\ 0 & 0.3617 & 0 \\ 0 & 0 & 0.31915 \end{pmatrix}$$

Then, the variance may be expressed as

$$\sigma_y^2 = \mathbf{a}^T S \mathbf{a} - (\mathbf{s}^T \mathbf{a})^2 \quad (\text{II.15})$$

For the example of Figure 2.2, the output map is given by $\mathbf{a} = (-2.0 \quad 1.5 \quad 0.3)^T$.

From a previous section of this chapter, the stationary distribution was found to be $\mathbf{s} = (.31915 \quad .3617 \quad .31915)^T$. Thus,

$$P[y(n) = -1] = 0.31915,$$

$$P[y(n) = 0] = 0.3617,$$

$$\text{and } P[y(n) = 1] = 0.31915.$$

The mean of this process is given by

$$\begin{aligned}
\mu_y &= \sum_{i=1}^N s_i a_i \\
&= (0.31915)(-2.0) + (0.3617)(1.5) + (0.31915)(0.3) \\
&= 0
\end{aligned}$$

and the variance is

$$\begin{aligned}
\sigma_y^2 &= \sum_{i=1}^3 s_i a_i^2 - \mu_y^2 \\
&= (0.31915)(-2.0)^2 + (0.3617)(1.5)^2 + (0.31915)(0.3)^2 - 0 \\
&= 2.1191.
\end{aligned}$$

It should be noted that, although the MCRP in this example has zero mean, it is possible for an MCRP to have a non-zero mean.

Autocorrelation of MCRPs

Proceeding in a similar fashion, the autocorrelation can be derived as follows: for $n \geq 0$,

$$\begin{aligned}
\phi(n) &= E\{y(k+n)y(k)\} \\
&= \sum_y P[y(k+n) = a_j, y(k) = a_i] a_j a_i \\
&= \sum_y P[x(k+n) = j, x(k) = i] a_j a_i \\
&= \sum_{ij} P[x(k+n) = j | x(k) = i] P[x(k) = i] a_j a_i.
\end{aligned}$$

Now, recall from a previous section that, by definition, $t_{ij}(n) = P[x(k+n) = j | x(k) = i]$.

Also, $P[x(k) = i] = s_i$ since the process is stationary. Then

$$\phi(n) = \sum_{ij} t_{ij}(n) s_i a_j a_i$$

But $t_{i,j}(n)$ is $[T^n]_{i,j}$. Thus,

$$\phi(n) = \sum_y [T^n]_{ij} s_i a_j a_i. \quad (\text{II.16})$$

From this Equation, the autocorrelation can be expressed in matrix notation. To do this, note

$$\mathbf{T}^n \mathbf{a} = \begin{pmatrix} \sum_j [T^n]_{1,j} a_j \\ \sum_j [T^n]_{2,j} a_j \\ \cdot \\ \cdot \\ \cdot \\ \sum_j [T^n]_{N,j} a_j \end{pmatrix}$$

Recall from a previous section that \mathbf{S} was defined as an $N \times N$ diagonal matrix and the diagonal elements are the stationary state probabilities; i.e., $[\mathbf{S}]_{i,i} = s_i$. Then, note

$$\mathbf{a}\mathbf{S} = (a_1 s_1 \quad a_2 s_2 \quad \cdots \quad a_N s_N)$$

Combining these terms,

$$\begin{aligned} \mathbf{a}^T \mathbf{S} \mathbf{T}^n \mathbf{a} &= (a_1 s_1 \quad a_2 s_2 \quad \cdots \quad a_N s_N) \begin{pmatrix} \sum_j [T^n]_{1,j} a_j \\ \sum_j [T^n]_{2,j} a_j \\ \cdot \\ \cdot \\ \cdot \\ \sum_j [T^n]_{N,j} a_j \end{pmatrix} \\ &= \sum_j [T^n]_{j,j} s_j a_j a_j \\ &= \phi(n). \end{aligned}$$

Also, since the autocorrelation is an even function, $\phi(n) = \phi(-n)$. Thus,

$$\phi(n) = \mathbf{a}^T \mathbf{S} \mathbf{T}^{|n|} \mathbf{a} \quad \forall n. \quad (\text{II.17})$$

At this point, it is necessary to digress and distinguish between cases when the process has zero mean and when it doesn't. From Equation (II.14), the mean of the process is given by $\underline{\mu}_y = \underline{s}^T \underline{a}$. Now, define $\tilde{a}_i = a_i - \mu_y$, and $\tilde{\underline{a}} = \underline{a} - \underline{\mu}$, where $\underline{\mu}$ is a vector of length N where every element is μ_y . Then, making the substitution $\underline{a} = \tilde{\underline{a}} + \underline{\mu}$,

$$\begin{aligned}\phi(n) &= (\tilde{\underline{a}} + \underline{\mu})^T \underline{S} T^{n-1} (\tilde{\underline{a}} + \underline{\mu}) \\ &= \tilde{\underline{a}}^T \underline{S} T^{n-1} \tilde{\underline{a}} + \underline{\mu}^T \underline{S} T^{n-1} \tilde{\underline{a}} + \tilde{\underline{a}}^T \underline{S} T^{n-1} \underline{\mu} + \underline{\mu}^T \underline{S} T^{n-1} \underline{\mu}\end{aligned}\quad (\text{II.18})$$

Now, note that $T^{n-1} \underline{\mu} = \underline{\mu}$ since T is a stochastic matrix. Substituting this in the above equation,

$$\begin{aligned}\tilde{\underline{a}}^T \underline{S} T^{n-1} \underline{\mu} &= \tilde{\underline{a}}^T \underline{S} \underline{\mu} \\ &= \sum_{i=1}^N \tilde{a}_i s_i \mu_y \\ &= \mu_y \sum_{i=1}^N (a_i - \mu_y) s_i \\ &= \mu_y \left(\sum_{i=1}^N a_i s_i - \mu_y \sum_{i=1}^N s_i \right) \\ &= \mu_y (\mu_y - \mu_y) \\ &= 0.\end{aligned}$$

Similarly, since $\underline{\mu}^T \underline{S} = \mu_y \underline{s}^T$ and $\underline{s}^T T^{n-1} = \underline{s}^T$,

$$\begin{aligned}\underline{\mu}^T \underline{S} T^{n-1} \tilde{\underline{a}} &= \mu_y \underline{s}^T T^{n-1} \tilde{\underline{a}} \\ &= \mu_y \underline{s}^T \tilde{\underline{a}} \\ &= \sum_{i=1}^N \mu_y s_i \tilde{a}_i \\ &= 0\end{aligned}$$

as above. Finally, since $\underline{\mu}^T \underline{S} = \mu_y \underline{s}^T$ and $T^{n-1} \underline{\mu} = \underline{\mu}$,

$$\begin{aligned}
\boldsymbol{\mu}^T \mathbf{S} \mathbf{T}^{n-1} \boldsymbol{\mu} &= \mu_y \mathbf{s}^T \boldsymbol{\mu} \\
&= \mu_y \sum_{i=1}^3 s_i \mu_y \\
&= \mu_y^2.
\end{aligned}$$

Thus, Equation (II.18) simplifies to

$$\phi(n) = \tilde{\mathbf{a}}^T \mathbf{S} \mathbf{T}^{n-1} \tilde{\mathbf{a}} + \mu_y^2. \quad (\text{II.19})$$

So, processes with a nonzero mean have a constant term in the autocorrelation expression, whereas those with a zero mean do not.

Returning to the example of Figure 2.2, recall that

$$\mathbf{T} = \begin{pmatrix} .15 & .8 & .05 \\ 0 & .25 & .75 \\ .85 & .05 & .1 \end{pmatrix},$$

$$\mathbf{a} = (-2.0 \quad 1.5 \quad 0.3)^T,$$

$$\text{and } \mathbf{s} = (.31915 \quad .3617 \quad .31915)^T.$$

So, $\mu_y = \sum_i a_i s_i = 0$, and thus $\tilde{\mathbf{a}} = \mathbf{a}$. The autocorrelation for this 3-state system is then

given by

$$\phi(n) = (-2.0 \quad 1.5 \quad 0.3) \begin{pmatrix} .3191 & 0 & 0 \\ 0 & .3617 & 0 \\ 0 & 0 & .3191 \end{pmatrix} \begin{pmatrix} .15 & .8 & .05 \\ 0 & .25 & .75 \\ .85 & .05 & .1 \end{pmatrix}^{n-1} \begin{pmatrix} -2.0 \\ 1.5 \\ 0.3 \end{pmatrix}.$$

A plot of the autocorrelation is shown in Figure 2.4.

Although the matrix representation of $\phi(n)$ is correct, it gives little insight to the shape of the autocorrelation sequence. A more perceptive representation of $\phi(n)$ can be gained through the spectral representation of \mathbf{T} . Let us consider the special case when \mathbf{T} is diagonalizable. The case of non-diagonalizable \mathbf{T} will be discussed in a later section of this chapter. The spectral representation of a diagonalizable $N \times N$ matrix \mathbf{T} is given by [Lan85]

$$T = \sum_{k=0}^{N-1} (\mathbf{v}_k \mathbf{u}_k^T) \lambda_k \quad (\text{II.20})$$

where $\{\lambda_k\}$ are the eigenvalues of T and $\{\mathbf{v}_k\}$ ($\{\mathbf{u}_k\}$) are the corresponding right (left) eigenvectors properly normalized so that $\mathbf{u}_k^T \mathbf{v}_k = 1$ for all k . More generally, T^n is given by $\sum_{k=0}^{N-1} (\mathbf{v}_k \mathbf{u}_k^T) \lambda_k^n$ for $n > 0$. The autocorrelation can then be expressed as follows:

$$\begin{aligned} \phi(n) &= \tilde{\mathbf{a}}^T S T^n \tilde{\mathbf{a}} + \mu_y^2 \\ &= \tilde{\mathbf{a}}^T S \left\{ \sum_{k=0}^{N-1} (\mathbf{v}_k \mathbf{u}_k^T) \lambda_k^{|n|} \right\} \tilde{\mathbf{a}} + \mu_y^2 \\ &= \sum_{k=0}^{N-1} (\tilde{\mathbf{a}}^T S \mathbf{v}_k \mathbf{u}_k^T \tilde{\mathbf{a}}) \lambda_k^{|n|} + \mu_y^2 \\ &= \sum_{k=0}^{N-1} \beta_k \lambda_k^{|n|} + \mu_y^2 \end{aligned} \quad (\text{II.21})$$

$$\text{where } \beta_k \equiv \tilde{\mathbf{a}}^T S \mathbf{v}_k \mathbf{u}_k^T \tilde{\mathbf{a}}. \quad (\text{II.22})$$

From Equation (II.21), the autocorrelation of an MCRP is influenced chiefly by the eigenvalues of its transition probability matrix. The eigenvectors of T and the output map determine the coefficients $\{\beta_k\}$.

For the example, the eigenvalues of T are 1 and $0.7053e^{\pm 1.933j}$. Consequently, after some calculations, the autocorrelation can be expressed as

$$\phi(n) = (0) (1)^{|n|} + (1.06 - 0.09j) (0.706e^{1.93j})^{|n|} + (1.06 + 0.09j) (0.706e^{-1.93j})^{|n|}$$

Note in particular that, for this example, β_0 , the coefficient associated with $\lambda_0 = 1$, is zero.

In fact, β_0 will always be zero. To see why, we shall compare the definition of $\phi(n)$ with the expression for $\phi(n)$ given in Equation (II.21). First recall the definition of $\phi(n)$: $\phi(n) = E\{y(k+n)y(k)\}$. Now, consider the limit of this expression as n goes to infinity. For regular Markov chains, it has been established that the limiting state probabilities are independent of the initial state probabilities. Because of this, the state of the Markov chain at time $k+n$ becomes independent of the state at time k as n increases without

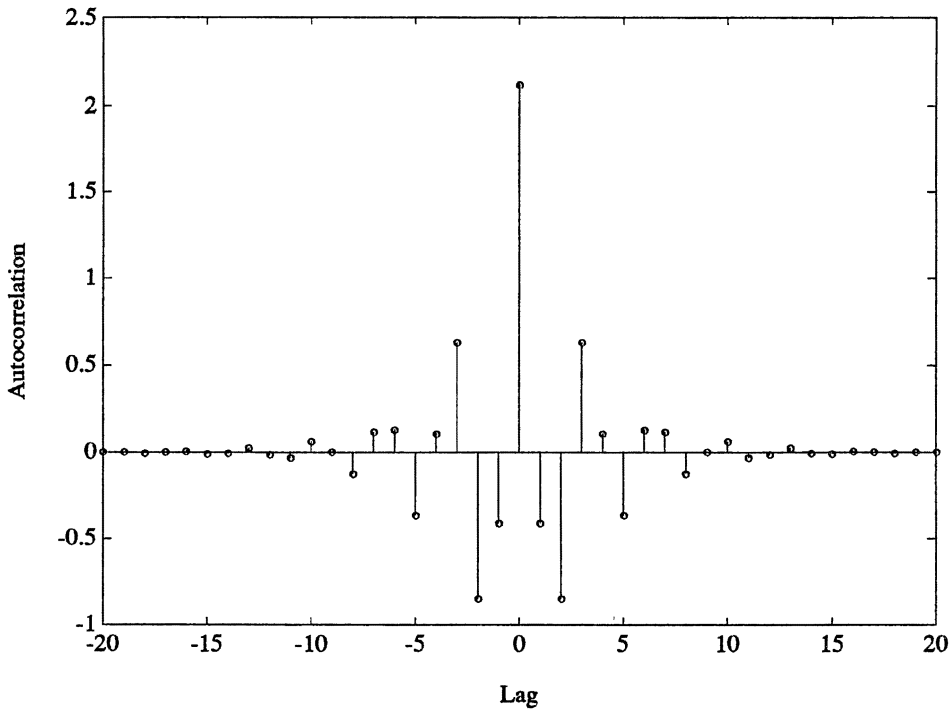


Figure 2.4. Autocorrelation of Example MCRP.

bounds. Hence, $y(k+n)$ and $y(k)$ will become independent as n goes to infinity. Therefore $\lim_{n \rightarrow \infty} \phi(n) = E\{y(k)\}E\{y(k+n)\}$. But $E\{y(k)\} = E\{y(k+n)\} = \mu_y$, so $\lim_{n \rightarrow \infty} \phi(n) = \mu_y^2$.

Now consider the eigenvalue representation of $\phi(n)$ in Equation (II.21). As n goes to infinity, λ_k^n goes to zero for all $|\lambda_k| < 1$. Thus, since $\lambda_0 = 1$ is the only eigenvalue with magnitude greater than or equal to 1, $\lim_{n \rightarrow \infty} \phi(n) = \lim_{n \rightarrow \infty} \beta_0 1^n + \mu_y^2 = \beta_0 + \mu_y^2$. It follows that β_0 must equal zero since, by the argument of the previous paragraph, $\phi(n)$ must converge to μ_y^2 . Consequently, since $\beta_0 = 0$, the associated term may be removed from the autocorrelation expression, which yields

$$\phi(n) = \sum_{k=1}^{N-1} \beta_k \lambda_k^{|n|} + \mu_y^2 \quad (\text{II.23})$$

Power Spectrum of MCRPs: General Expression

The power spectrum is a frequency representation of the autocorrelation. The power spectrum is found by taking the Fourier transform of the autocorrelation sequence, which is usually stated as the Z-transform of the sequence evaluated on the unit circle.

Namely,

$$\Phi(\omega) = \sum_{n=-\infty}^{\infty} \phi(n) z^{-n} \Big|_{z=e^{j\omega}}$$

As the power spectrum is a linear operator on the autocorrelation, it is possible to consider the components of $\phi(n)$ independently. That is, as $\phi(n) = \tilde{\mathbf{a}}^T \mathbf{S} \mathbf{T}^n \tilde{\mathbf{a}} + \mu_y^2$, let $\phi_1(n) = \tilde{\mathbf{a}}^T \mathbf{S} \mathbf{T}^n \tilde{\mathbf{a}}$ and $\phi_2(n) = \mu_y^2$ so that $\phi(n) = \phi_1(n) + \phi_2(n)$. The power spectrum of $\phi(n)$ is then given by the sum of the power spectrum of $\phi_1(n)$ and $\phi_2(n)$.

Consider first the power spectrum of $\phi_1(n)$. Before the power spectrum of the MCRP can be found, the z-transform of the autocorrelation needs to be studied. As is common, the z-transform is split into sections as follows:

$$\begin{aligned} \Phi_1(z) &= \sum_{n=-\infty}^{\infty} \phi_1(n) z^{-n} \\ &= \sum_{n=-\infty}^0 \phi_1(n) z^{-n} + \sum_{n=0}^{\infty} \phi_1(n) z^{-n} - \phi_1(0) \\ &= \Phi_1^-(z) + \Phi_1^+(z) - \phi_1(0). \end{aligned} \quad (\text{II.24})$$

However, since $\phi_1(-n) = \phi_1(n)$, $\Phi_1^-(z) = \sum_{n=0}^{\infty} \phi_1(n) z^n = \Phi_1^+(z^{-1})$. Therefore if $\Phi_1^+(z)$ or

$\Phi_1^-(z)$ is known, $\Phi_1(z)$ may be found.

Consider $\Phi_1^+(z)$:

$$\begin{aligned}\Phi_1^+(z) &= \sum_{n=0}^{\infty} \phi_1(n) z^{-n} \\ &= \sum_{n=0}^{\infty} \tilde{\mathbf{a}}^T S T^n \tilde{\mathbf{a}} z^{-n}.\end{aligned}$$

It is tempting to express this as $\tilde{\mathbf{a}}^T S (\sum_{n=0}^{\infty} T^n z^{-n}) \tilde{\mathbf{a}}$. However, $\sum_{n=0}^{\infty} T^n z^{-n}$ does not converge for $z=1$, and it is the unit circle on which we need $\Phi_1(z)$ to converge. Still, $\Phi_1(z)$ does converge on the unit circle, but to show this requires a little manipulation. First, consider the Jordan decomposition of T , which can be expressed as $T = VJV^{-1}$, where J is the block-diagonal matrix

$$J = \begin{pmatrix} 1 & & & \\ & J_1 & & \\ & & \ddots & \\ & & & J_M \end{pmatrix}$$

and where J_i is the Jordan block for the i^{th} eigenvalue of T , λ_i . Here, the eigenvalues of T are $\{1, \lambda_1, \dots, \lambda_M\}$. Since J is block-diagonal, T^n may be expressed as

$$T^n = V \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} V^{-1} + V \begin{pmatrix} 0 & & & \\ & J_1^n & & \\ & & \ddots & \\ & & & J_M^n \end{pmatrix} V^{-1} \quad (\text{II.25})$$

for $n > 0$. The first term of the right hand side is just $\nu_0 \mathbf{u}_0^T$, where ν_0 (\mathbf{u}_0) is the right (left) eigenvector of T associated with $\lambda_0 = 1$. Recall from a previous section of this chapter that $T^\infty = \lim_{n \rightarrow \infty} T^n = \nu_0 \mathbf{u}_0^T$. Thus, the first term in Equation (II.25) is T^∞ . Equation (II.25) can then be rewritten as

$$T^n = T^\infty + \tilde{T}^n$$

for $n > 0$, where \tilde{T} is defined as

$$\tilde{T} \equiv V \begin{pmatrix} 0 & & & & \\ & J_1 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & J_M \end{pmatrix} V^{-1}$$

$$= T - T^\infty$$

Note in particular that the eigenvalues of \tilde{T} are $\{0, \lambda_1, \dots, \lambda_M\}$. Since T was assumed to be regular, it follows that $|\lambda_i| < 1$ for $1 \leq i \leq M$. Thus, all of \tilde{T} 's eigenvalues are strictly less than unity in magnitude.

Now reconsider the autocorrelation: substituting $T^n = \tilde{T}^n + T^\infty$ yields

$$\begin{aligned} \phi_1(n) &= \tilde{\mathbf{a}}^T S T^n \tilde{\mathbf{a}} \\ &= \tilde{\mathbf{a}} S T^\infty \tilde{\mathbf{a}}^T + \tilde{\mathbf{a}} S \tilde{T}^n \tilde{\mathbf{a}}^T \end{aligned} \quad (\text{II.26})$$

for $n > 0$. But, $T^\infty = \nu_0 \mathbf{u}_0^T$ so $T^\infty \tilde{\mathbf{a}} = \nu_0 \mathbf{u}_0^T \tilde{\mathbf{a}}$. Further, recall from a previous section of this chapter that $\mathbf{s} = \mathbf{u}_0$ for regular Markov chains, so $T^\infty \tilde{\mathbf{a}} = \nu_0 \mathbf{s}^T \tilde{\mathbf{a}}$. Now, $\mathbf{s}^T \tilde{\mathbf{a}} = 0$:

$$\begin{aligned} \mathbf{s}^T \tilde{\mathbf{a}} &= \sum_i s_i (a_i - \mu_y) \\ &= \sum_i s_i a_i - \sum_i s_i \mu_y \\ &= \mu_y - \mu_y \\ &= 0. \end{aligned}$$

Hence, $\tilde{\mathbf{a}}^T S T^\infty \tilde{\mathbf{a}} = 0$. Thus, $\phi_1(n) = \tilde{\mathbf{a}}^T S \tilde{T}^n \tilde{\mathbf{a}}$ for $n > 0$. But, by convention, $\tilde{T}^0 = T^0 = I$, so $\phi_1(0) = \tilde{\mathbf{a}}^T S T^0 \tilde{\mathbf{a}} = \tilde{\mathbf{a}}^T S \tilde{T}^0 \tilde{\mathbf{a}}$. Therefore,

$$\phi_1(n) = \tilde{\mathbf{a}}^T S \tilde{T}^n \tilde{\mathbf{a}} \quad \forall n. \quad (\text{II.27})$$

The crucial difference here is that we have replaced T by \tilde{T} , and all of \tilde{T} 's eigenvalues are less than unity in magnitude.

Now, $\Phi_1^+(z)$ may be expressed as

$$\Phi_1^+(z) = \tilde{\mathbf{a}}^T S \left(\sum_{n=0}^{\infty} (\tilde{T}z^{-1})^n \right) \tilde{\mathbf{a}}. \quad (\text{II.28})$$

Consider the summation $\sum_{n=0}^{\infty} z^{-n} \tilde{T}^n$ as a power series of \tilde{T} . To study the convergence of this power series, consider the following theorem [Bro85]:

Theorem: For A , an $N \times N$ matrix with eigenvalues $\{\lambda_k\}$, if the power series $\sum_{n=0}^{\infty} b_n x^n$ converges for each $x = \lambda_k$, then $\sum_{n=0}^{\infty} b_n A^n$ converges.

In our case, the power series coefficients are $b_n = z^{-n}$. Hence, the power series is $\sum_{n=0}^{\infty} (z^{-n}) x^n$ which will converge for $|x/z| < 1$ or equivalently $|x| < |z|$ [Buc78]. Now, the eigenvalues of \tilde{T} are $0, \lambda_1, \dots, \lambda_{N-1}$. Hence, the power series will converge for all of the eigenvalues of \tilde{T} if $|z| > \max_{k \neq 0} |\lambda_k|$. All of the eigenvalues of \tilde{T} are strictly less than 1 in magnitude, and therefore the convergence region includes the unit circle. Hence, $\sum_{n=0}^{\infty} z^{-n} \tilde{T}^n$ converges for $|z| > \max_{k \neq 0} |\lambda_k|$, which includes the unit circle. Also, since convergence has been established, $\sum_{n=0}^{\infty} (\tilde{T}z^{-1})^n$ may be expressed as $(I - \tilde{T}z^{-1})^{-1}$ since

$$\begin{aligned} (I - \tilde{T}z^{-1}) \sum_{n=0}^{\infty} (\tilde{T}z^{-1})^n &= \sum_{n=0}^{\infty} (\tilde{T}z^{-1})^n - \sum_{n=1}^{\infty} (\tilde{T}z^{-1})^n \\ &= (\tilde{T}z^{-1})^0 \\ &= I. \end{aligned}$$

Then $\Phi_1^+(z) = \tilde{\mathbf{a}}^T S (I - \tilde{T}z^{-1})^{-1} \tilde{\mathbf{a}}$. By symmetry, $\Phi_1^-(z) = \tilde{\mathbf{a}}^T S (I - \tilde{T}z)^{-1} \tilde{\mathbf{a}}$ and converges

for $|z| < 1/\max_{k \neq 0} |\lambda_k|$. Note also that $\phi_1(0) = \tilde{\mathbf{a}}^T S T^0 \tilde{\mathbf{a}} = \tilde{\mathbf{a}}^T S \mathbf{a}$. Consequently,

$$\Phi_1(z) = \tilde{\mathbf{a}}^T S (I - \tilde{T}z)^{-1} \tilde{\mathbf{a}} + \tilde{\mathbf{a}}^T S (I - \tilde{T}z^{-1})^{-1} \tilde{\mathbf{a}} - \tilde{\mathbf{a}}^T S \tilde{\mathbf{a}} \quad (\text{II.29})$$

which is defined on the annulus $\max_{k \neq 0} |\lambda_k| < |z| < 1/\max_{k \neq 0} |\lambda_k|$, which includes the unit circle. The power spectrum is defined by a substitution of z in $\Phi_1(z)$:

$$\begin{aligned}\Phi_1(\omega) &\equiv \Phi_1(z) \Big|_{z=e^{j\omega}} \\ &= \tilde{\mathbf{a}}^T \mathbf{S}(\mathbf{I} - \tilde{\mathbf{T}}e^{j\omega})^{-1} \tilde{\mathbf{a}} + \tilde{\mathbf{a}}^T \mathbf{S}(\mathbf{I} - \tilde{\mathbf{T}}e^{-j\omega})^{-1} \tilde{\mathbf{a}} - \tilde{\mathbf{a}}^T \mathbf{S} \tilde{\mathbf{a}}.\end{aligned}\quad (\text{II.30})$$

Now consider the power spectrum for $\phi_2(n)$. Recall that $\phi_2(n) = \mu_y^2$. For $\phi_2(n)$, the power spectrum is given by

$$\Phi_2(\omega) = 2\pi\mu_y^2 \sum_{k=-\infty}^{\infty} \delta(\omega - 2\pi k) \quad (\text{II.31})$$

where $\delta(\omega)$ is the unit impulse function; i.e., $\int_{-\infty}^{\infty} f(t)\delta(t)dt = f(0)$ for any function f which is continuous at $t=0$. This can be confirmed by taking the inverse transform of the resultant:

$$\begin{aligned}\frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \Phi_2(\omega) e^{j\omega} d\omega &= \frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sum_{k=-\infty}^{\infty} 2\pi\mu_y^2 \delta(\omega - 2\pi k) e^{j\omega} d\omega \\ &= \mu_y^2 \sum_{k=-\infty}^{\infty} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \delta(\omega - 2\pi k) e^{j\omega} d\omega \\ &= \mu_y^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \delta(\omega) d\omega \\ &= \mu_y^2 \\ &= \phi_2(n).\end{aligned}$$

The power spectrum of the MCRP is then found by adding the terms in Equations (II.30) and (II.31):

$$\begin{aligned}\Phi(\omega) &= \Phi_1(\omega) + \Phi_2(\omega) \\ &= \tilde{\mathbf{a}}^T \mathbf{S}(\mathbf{I} - \tilde{\mathbf{T}}e^{j\omega})^{-1} \tilde{\mathbf{a}} + \tilde{\mathbf{a}}^T \mathbf{S}(\mathbf{I} - \tilde{\mathbf{T}}e^{-j\omega})^{-1} \tilde{\mathbf{a}} - \tilde{\mathbf{a}}^T \mathbf{S} \tilde{\mathbf{a}} + \sum_{k=-\infty}^{\infty} 2\pi\mu_y^2 \delta(\omega - 2\pi k)\end{aligned}\quad (\text{II.32})$$

As an example, the power spectrum for the MCRP illustrated previously in Figure 2.2 is given below in Figure 2.5. A few comments are perhaps in order. The power spectrum is fairly ordinary looking: there are no sharp corners or jagged edges. In fact, it looks like the frequency response one might obtain from a band-pass filter. But then, exactly what were we expecting? For, while the power spectrum expression in Equation (II.32) is correct, it is however difficult to interpret: exactly how do T and \mathbf{a} influence the power spectrum? To help answer this question, consider the special case explored in the following section.

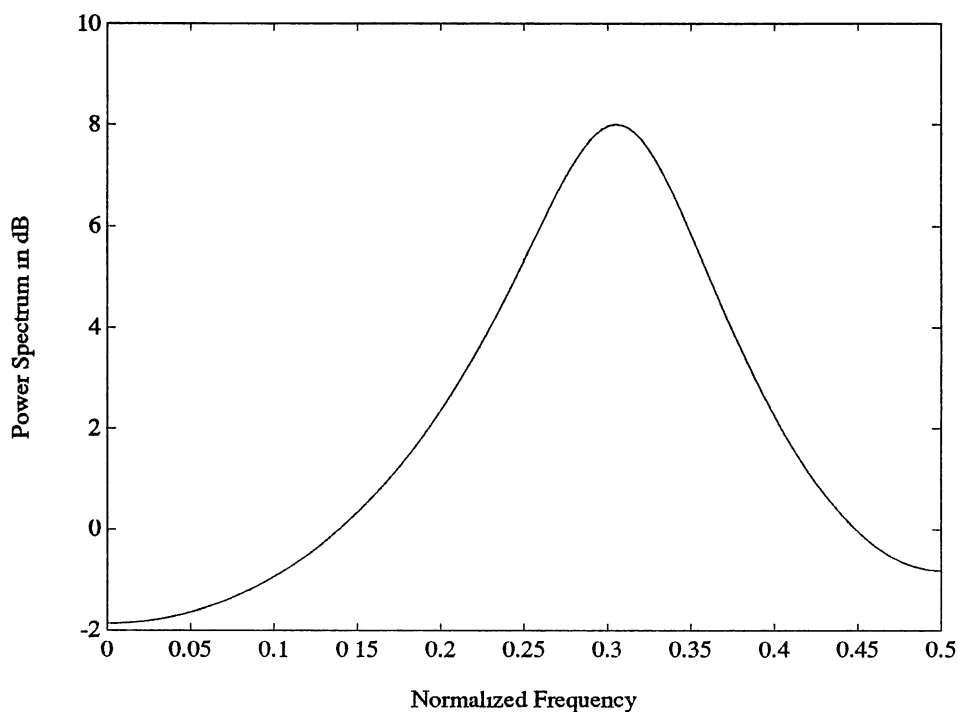


Figure 2.5. Power Spectrum of Example MCRP.

Power Spectrum of MCRPs:

Diagonable Transition Probability Matrix

Suppose that, for a given MCRP, the transition probability matrix T is diagonable.

Assume without loss of generality that $\mu_y = 0$. For this situation it is advantageous to

begin with the autocorrelation equation $\phi(n) = \sum_{k=1}^N \beta_k \lambda_k^{|n|}$ as in Equation (II.23). Then,

$$\begin{aligned} \Phi^+(z) &= \sum_{n=0}^{\infty} \left(\sum_{k=1}^{N-1} \beta_k \lambda_k^n \right) z^{-n} \\ &= \sum_{k=1}^{N-1} \beta_k \sum_{n=0}^{\infty} (\lambda_k z^{-1})^n. \end{aligned} \quad (\text{II.33})$$

Noting that $\sum_{n=0}^{\infty} r^n = 1/(1-r)$ for $|r| < 1$, $\sum_{n=0}^{\infty} (\lambda_k z^{-1})^n = 1/(1 - \lambda_k z^{-1})$ for $|z| > |\lambda_k|$.

Thus,

$$\Phi^+(z) = \sum_{k=1}^{N-1} \frac{\beta_k}{1 - \lambda_k z^{-1}}$$

for $|z| > \max_{k \neq 0} |\lambda_k|$.

Since $\Phi^-(z) = \Phi^+(z^{-1})$, it follows that $\Phi^-(z) = \sum_{k=1}^{N-1} \beta_k / (1 - \lambda_k z^{-1})$ for

$|z| < 1/(\max_{k \neq 0} |\lambda_k|)$. Thus,

$$\Phi(z) = \sum_{k=1}^{N-1} \frac{\beta_k}{1 - \lambda_k z^{-1}} + \sum_{k=1}^{N-1} \frac{\beta_k}{1 - \lambda_k z} - \phi(0)$$

for $\max_{k \neq 0} |\lambda_k| < |z| < 1/(\max_{k \neq 0} |\lambda_k|)$. Now, noting that $\phi(0) = \sum_{k=1}^{N-1} \beta_k \lambda_k^0 = \sum_{k=1}^{N-1} \beta_k$, this

can be expressed as

$$\Phi(z) = \sum_{k=1}^{N-1} \beta_k \left(\frac{1}{1 - \lambda_k z^{-1}} + \frac{1}{1 - \lambda_k z} - 1 \right)$$

The power spectrum is then given by

$$\Phi(\omega) = \sum_{k=1}^{N-1} \beta_k \left(\frac{1}{1 - \lambda_k e^{j\omega}} + \frac{1}{1 - \lambda_k e^{-j\omega}} - 1 \right). \quad (\text{II.34})$$

While the power spectrum expression above may not appear familiar, the following is true: one would expect the power spectrum to have maxima at frequencies corresponding to the phase angles of the eigenvalues. To explain, suppose $\lambda_k = r_k e^{j\theta_k}$. Then, $1 - \lambda_k e^{-j\omega} = 1 - r_k e^{j(\theta_k - \omega)}$, which will be minimum at $\omega = \theta_k$. Since the power spectrum equation includes the term $1/(1 - \lambda_k e^{-j\omega})$, one would expect the power spectrum to have maxima at frequencies corresponding to the phase angles of the eigenvalues. Further, the magnitude of λ_k will determine how sharp the maxima is at frequencies around θ_k : as $|\lambda_k|$ increases, the maxima will become more narrow. Having seen how the eigenvalues influence the shape of the power spectrum, note that the $\{\beta_k\}$ coefficients can be thought of as weights which control the influence of the various eigenvalues.

This analysis of the MCRP power spectrum leads to a rather interesting interpretation: the eigenvalues of the transition probability matrix behave like "poles" of the process. That is, there exists a connection between the eigenvalues of T and the poles of a linear transfer function for a discrete-time system. This fact shall be explored in depth in Chapter IV. However, this linkage is noted at this time to inform the reader that while the power spectrum expression in Equation (II.34) is perhaps unfamiliar in form, it should not be unfamiliar in function.

Let us now consider the power spectrum for the MCRP illustrated previously. The power spectrum for this MCRP was shown in Figure 2.5 in the previous section. The transition probability matrix for this MCRP is diagonalizable, and its eigenvalues are $\lambda_0 = 1$, $\lambda_1 = .7058e^{1.93j}$, and $\lambda_2 = .7058e^{-1.93j}$. Thus, a peak in the power spectrum is expected at a normalized frequency of $1.93/2\pi \approx .31$. From Figure 2.5, one can see that this is in fact where the power spectrum obtains its maximum. Also, note that, had the eigenvalue magnitude been closer to unity, the peak would have been more pronounced.

Power Spectrum of MCRPs:

Non-Diagonable Transition Probability Matrix

Since we have explored the case when T is diagonal, a natural concern is the effect on the power spectrum when T is not diagonal. However, when T is not diagonal, the expression for the power spectrum is somewhat more complicated. To illustrate, consider a simple, hypothetical example. Suppose T is a 3x3 transition probability matrix with Jordan form

$$J = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$$

That is, T has two eigenvalues: 1, which has unit multiplicity; and λ , which has algebraic multiplicity 2, but geometric multiplicity 1. Further, suppose the modal matrix of T is given in block vector notation as

$$V = (\mathbf{v}_0 \quad \mathbf{v}_1 \quad \mathbf{v}_2)$$

where $\{\mathbf{v}_i\}$ are right eigenvectors and

$$V^{-1} = \begin{pmatrix} \mathbf{u}_0^T \\ \mathbf{u}_1^T \\ \mathbf{u}_2^T \end{pmatrix}$$

where $\{\mathbf{u}_i\}$ are left eigenvectors. Then, for $n \geq 0$, T^n may be expressed as

$$T^n = (\mathbf{v}_0 \quad \mathbf{v}_1 \quad \mathbf{v}_2) \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda^n & n\lambda^{n-1} \\ 0 & 0 & \lambda^n \end{pmatrix} \begin{pmatrix} \mathbf{u}_0^T \\ \mathbf{u}_1^T \\ \mathbf{u}_2^T \end{pmatrix}$$

The spectral representation of T^n is then

$$T^n = \mathbf{v}_0 \mathbf{u}_0^T (1) + (\mathbf{v}_1 \mathbf{u}_1^T + \mathbf{v}_2 \mathbf{u}_2^T) (\lambda^n) + \mathbf{v}_1 \mathbf{u}_2^T (n\lambda^{n-1}).$$

The key difference from the diagonal case is the presence of the $n\lambda^{n-1}$ term in addition to the λ^n term. Using the spectral representation, the autocorrelation is

$$\begin{aligned}\phi(n) &= \mathbf{a}^T \mathbf{S} \mathbf{T}^n \mathbf{a} \\ &= (\mathbf{a}^T \mathbf{S} \mathbf{v}_0 \mathbf{u}_0^T \mathbf{a}) (1) + [\mathbf{a}^T \mathbf{S} (\mathbf{v}_1 \mathbf{u}_1^T + \mathbf{v}_2 \mathbf{u}_2^T) \mathbf{a}] (\lambda^n) + (\mathbf{a} \mathbf{S} \mathbf{v}_1 \mathbf{u}_2^T \mathbf{a}) (n\lambda^{n-1}) \\ &= \beta_0(1) + \beta_1(\lambda^n) + \beta_2(n\lambda^{n-1}).\end{aligned}\quad (\text{II.35})$$

As before, $\beta_0 = 0$ by the properties of the eigenvalues of T associated with the eigenvalue $\lambda_0 = 1$. Therefore, the autocorrelation reduces to $\phi(n) = \beta_1 \lambda^n + \beta_2 n \lambda^{n-1}$. Now consider the one-sided z-transform of $\phi(n)$:

$$\begin{aligned}\Phi^+(z) &= \sum_{n=0}^{\infty} \phi(n) z^{-n} \\ &= \sum_{n=0}^{\infty} \beta_1 \lambda^n z^{-n} + \sum_{n=0}^{\infty} \beta_2 n \lambda^{n-1} z^{-n}\end{aligned}$$

Now, using the fact that $\sum_{n=0}^{\infty} x^n = 1/(1-x)$ and a little manipulation,

$$\begin{aligned}\Phi^+(z) &= \frac{\beta_1}{1-\lambda z^{-1}} + \beta_2 \frac{\partial}{\partial \lambda} \left(\sum_{n=0}^{\infty} \lambda^n z^{-n} \right) \\ &= \frac{\beta_1}{1-\lambda z^{-1}} + \beta_2 \frac{\partial}{\partial \lambda} \left(\frac{1}{1-\lambda z^{-1}} \right) \\ &= \frac{\beta_1}{1-\lambda z^{-1}} + \frac{\beta_2 z^{-1}}{(1-\lambda z^{-1})^2}.\end{aligned}$$

Next, using $\Phi(z) = \Phi^+(z) + \Phi^+(z^{-1}) - \phi(0)$, and $\Phi(\omega) = \Phi(z) |_{z=e^{j\omega}}$,

$$\Phi(\omega) = \beta_1 \left(\frac{1}{1-\lambda e^{-j\omega}} + \frac{1}{1-\lambda e^{j\omega}} \right) + \beta_2 \left(\frac{e^{-j\omega}}{(1-\lambda e^{-j\omega})^2} + \frac{e^{j\omega}}{(1-\lambda e^{j\omega})^2} \right) - \phi(0) \quad (\text{II.36})$$

Again, the eigenvalues function as "poles" of the process. However, they behave somewhat differently than before because T is not diagonal. Still, there is a correspondence between these MCRPs and processes derived from linear filters. In this case, the

relationship is to a linear filter which has a repeated pole. Again, this parallel between MCRPs and processes derived from linear filters is mentioned only in passing at this time; a more thorough treatment shall be given in Chapter IV.

Hidden Markov Models

Hidden Markov Models (HMMs) are in some sense similar to Markov chain random processes: both are extensions of Markov chains. However, whereas an MCRP associates a real number with each state of a Markov chain, an HMM associates an entire random process with each state. Typically, the HMM remains in each state for a fixed duration. During that time, the output of the HMM is a sample from the random process associated with the current state.

The model is termed "hidden" because the underlying Markov structure cannot be determined through direct measurements of the output of the process. However, the output does provide clues as to the current state. HMMs, then, possess two layers of randomness: the Markov chain connecting the states; and the random processes composing the states. Processes such as HMMs which have two layers of randomness are termed *doubly stochastic*.

HMMs have proven to be a very useful tool in modeling certain types of nonstationary processes. Specifically, HMMs are best suited to nonstationary processes which possess "short-time" stationarity. That is, although the process is nonstationary, the statistics of the process change slowly enough that, over small segments of time, the statistics are relatively constant. For these types of processes, HMMs can serve as a structure to connect the short-time segments. In this case, the states of the HMM are models of the short-time sections of the nonstationary process.

As an example, speech is a nonstationary signal which exhibits short-time stationarity. Because of this, it is possible to use an HMM to model the interconnection between

the short-time sections of speech. In fact, HMMs have proven very useful in the areas of speech recognition and speaker identification (see, for example [Rab89]) and have recently been exploited for speech coding purposes [Far88, Dea89].

There are several common Markov structures commonly used for Hidden Markov Models. A few of these are illustrated below in Figure 2.6. The figure in (a) is known as a left-right model and is commonly used when the process displays no long-term correlation. On the other hand, a model such as in (b) would be more applicable if the process has only a few "states" which are recurring. For more information on HMMs, consult either of Rabiner's papers [Rab86, Rab89].

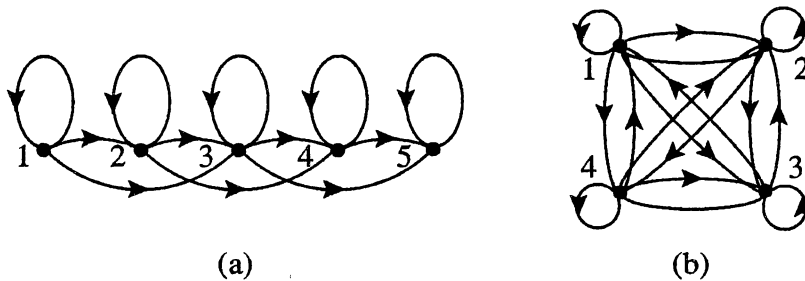


Figure 2.6. Common Hidden Markov Model Structures.

CHAPTER III

SYNTHESIS OF MARKOV CHAIN

RANDOM PROCESSES

This chapter details MCRP synthesis. The objectives are in three separate but related areas of synthesis: probability distribution function; power spectral density; and joint probability distribution/power spectral density. In all cases the constraints to be met are assumed to be known; estimation of probability distribution or power spectrum, if required, is not discussed in this chapter.

Probability Distribution Synthesis

Consider trying to locate an MCRP which has a specified probability distribution function. Since MCRPs are discrete-valued processes only probability functions from discrete-valued processes can be matched exactly. While it is possible to approximate the probability distribution of a continuous-valued process, consider synthesis for a discrete-valued processes first.

Suppose the desired discrete-valued process assumes the values $\{y_i\}$ with probability $\{p_i\}$ for $1 \leq i \leq N$. Clearly, the output map for the MCRP should be $\mathbf{a} = (y_1 \ y_2 \ \cdots \ y_N)^T$ and the stationary distribution should be $\mathbf{s} = (p_1 \ p_2 \ \cdots \ p_N)^T$. It only remains to find a transition probability matrix T which meets this stationary distribution. However, this problem does not have a unique solution as there are many transition probability matrices which will satisfy the conditions. One simple solution to this problem is the transition probability matrix

$$T = \begin{pmatrix} p_1 & p_2 & \cdots & p_N \\ p_1 & p_2 & \cdots & p_N \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ p_1 & p_2 & \cdots & p_N \end{pmatrix}$$

This matrix can be more compactly represented as $T = \underline{\mathbf{1}}_N s^T$, where $\underline{\mathbf{1}}_N$ is the usual vector of all ones. Then, s is the stationary distribution since $s^T T = s^T (\underline{\mathbf{1}}_N s^T) = (s^T \underline{\mathbf{1}}_N) s^T = s^T$.

Now consider the more difficult problem of approximating the probability distribution function of a continuous-valued process with a discrete-valued MCRP. The more conventional approach to this problem is to pass a white noise process through a properly chosen memoryless nonlinearity. However, consider the use of MCRPs in this endeavor.

Of the numerous ways which one could approach this problem, the following method appears to have the most merit. Let $f(x)$ be a probability density function of a continuous-valued stationary random process. Suppose the MCRP is to have N states. The basic idea is to divide the x -axis into N intervals and assign one member of the output map to each interval. Again, there are many ways which this could be done. However, suppose the intervals are chosen so as to yield intervals of equal probability. That is, a set $\{x_i\}$ is found such that

$$\int_{-\infty}^{x_1} f(x) dx = \int_{x_1}^{x_2} f(x) dx = \cdots = \int_{x_{N-2}}^{x_{N-1}} f(x) dx = \int_{x_{N-1}}^{\infty} f(x) dx = \frac{1}{N} \quad (\text{III.1})$$

For convenience of notation, let $x_0 = -\infty$, and $x_N = \infty$. Then the i^{th} interval is (x_{i-1}, x_i) for $1 \leq i \leq N$. Based on this set of intervals, a logical choice for a_i is the centroid of the i^{th} interval:

$$\begin{aligned}
a_i &= \frac{\int_{x_{i-1}}^{x_i} xf(x)dx}{\int_{x_{i-1}}^{x_i} f(x)dx} \\
&= N \int_{x_{i-1}}^{x_i} xf(x)dx \tag{III.2}
\end{aligned}$$

Next, let T be any regular *doubly stochastic* matrix. A doubly stochastic matrix is a matrix for which both T and T^T are stochastic; that is, every row and every column of T sum to unity, respectively. Note that the stationary distribution for any regular, doubly stochastic matrix is $s = (1/N \quad 1/N \quad \cdots \quad 1/N)^T$.

Under the above conditions, (T, a, s) defines a stationary random process whose probability distribution function should approximate that of the given process. In fact, by definition, the mean of the MCRP is equivalent to the mean of the given process:

$$\begin{aligned}
\mu_{\text{MCRP}} &= \sum_{i=1}^N s_i a_i \\
&= \sum_{i=1}^N \left(\frac{1}{N} \right) N \int_{x_{i-1}}^{x_i} xf(x)dx \\
&= \int_{-\infty}^{\infty} xf(x)dx \\
&= \mu_y
\end{aligned}$$

As an example, consider a random process which has a Laplacian density function with unit variance:

$$f(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \quad \text{for } -\infty < x < \infty.$$

Suppose without loss of generality that N is even. In this case, the output map can be shown to be

$$a_i = \begin{cases} \frac{1}{\sqrt{2}} \left\{ \log\left(\frac{2}{N}\right) - 1 \right\} & \text{for } i = 1 \\ \frac{i}{\sqrt{2}} \left\{ \log\left(\frac{2i}{N}\right) - 1 \right\} - \frac{(i-1)}{\sqrt{2}} \left\{ \log\left(\frac{2(i-1)}{N}\right) - 1 \right\} & \text{for } i = 2, \dots, \frac{N}{2} \\ -a_{N-i+1} & \text{for } i = \frac{N}{2} + 1, \dots, N \end{cases}$$

Figure 3.1 gives sample outcomes from the Laplacian process (top plot) and MCRP with $N=20$ (bottom plot). As can be seen, the waveforms appear similar. The only noticeable difference is that the MCRP waveform appears to be clipped. This is because the MCRP has a definite maximum value, which is equal to the largest a_i . In contrast, the Laplacian process has no such limit.

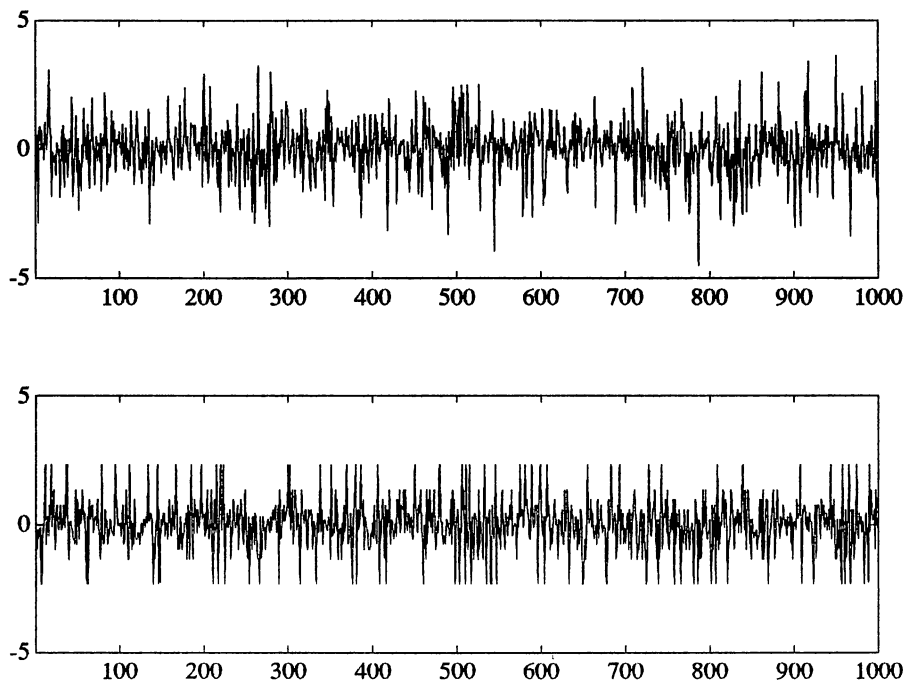


Figure 3.1. Waveforms for Laplacian Process (top) and MCRP (bottom).

As stated above, the mean of the MCRP equals the mean of the given process.

However, the higher order moments are not guaranteed to be equal. For the above example, Figure 3.2 plots the variance, $E\{(y(n) - \mu_y)^2\}$, and skew, $E\{|y(n)|^3\}$, as a function of the number of states N . The true value of the variance is unity and the skew is $\sqrt{4.5} \approx 2.12$. Note that, as N increases, it appears that the MCRP moments become closer to the correct values. While this approximation technique has not been tested on a great number of processes, the author believes that this statement is true in general.

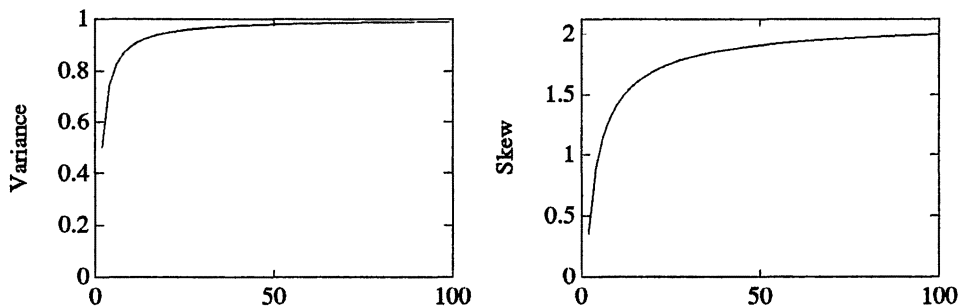


Figure 3.2. Variance and Skew of MCRP as a Function of N .

Power Spectral Density Synthesis

The second task of this chapter is to explore synthesis using the power spectral density. Specifically, we wish to find a random process which possesses a desired power spectrum. A standard solution to this problem is to assume the process to be the output of a linear filter driven by white noise. The core of the problem is to determine the filter parameters to achieve the desired spectrum.

However, consider trying to locate an MCRP which possesses a given power spectral density. In general, one would probably approach this problem in two steps, as illustrated below in Figure 3.3. First, the power spectrum would need to be expressed in a form similar to the MCRP power spectrum. Specifically, if T is required to be diagonal, the given power spectrum must conform the MCRP power spectrum equation

$$\Phi(\omega) = \sum_{k=1}^{N-1} \beta_k \left(\frac{1}{1 - \lambda_k e^{j\omega}} + \frac{1}{1 - \lambda_k e^{-j\omega}} - 1 \right) \quad (\text{III.3})$$

where it has been assumed without loss of generality that $\mu = 0$ for the desired process.

Manipulating the power spectrum into the correct form is equivalent to determining the necessary MCRP spectral parameters $\{\lambda_k\}$ and $\{\beta_k\}$. Now, the actual determination of these parameters is an application-specific problem. Consequently, for the purposes of this section, the desired $\{\lambda_k\}$ and $\{\beta_k\}$ shall be assumed to be known.

The second phase of this synthesis problem involves finding T and a which yield the desired eigenvalues $\{\lambda_k\}$ and spectral coefficients $\{\beta_k\}$. It is this problem which is the focus of this section.

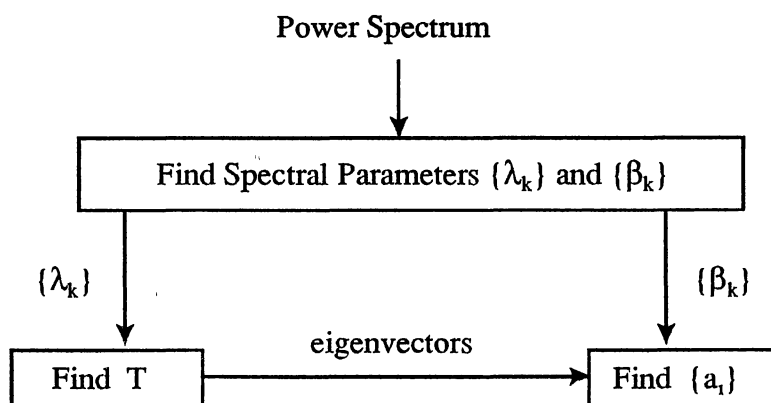


Figure 3.3. Block Diagram of Power Spectrum Synthesis.

Let us take a closer look at this problem. Recall from Chapter II that if T is diagonalizable it can be expressed as $T = \sum_k (\mathbf{v}_k \mathbf{u}_k^T) \lambda_k$ where \mathbf{v}_k (\mathbf{u}_k) is the right (left) eigenvector of T associated with the eigenvalue λ_k . Also from Chapter II, recall that the spectral coefficients $\{\beta_k\}$ were given by $\beta_k = \tilde{\mathbf{a}}^T S \mathbf{v}_k \mathbf{u}_k^T \tilde{\mathbf{a}}$ where S is a diagonal matrix of the stationary state probabilities, and $\tilde{\mathbf{a}} = (\tilde{a}_1 \quad \tilde{a}_2 \quad \cdots \quad \tilde{a}_N)$ where $\tilde{a}_i = a_i - \mu_y$ and where μ_y is the mean of the process. By a previous assumption, $\mu_y = 0$ so $\tilde{\mathbf{a}} = \mathbf{a}$.

In essence, one is left with two separate but related problems: finding T with eigenvalues $\{\lambda_k\}$; and finding \mathbf{a} which results in $\{\beta_k\}$. However, as the spectral coefficients are functions of the eigenvectors of T , it is clear that the latter problem cannot be solved until T is found. Consequently, T and its eigenvalues must first be examined.

A Stochastic Eigenvalue Problem

Consider trying to find an $N \times N$ stochastic matrix T which possesses a given set of N eigenvalues. If T could be any $N \times N$ matrix this would be an almost trivial problem. However, T must be a stochastic matrix. Because of this, finding a transition probability matrix with a given set of eigenvalues is a very difficult problem. In fact, there may be no solution at all.

One obstacle may be that one or more of the specified values are not eigenvalues for any $N \times N$ stochastic matrix. The Russian mathematician F.I. Karpelevic [Kar51] has shown that the eigenvalues of stochastic matrices are limited to a closed set in the complex plane which is a function of N . This set is quite difficult to describe exactly. However, a good approximation can be made fairly easily, and we shall proceed to define this set. (For an exact specification, see Karpelevic's original paper [Kar51], or the description by de Oliveira [Oli68].) To describe this set, define P_L to be the convex set in the complex plane for which the boundary is the polygon whose vertices are the L^{th} roots of unity. That is,

$$\mathbf{P}_L = \left\{ \sum_{k=0}^{L-1} \alpha_k e^{j2\pi k/L} : \sum_{k=0}^{L-1} \alpha_k = 1, 0 \leq \alpha_k \leq 1 \forall k \right\}$$

Then, the set M_N of all complex numbers which are eigenvalues for some stochastic $N \times N$ matrix can be approximated by

$$M_N = \bigcup_{i=2}^N \mathbf{P}_i. \quad (\text{III.4})$$

These sets are illustrated in Figure 3.4 for several values of N . Note that, as N increases, M_N contains more and more of the unit disc. In fact, in the limit as N goes to infinity, the set M_N approaches the union of all points interior to the unit disc and all points $e^{j2\pi k/L}$ on the unit circle where k and L are integers.

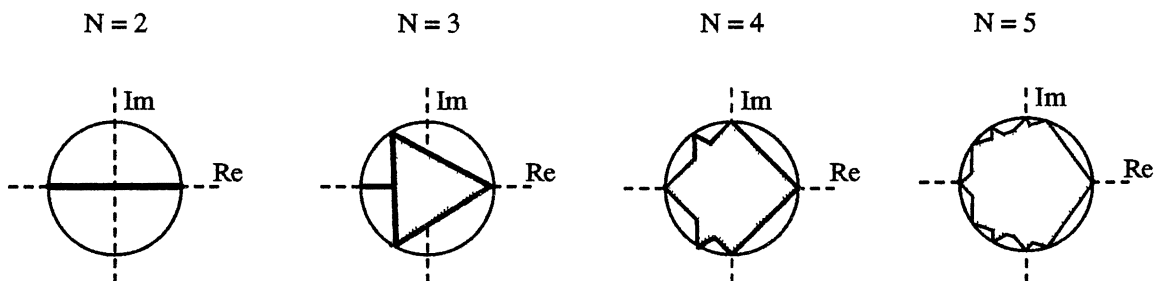


Figure 3.4. Domain of Eigenvalues for Stochastic $N \times N$ Matrices.

Unfortunately, even if all N of the specified eigenvalues lie in M_N , it is still possible that there is no $N \times N$ stochastic matrix which possesses these eigenvalues. Consider the region D_N of complex N -dimensional space defined by the requirement that $(\lambda_0, \lambda_1, \dots, \lambda_{N-1}) \in D_N$ if and only if there exists an $N \times N$ stochastic matrix for which $\lambda_0, \lambda_1, \dots,$ and λ_{N-1} are eigenvalues. However, determination of D_N is a very difficult

problem, and published papers have limited themselves to the real eigenvalues of stochastic matrices [Sul49, Per64]. Since this research is equally interested in complex eigenvalues, results from this area of investigation offer little help.

Another approach to this stochastic eigenvalue problem is to determine bounds on the eigenvalues in terms of the elements of T . Such is the approach with Gersgorin's Theorem [Gan64, Lan85], which states that any eigenvalue of a matrix $T = (t_{ij})$ must lie in at least one of the disks in the complex plane defined by $|z - t_{ii}| \leq \sum_{j \neq i} |t_{ij}|$. There are also other bounds such as the Deutsch bound [Zen72, Sen81] for the non-unity eigenvalues, which takes the form

$$|\lambda| \leq \frac{1}{2} \max_{i,i'} \sum_{j=1}^N |t_{ij} - t_{i'j}|.$$

However, it is not clear how one could make use of these bounds to find T given a set of eigenvalues.

The consequences of the above paragraphs are that, given current knowledge, it is not possible to solve directly for an $N \times N$ stochastic matrix T given a set of N eigenvalues $\{\lambda_k\}$. Further, the discussion implies that no solution exists for certain sets of eigenvalues which, at this time, have not yet been clearly defined.

Despite this, the author offers the following conjecture here, which shall be restated and proved as a theorem in a later section of this chapter:

Conjecture: Given any set of N eigenvalues $\{\lambda_k\}_{k=0}^{N-1}$ where $\lambda_0 = 1$ and $|\lambda_k| < 1$ for $k \neq 0$, there exists an $N' \times N'$ diagonalizable, regular, stochastic matrix T with eigenvalues $\{\lambda'_k\}_{k=0}^{N'-1}$ such that

- (1) $N' \geq N$,
- (2) $\lambda'_0 = \lambda_0$ and $\lambda'_k \approx \lambda_k$ for $0 < k < N$,

and (3) $|\lambda'_k| < 1$ for $0 < k < N'$.

In simpler terms, this conjecture says that the desired eigenvalues will be a subset of the eigenvalues from a larger stochastic matrix. Thus, by increasing the size of T , the limita-

tions imposed by the domain of eigenvalues can be overcome. It should be noted that a disadvantage of this approach is that there will likely be undesirable eigenvalues.

However, as far as the autocorrelation and the power spectrum are concerned, if $\beta_k = 0$ for any undesired eigenvalues, the result will be the same as desired.

Although the conjecture claims the existence of a desirable stochastic matrix, it does not state how such a matrix may be found. To find such a matrix, let us consider several classes of stochastic matrices which have desirable eigen decompositions. Ideally, one would like a class of matrices for which there is a simple and well-behaved relation between the eigenvalues and the elements of T . Such matrix classes would not only let us achieve the desired eigenvalues $\{\lambda_k\}$, as well as manipulate the spectral coefficients $\{\beta_k\}$ in a simple fashion. Classes of matrices with these properties do exist as shall be shown in the following sections.

A Circulant Structure

The use of circulant stochastic matrices for MCRPs was proposed by Mullis and Steiglitz [Mul72], and the material in this section is based on their work. For an $N \times N$ circulant matrix C , the elements are related by the equation

$$[C]_{i,j} = c_{(j-i) \bmod N}. \quad (\text{III.5})$$

For example, a 3×3 circulant matrix would be given by

$$C = \begin{pmatrix} c_0 & c_1 & c_2 \\ c_2 & c_0 & c_1 \\ c_1 & c_2 & c_0 \end{pmatrix}.$$

The eigenvalue-eigenvector decomposition for an $N \times N$ circulant matrix C is well-known and is given by [Dav79, Her86]

$$C = F_{DFT} \Lambda F_{DFT}^{-1}. \quad (\text{III.6})$$

In this equation, F_{DFT} is the $N \times N$ matrix

$$F_{DFT} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & e^{-j2\pi/N} & e^{-j4\pi/N} & \cdots & e^{-j2(N-1)\pi/N} \\ 1 & e^{-j4\pi/N} & e^{-j8\pi/N} & \cdots & e^{-j4(N-1)\pi/N} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & e^{-j2(N-1)\pi/N} & e^{-j4(N-1)\pi/N} & \cdots & e^{-j2(N-1)(N-1)\pi/N} \end{pmatrix} \quad (\text{III.7})$$

and Λ is the diagonal matrix

$$\Lambda = \begin{pmatrix} \lambda_0 & & & & \\ & \lambda_1 & & & \\ & & \lambda_2 & & \\ & & & \ddots & \\ & & & & \lambda_{N-1} \end{pmatrix} \quad (\text{III.8})$$

where $\lambda_k = \sum_{i=0}^{N-1} c_i e^{-j2\pi k i/N}$, which is just the discrete Fourier transform (DFT) of

$\{c_0, c_1, \dots, c_{N-1}\}$. While the above applies to all circulant matrices, if the matrix is also stochastic, then note that, like all stochastic matrices, there is an eigenvalue equal to one:

$$\lambda_0 = \sum_{i=0}^{N-1} c_i e^{j0} = 1.$$

For a circulant stochastic matrix T_C , the Markov chain is irreducible as long as

$c_0 \neq 1$. Now consider the eigenvalues of a circulant stochastic matrix. For $k > 0$, since

$\lambda_k = \sum_{i=0}^{N-1} c_i e^{-j2\pi k i/N}$, λ_k is just a convex combination of the points

$\{1, e^{-j2\pi k/N}, e^{-j2\pi 2k/N}, \dots, e^{-j2\pi(N-1)k/N}\}$ in complex space. Further, since $0 \leq c_i \leq 1$, and

$\sum_{i=0}^{N-1} c_i = 1$, λ_k must clearly lie on or interior to the polygon in complex space formed by connecting the points $\{1, e^{-j2\pi k/N}, e^{-j2\pi 2k/N}, \dots, e^{-j2\pi(N-1)k/N}\}$, which all lie on the unit circle.

If only one element of $\{c_0, c_1, \dots, c_{N-1}\}$ is nonzero, λ_k must equal one of these points. In this case, $|\lambda_k| = 1$, which implies that the Markov chain would not be regular. Con-

versely, if two or more elements of $\{c_0, c_1, \dots, c_{N-1}\}$ are nonzero, λ_k will lie in the interior of the unit circle, so that $|\lambda_k| < 1$. Thus, a necessary condition for the circulant Markov chain to be regular is that at least two elements of $\{c_0, c_1, \dots, c_{N-1}\}$ are nonzero.

As shown in Chapter II, \mathbf{s} is the left eigenvector associated with the eigenvalue $\lambda_0 = 1$. For circulant stochastic matrices, this eigenvector is the first column of \mathbf{F}_{DFT}^{-1} . But, it is true that $\mathbf{F}_{DFT}^{-1} = (1/N)\mathbf{F}_{DFT}^*$, where $*$ refers to complex conjugate transpose. Consequently,

$$\begin{aligned}\mathbf{s} &= \left(\frac{1}{N} \quad \frac{1}{N} \quad \dots \quad \frac{1}{N} \right)^T \\ &= \frac{1}{N} \mathbf{1}_N,\end{aligned}$$

and therefore, the diagonal matrix \mathbf{S} consisting of stationary probabilities is given by

$$\mathbf{S} = (1/N)\mathbf{I}_N.$$

Having determined the stationary probabilities, the spectral coefficients $\{\beta_k\}$ may be found. For λ_k , the right eigenvector is $\mathbf{v}_k = (\omega^0 \quad \omega^1 \quad \omega^2 \quad \dots \quad \omega^{N-1})^T$ where $\omega = e^{-j2\pi k/N}$ and the left eigenvector is $\mathbf{u}_k = (1/N)(\mathbf{v}_k^*)^T$. Thus,

$$\begin{aligned}\beta_k &= \mathbf{a}^T \mathbf{S} \mathbf{v}_k \mathbf{u}_k^T \mathbf{a} \\ &= \frac{1}{N} (\mathbf{a}^T \mathbf{v}_k) (\mathbf{u}_k^T \mathbf{a}).\end{aligned}$$

This can be further simplified by using DFT notation: let

$$A(k) = \sum_{i=0}^{N-1} a_{i+1} e^{-j2\pi k i/N},$$

which is just the DFT of $\{a_i\}$. Note that $\mathbf{a}^T \mathbf{v}_k = A(k)$ and $\mathbf{u}_k^T \mathbf{a} = A^*(k)/N$. Making these substitutions in the above equation for β_k yields

$$\begin{aligned}\beta_k &= \frac{A^*(k)A(k)}{N^2} \\ &= \frac{|A(k)|^2}{N^2}.\end{aligned}\tag{III.9}$$

The power spectrum for MCRPs with circulant transition probability matrices may therefore be expressed as

$$\Phi(\omega) = \sum_{k=1}^{N-1} \frac{|A(k)|^2}{N^2} \left(\frac{1}{1 - \lambda_k e^{-j\omega}} + \frac{1}{1 - \lambda_k e^{j\omega}} - 1 \right).\tag{III.10}$$

The benefit of MCRP with circulant transition probability matrices is that there is a straightforward relationship between the spectral parameters $\{\lambda_k\}$ and $\{\beta_k\}$ and the MCRP parameters in T and a .

From the previous discussion, the eigenvalues of a circulant matrix are given by the DFT of $\{c_0, c_1, \dots, c_{n-1}\}$. It follows that, given a set of eigenvalues $\{\lambda_k\}$, the set $\{c_i\}$ can be found by taking the inverse DFT of the eigenvalues:

$$c_i = \frac{1}{N} \sum_{k=0}^{N-1} \lambda_k e^{j2\pi k i / N}.\tag{III.11}$$

While this seems straightforward enough, there is one problem: for an arbitrary set of eigenvalues, there is no guarantee that the resulting values for $\{c_i\}$ will satisfy $0 \leq c_i \leq 1$ and $\sum_{i=0}^{N-1} c_i = 1$. However, Siegel et. al. [Sie76] have developed a linear programming algorithm which reportedly results in a set of eigenvalues close to the specified eigenvalues. Here, Siegel made implicit use of the above conjecture by making the size of the matrix much larger than the number of eigenvalues given. For example, Siegel reported using a 256-state MCRP to match 5 pairs of complex eigenvalues.

Another problem with circulant matrices is that the β_k values are restricted to being real and positive: in general, these constants can be complex. In fact, for the MCRP

example in Chapter II, there were complex constants: $\beta_1 = 1.06 + 0.09j$ and $\beta_2 = \beta_1^*$. As shall be seen in the next chapter, this restriction hinders the ability to match data and MCRP power spectra.

It is worthwhile comparing \mathbf{M}_N , the set composed of all eigenvalues of $N \times N$ stochastic matrices, with the corresponding region, denoted here as \mathbf{B}_N , restricted to circulant stochastic matrices. Whereas $\mathbf{M}_N = \cup_{i=2}^N \mathbf{P}_i$, $\mathbf{B}_N = \mathbf{P}_N$. The differences between the sets \mathbf{M}_N and \mathbf{B}_N are illustrated in Figure 3.5 for a few values of N . Note that the difference between the two sets is small. Thus, circulant stochastic matrices are nearly optimal in the sense that they generate nearly all possible eigenvalues.

Having seen the usefulness and limitations of MCRPs with circulant stochastic matrices, a natural question is whether there are other classes of matrices which would prove useful. Steinhardt, for one, has considered the class of normal matrices [Ste83], which are diagonalizable matrices such that the modal matrix V satisfies $V^{-1} = V^*$; i.e., $T = V\Lambda V^*$, Λ is a diagonal matrix of eigenvalues. However, Steinhardt found that these matrices also require β_k to be real and positive. Additionally, he offered no method to find T given a set of eigenvalues. Consequently, since the β_k values must be real and positive, and since a linear programming algorithm must be used to find the elements of T_C for circulant stochastic matrices, there is certainly reason to search for other classes of matrices.

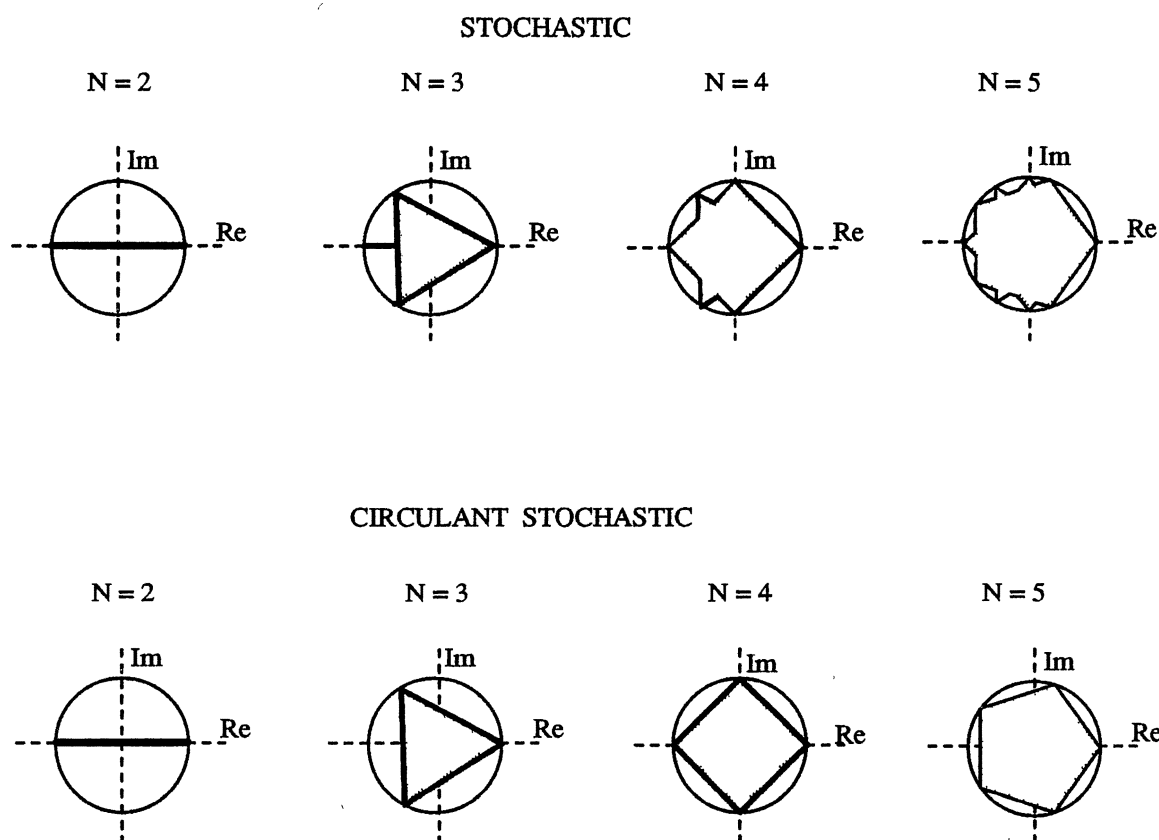


Figure 3.5. Domain of Eigenvalues for Stochastic and Circulant Stochastic $N \times N$ Matrices.

A Summed Structure

One of the main problems with the circulant structure is that it is difficult to synthesize a process which requires a large number of eigenvalues. One approach to solving complex problems such as this is to try to separate the problem into smaller, hopefully simpler, problems. Such is the approach with the summed structure discussed in this section.

The summed structure is based on the fact that if several independent processes are summed, the resulting power spectrum is the sum of the power spectrum for each pro-

cess. Formally, if $\{y_i(n)\}$ are independent random processes, and $w(n) = \sum_i y_i(n)$, then $\Phi_w(\omega) = \sum_i \Phi_{y_i}(\omega)$. The concept here is to partition the set of spectral parameters, $\{\lambda_k\}$ and $\{\beta_k\}$, and find MCRPs which meet the parameters for each partition. Then, these MCRPs, which for clarity shall be called *subchains*, can be summed to yield the desired process.

One drawback of this approach is that it would require the operation of several chains in parallel. Often, one would rather have to deal with only one process than many. As it turns out, it is possible to combine these subchains into one larger MCRP. First, if $\{T_i\}$ are the transition probability matrices for the subchains, then let

$$T = T_1 \otimes T_2 \otimes \cdots \otimes T_M$$

where \otimes is the Kronecker product. If T_i is $N_i \times N_i$, then T will be $N \times N$ where $N = \prod_i N_i$.

That is, T will have $\prod_i N_i$ states. The output map for T is more difficult to annotate.

However, as the Kronecker product is the element-by-element product of the T_i matrices, the desired output map for T is found from an element-by-element sum of the output maps $\{a_i\}$.

Perhaps this is best illustrated by an example. Suppose that we have two subchains; one has three states and the other has two states. Let the transition probability matrices be given by

$$T_1 = \begin{pmatrix} t_{1,1} & t_{1,2} & t_{1,3} \\ t_{2,1} & t_{2,2} & t_{2,3} \\ t_{3,1} & t_{3,2} & t_{3,3} \end{pmatrix}, \quad T_2 = \begin{pmatrix} \tau_{1,1} & \tau_{1,2} \\ \tau_{2,1} & \tau_{2,2} \end{pmatrix},$$

and the output maps by $a_1 = (a_1 \ a_2 \ a_3)^T$ and $a_2 = (\alpha_1 \ \alpha_2)^T$. Then,

$$T = T_1 \otimes T_2 = \begin{pmatrix} t_{1,1}\tau_{1,1} & t_{1,1}\tau_{1,2} & t_{1,2}\tau_{1,1} & t_{1,2}\tau_{1,2} & t_{1,3}\tau_{1,1} & t_{1,3}\tau_{1,2} \\ t_{1,1}\tau_{2,1} & t_{1,1}\tau_{2,2} & t_{1,2}\tau_{2,1} & t_{1,2}\tau_{2,2} & t_{1,3}\tau_{2,1} & t_{1,3}\tau_{2,2} \\ t_{2,1}\tau_{1,1} & t_{2,1}\tau_{1,2} & t_{2,2}\tau_{1,1} & t_{2,2}\tau_{1,2} & t_{2,3}\tau_{1,1} & t_{2,3}\tau_{1,2} \\ t_{2,1}\tau_{2,1} & t_{2,1}\tau_{2,2} & t_{2,2}\tau_{2,1} & t_{2,2}\tau_{2,2} & t_{2,3}\tau_{2,1} & t_{2,3}\tau_{2,2} \\ t_{3,1}\tau_{1,1} & t_{3,1}\tau_{1,2} & t_{3,2}\tau_{1,1} & t_{3,2}\tau_{1,2} & t_{3,3}\tau_{1,1} & t_{3,3}\tau_{1,2} \\ t_{3,1}\tau_{2,1} & t_{3,1}\tau_{2,2} & t_{3,2}\tau_{2,1} & t_{3,2}\tau_{2,2} & t_{3,3}\tau_{2,1} & t_{3,3}\tau_{2,2} \end{pmatrix},$$

and

$$a = \begin{pmatrix} a_1 + \alpha_1 \\ a_1 + \alpha_2 \\ a_2 + \alpha_1 \\ a_2 + \alpha_2 \\ a_3 + \alpha_1 \\ a_3 + \alpha_2 \end{pmatrix}$$

Note that the diagonal elements of T are similar to the elements of a . For example,

$[T]_{3,3} = t_{2,2}\tau_{1,1}$ while $[a]_3 = a_2 + \alpha_1$. In general, if $[T]_{i,i} = t_{j,j}\tau_{k,k}$ then $[a]_i = a_j + \alpha_k$.

Perhaps the most significant problem with this approach is the extremely large number of states. Since the number of states is $N = \prod_i N_i$, T can be very large even for a relatively small number of subchains. The structure introduced in the following section avoids this problem while maintaining the same basic approach.

A Ring Structure

The ring structure discussed in this section uses a similar approach to the summed structure but with relatively fewer states. This ring structure is illustrated in Figure 3.6. As before, several MCRPs are combined to form a larger process. However, in this case the processes are not added, but are connected together by a Markov chain. To explain further, refer to Figure 3.6 and suppose at some instant the chain is in any state of subchain 1. There is a probability of $(1 - \alpha_1)$ that the next state will also be in subchain 1; in

this case, the particular state depends upon T_1 . At the same time, there is a probability of α_1 that the next state will be in subchain 2; in this case, the particular state depends upon s_2 , the stationary distribution for subchain 2.

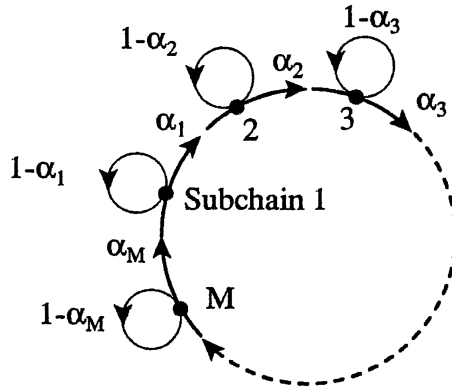


Figure 3.6. MCRP Ring Structure.

Using block matrix notation, the transition probability matrix for this structure can be expressed as

$$T = \begin{pmatrix} (1-\alpha_1)T_1 & \alpha_1 \mathbf{1}_{N_1} s_2^T & 0 & \cdots & \cdots & 0 \\ 0 & (1-\alpha_2)T_2 & \alpha_2 \mathbf{1}_{N_2} s_3^T & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \vdots & \vdots & \vdots & (1-\alpha_{M-1})T_{M-1} & \alpha_{M-1} \mathbf{1}_{N_{M-1}} s_M^T \\ \alpha_M \mathbf{1}_{N_M} s_1^T & 0 & 0 & \cdots & 0 & (1-\alpha_M)T_M \end{pmatrix} \quad (\text{III.12})$$

where T_i is the $N_i \times N_i$ transition probability matrix for the i^{th} subchain, s_i is the stationary distribution for the i^{th} subchain, $\mathbf{1}_{N_i}$ is a vector of all ones and of length N_i , and $0 < \alpha_i < 1$ for all i . As before, the total number of states in T shall be designated by N : $N = \sum_i N_i$.

Recall that for a matrix T to be stochastic, every element must be nonnegative and every row must sum to unity. Since $0 < \alpha_i < 1$, both α_i and $(1 - \alpha_i)$ are nonnegative for all i , the matrices $\{(1 - \alpha_i)T_i\}$ and $\{\alpha_i \mathbf{1} s_i^T\}$ consist of nonnegative elements. Hence, every element of T is nonnegative. To illustrate that every row of T sums to unity, consider the sum of the elements in the first row of T :

$$\begin{aligned} \sum_i [T]_{1,i} &= \sum_{i=1}^{N_1} (1 - \alpha_1) [T_1]_{1,i} + \sum_{i=1}^{N_1} \alpha_1 [\mathbf{1}_{N_1} s_2]_{1,i} \\ &= (1 - \alpha_1) \sum_{i=1}^{N_1} [T_1]_{1,i} + \alpha_1 \sum_{i=1}^{N_1} [s_2]_i \\ &= (1 - \alpha_1) + \alpha_1 \\ &= 1. \end{aligned}$$

Similarly, the other rows of T also sum to unity. Hence, T is a stochastic matrix.

For our purposes there are a few restrictions which must be imposed on the subchains. First, each of the T_i matrices must be diagonalizable. In Chapter II it was shown that regular Markov chains have desirable properties. The restrictions on the T_i matrices, however, can be relaxed: it is possible to obtain useful results even if some of these matrices are not regular. However, if a given T_i is not regular, and is in fact reducible, its stationary distribution s_i will not be unique. In this case, the stationary distribution s_i must be chosen such that every element of s_i is nonzero. This restriction is necessary to insure that T will be irreducible.

Before discussing the statistical properties of this structure, note that an MCRP of this type is a doubly stochastic process: each subchain is a random process, and, at a higher level, these subchains are connected together by a random process. In fact, this process can be interpreted as a Hidden Markov Model (HMM). Moreover, the process is doubly Markov, as both layers of processes are Markov.

To define the statistical properties of this structure, some notation needs to be developed. Let $\mathbf{a}_i = (a_{i,1} \ a_{i,2} \ \cdots \ a_{i,N_i})^T$ be the output map for the i^{th} subchain, so that $\mathbf{a} = (\mathbf{a}_1^T \ \mathbf{a}_2^T \ \cdots \ \mathbf{a}_M^T)^T$ for the entire chain. Also, let $\{\lambda_{i,k}\}$ for $0 \leq k \leq N_i - 1$ be the eigenvalues of T_i , with associated right and left eigenvectors $\{\mathbf{v}_{i,k}\}$ and $\{\mathbf{u}_{i,k}\}$, respectively. By convention, $\lambda_0 = 1$ is the eigenvalue associated with the right eigenvector $\mathbf{u}_{i,0}$ which is also the stationary distribution s_i . Also, let S_i be a diagonal matrix of the stationary distribution s_i for the i th subchain. Assume without loss of generality that the mean for each subchain is zero so that $\tilde{\mathbf{a}}_i = \mathbf{a}_i$. In this case, the spectral coefficients for each subchain are given by $\beta_{i,k} = \tilde{\mathbf{a}}_i^T S_i \mathbf{v}_{i,k} \mathbf{u}_{i,k}^T \tilde{\mathbf{a}}$ for $0 < k < N_i$, and for $1 \leq i \leq M$.

Using this notation, the stationary state probabilities for this structure can be expressed in block vector notation as

$$\mathbf{s} = \frac{1}{\sum_j \frac{1}{\alpha_j}} \left(\frac{1}{\alpha_1} s_1^T \quad \frac{1}{\alpha_2} s_2^T \quad \cdots \quad \frac{1}{\alpha_M} s_M^T \right)^T. \quad (\text{III.13})$$

where the $\sum_j (1/\alpha_j)$ term has been included as a scaling factor so that the elements of \mathbf{s} will sum to unity. The only eigenvalues of T which can have non-zero spectral coefficients are $\{\lambda'_{i,k}\}$ where

$$\lambda'_{i,k} = (1 - \alpha_i) \lambda_{i,k} \quad (\text{III.14})$$

for $0 < k < N_i$ and $1 \leq i \leq M$. The associated spectral coefficients for these eigenvalues are $\{\beta'_{i,k}\}$ where

$$\beta'_{i,k} = \frac{\beta_{i,k}}{\alpha_i \sum_j \frac{1}{\alpha_j}} \quad (\text{III.15})$$

for $0 < k < N_i$ and $1 \leq i \leq M$. Thus, for this structure, the spectral parameters can be found very easily from the eigenvalues and spectral parameters of the subchains. It

should be pointed out that T does have eigenvalues other than those described above; however, the spectral coefficients of these eigenvalues are identically zero. The proof of the assertions in Equations (III.13), (III.14), and (III.15) are given in the appendix.

As an example of this structure, suppose there are 3 subchains. Let

$$T_1 = \begin{pmatrix} .4 & .6 & 0 \\ 0 & .4 & .6 \\ .6 & 0 & .4 \end{pmatrix},$$

$$T_2 = \begin{pmatrix} .2 & .8 \\ .8 & .2 \end{pmatrix},$$

and $T_3 = \begin{pmatrix} .7 & .3 \\ .3 & .7 \end{pmatrix}.$

As all these matrices are all circulant matrices, $s_1 = (1/3 \ 1/3 \ 1/3)^T$, $s_2 = (1/2 \ 1/2)^T$, and $s_3 = (1/2 \ 1/2)^T$. Now suppose that $\alpha_1 = .1$, $\alpha_2 = .3$, and $\alpha_3 = .6$. The transition probability matrix for this example is then

$$T = \begin{pmatrix} (1 - \alpha_1)T_1 & \alpha_1 \mathbf{1}_3 s_2^T & 0 \\ 0 & (1 - \alpha_2)T_2 & \alpha_2 \mathbf{1}_2 s_3^T \\ \alpha_3 \mathbf{1}_2 s_1^T & 0 & (1 - \alpha_3)T_3 \end{pmatrix}$$

$$= \begin{pmatrix} .36 & .54 & 0 & .05 & .05 & 0 & 0 \\ 0 & .36 & .54 & .05 & .05 & 0 & 0 \\ .54 & 0 & .36 & .05 & .05 & 0 & 0 \\ 0 & 0 & 0 & .14 & .56 & .15 & .15 \\ 0 & 0 & 0 & .56 & .14 & .15 & .15 \\ .2 & .2 & .2 & 0 & 0 & .28 & .12 \\ .2 & .2 & .2 & 0 & 0 & .12 & .28 \end{pmatrix}.$$

Now suppose the output maps are $a_1 = (-2 \ .5 \ 1.5)^T$, $a_2 = (-.7 \ .7)^T$, and $a_3 = (-.9 \ .9)^T$, so that $a = (-2 \ .5 \ 1.5 \ -.7 \ .7 \ -.9 \ .9)^T$.

Then, according to Equation (III.13), the stationary state probabilities for T are given by

$$\begin{aligned}
\mathbf{s} &= \frac{1}{\sum_j \frac{1}{\alpha_j}} \left(\frac{1}{\alpha_1} s_1^T \quad \frac{1}{\alpha_2} s_2^T \quad \frac{1}{\alpha_3} s_3^T \right)^T \\
&= \frac{1}{\sum_j \frac{1}{\alpha_j}} \left(\frac{1}{3\alpha_1} \quad \frac{1}{3\alpha_1} \quad \frac{1}{3\alpha_1} \quad \frac{1}{2\alpha_2} \quad \frac{1}{2\alpha_2} \quad \frac{1}{2\alpha_3} \quad \frac{1}{2\alpha_3} \right)^T \\
&= \left(\frac{4}{18} \quad \frac{4}{18} \quad \frac{4}{18} \quad \frac{2}{18} \quad \frac{2}{18} \quad \frac{1}{18} \quad \frac{1}{18} \right)^T
\end{aligned}$$

since $\alpha_1 = .1$, $\alpha_2 = .3$, and $\alpha_3 = .6$. That is, the stationary probabilities are 4/18 for every state in the first subchain, 2/18 for every state in the second subchain, and 1/18 for states in the third subchain. Now consider the eigenvalues of T . For T_1 the relevant eigenvalues are

$$\lambda_{1,1} = .4(1) + .6(e^{-j2\pi/3}) + 0(e^{-j4\pi/3}) = .5292e^{-1.3807j},$$

$$\text{and } \lambda_{1,2} = .4(1) + .6(e^{-j4\pi/3}) + 0(e^{-j8\pi/3}) = .5292e^{1.3807j},$$

and the spectral coefficients for this subchain are

$$\beta_{1,1} = |A_1(1)|^2 / (3)^2 = .333,$$

$$\text{and } \beta_{1,2} = |A_1(2)|^2 / (3)^2 = .333,$$

$$\text{where } A_1(k) = -2(1) + .5(e^{-j2\pi k/3}) + 1.5(e^{-j4\pi k/3}).$$

For (T_2, \mathbf{a}_2) the relevant eigenvalue and spectral coefficient are

$$\lambda_{2,1} = .2(1) + .8(e^{-j2\pi/2}) = -.6,$$

$$\text{and } \beta_{2,1} = |A_2(1)|^2 / 2^2 = .49,$$

$$\text{where } A_2(1) = -.7(1) + .7(e^{-j2\pi/2}) = -1.4.$$

For (T_3, \mathbf{a}_3) the relevant eigenvalue and spectral coefficient are

$$\lambda_{2,1} = .7(1) + .3(e^{-j2\pi/2}) = .4,$$

$$\text{and } \beta_{2,1} = |A_2(1)|^2 / 2^2 = .81,$$

$$\text{where } A_2(1) = -.9(1) + .9(e^{-j2\pi/2}) = -1.8$$

It should be noted that T_1 , T_2 , and T_3 each have an eigenvalue of 1, but each of the spectral coefficients for these eigenvalues is zero. Then, according to Equations (III.14) and (III.15), the eigenvalues of T which have nonzero spectral coefficients and their corresponding spectral coefficients are given below:

$$\begin{aligned} \lambda'_{1,1} &= (1 - \alpha_1)\lambda_{1,1} = .4763e^{-1.3807j} & \beta'_{1,1} &= \frac{\beta_{1,1}}{\alpha_1 \sum_j \frac{1}{\alpha_j}} = .1667 \\ \lambda'_{1,2} &= (1 - \alpha_1)\lambda_{1,2} = .4763e^{-1.3807j} & \beta'_{1,2} &= \frac{\beta_{1,2}}{\alpha_1 \sum_j \frac{1}{\alpha_j}} = .1667 \\ \lambda'_{1,2} &= (1 - \alpha_2)\lambda_{2,1} = -.42 & \beta'_{2,1} &= \frac{\beta_{2,1}}{\alpha_2 \sum_j \frac{1}{\alpha_j}} = .1633 \\ \lambda'_{3,1} &= (1 - \alpha_2)\lambda_{3,1} = .16 & \beta'_{3,1} &= \frac{\beta_{3,1}}{\alpha_3 \sum_j \frac{1}{\alpha_j}} = .135 \end{aligned}$$

The power spectrum for this MCRP is given below in Figure 3.7.

The conjecture of an earlier section of this chapter can now be formally stated and proved:

Theorem: Let $\{\lambda_k\}_{k=0}^{N-1}$ be a set of N complex numbers where $\lambda_0 = 1$ and such that $|\lambda_k| < 1$ for $k \neq 0$. Let $\epsilon > 0$ be given. Then there exists a regular, diagonalizable, stochastic $N' \times N'$ matrix T with eigenvalues $\{\lambda'_k\}_{k=0}^{N'-1}$ such that

$$(1) N' \geq N,$$

$$(2) \lambda'_0 = \lambda_0 \text{ and } |\lambda'_k - \lambda_k| < \epsilon \text{ for } 0 < k < N,$$

$$\text{and } (3) |\lambda'_k| < 1 \text{ for } 0 < k < N'.$$

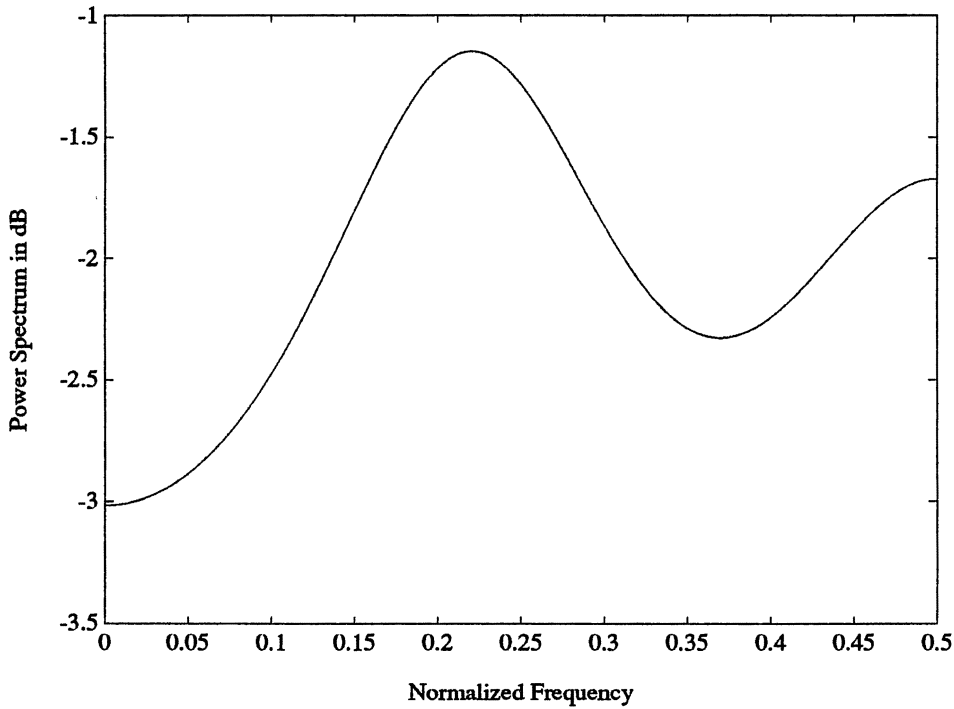


Figure 3.7. Power Spectrum for Example Ring-Structured MCRP.

To prove this theorem, MCRPs employing the ring-structure shall be used. The strategy used shall be to assign one subchain to each eigenvalue or complex-conjugate eigenvalue pair which is desired. Let us examine three separate cases: (1) complex-conjugate eigenvalue pair, (2) real, negative eigenvalue, and (3) real, positive eigenvalue.

If a complex-conjugate pair of eigenvalues is desired, let T_i be the circulant stochastic matrix where $c_{i,1} = 1$. For convenience, such a circulant matrix of size $N_i \times N_i$, shall be referred to as R_{N_i} . For example, if $N_i = 3$, this matrix is

$$R_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

The eigenvalues of this matrix are $\lambda_{i,k} = \sum_n c_{i,n} e^{-j2\pi kn/N_i} = e^{-j2\pi k/N_i}$, for $0 \leq k < N_i$. Consequently, $(1 - \alpha_i) e^{-j2\pi k/N_i}$ will be eigenvalues of T for $0 < k < N_i$. It is important to emphasize that while the MCRP with transition probability matrix R_{N_i} is not regular as it has more than one eigenvalue on the unit circle, the resulting T matrix will be regular since the eigenvalues $(1 - \alpha_i) e^{-j2\pi k/N_i}$ are within the unit circle, as $0 < \alpha_i < 1$.

Suppose the desired complex-conjugate eigenvalue pair is $r e^{\pm j\theta}$ where $0 < \theta < \pi$.

Then let $\alpha_i = 1 - r$ so that $(1 - \alpha_i) = r$, and suppose there exists positive integers k' and N_i such that $k' < N_i$ and $2\pi k'/N_i \approx \theta$. It follows that if $T_i = R_{N_i}$, then $\lambda'_{i,k'} = (1 - \alpha_i) e^{-j2\pi k'/N_i}$ and $\lambda'_{i,N_i-k'} = (1 - \alpha_i) e^{j2\pi k'/N_i}$ are eigenvalues of T . But $2\pi k'/N_i \approx \theta$ and $(1 - \alpha_i) = r$, so $\lambda'_{i,k'} \approx r e^{-j\theta}$ and $\lambda'_{i,N_i-k'} \approx r e^{j\theta}$.

In practice, the difference between θ and $2\pi k'/N_i$ can be made as small as desired by choosing k' and N_i appropriately. In general, for a given N_i , the best k' is given by

$$k' = \left[\frac{N_i \theta}{2\pi} \right] \quad (\text{III.16})$$

where $[\cdot]$ is the nearest integer operator. Hence, using this approach and given any $\varepsilon > 0$, a proper choice of α_i , k' and N_i will make the inequality $|r e^{j\theta} - (1 - \alpha_i) e^{j2\pi k'/N_i}| < \varepsilon$ hold.

The second case is when the desired eigenvalue is real and negative. In this situation, let $T_i = R_2$. The eigenvalues of T_i are $\lambda_{i,0} = 1$ and $\lambda_{i,1} = -1$. Hence, $(1 - \alpha_i)(-1)$ will be an eigenvalue of T . If the desired eigenvalue is $-r$, then let $\alpha_i = 1 - r$ so that $-r$ will be an eigenvalue of T . In this case, an approximation to the desired eigenvalue does not have to be made; the desired and resulting eigenvalue are equivalent.

The final case is when the desired eigenvalue is real and positive. In this case, let $T_i = I_2$; that is, T_i is a 2×2 identity matrix. In a situation like this in which T_i is reducible, s_i is not unique. As noted previously, s_i must be chosen so that all of its elements are nonzero. Although one could choose s_i to be most anything, let us use $s_i = (0.5 \quad 0.5)^T$.

The eigenvalues of T_i are $\lambda_{i,0} = 1$ and $\lambda_{i,1} = 1$. Hence, according to Equation (III.14) $\lambda'_{i,1} = (1 - \alpha_i)\lambda_{i,1} = (1 - \alpha_i)$ is an eigenvalue of T . If the desired real, positive eigenvalue is r , then one could choose $\alpha_i = 1 - r$ so that $(1 - \alpha_i) = r$ would be an eigenvalue of T . In this case, the desired and resulting eigenvalue are identical.

In summary, by using the ring-structured MCRP, any set of desired eigenvalues can be achieved within a given tolerance. This can be accomplished by assigning one subchain to each desired eigenvalue or complex-conjugate eigenvalue pair. For complex-conjugate eigenvalue pairs, use T_i be the circulant stochastic matrix R_{N_i} where $N_i > 2$. By properly choosing N_i and α_i for this subchain, the desired complex-conjugate pair can be approximated to within a given tolerance. For real, negative eigenvalues, use $T_i = R_2$ and choose α_i to be one minus the magnitude of the eigenvalue. In this case, the desired eigenvalue is matched exactly. Finally, if the desired eigenvalue is real and positive, use $T_i = I_2$ and choose α_i to be one minus the magnitude of the eigenvalue. In this case, the desired eigenvalue is also matched exactly. Of course, there are more exotic procedures which will also achieve the desired results.

This constructive proof solves one of the two problems in power spectrum synthesis: locating T which possesses a given set of eigenvalues. The second problem is to determine the necessary output map to achieve the desired spectral coefficients $\{\beta_k\}$. Following the procedure above, let us split this problem into three cases: (1) complex-conjugate eigenvalue pair; (2) real, negative eigenvalue; and (3) real, positive eigenvalue.

Suppose that $re^{\pm j\theta}$ is a desired complex-conjugate eigenvalue pair. As described above, one could assign T_i to be the circulant matrix R_{N_i} . Let us assume that α_i , k' , and N_i have been chosen so that $(1 - \alpha_i)e^{\pm j2\pi k'/N_i}$ approximates the desired eigenvalue pair $re^{\pm j\theta}$.

The eigenvalues of T which are derived from T_i are $\lambda'_{i,k} = (1 - \alpha_i)e^{-j2\pi k/N_i}$ for $0 < k < N_i$. However, only $\lambda'_{i,k'} = (1 - \alpha_i)e^{-j2\pi k'/N_i}$ and $\lambda'_{i,N_i-k'} = (1 - \alpha_i)e^{j2\pi k'/N_i}$ are desired. Consequently, one would like to make $\beta_{i,k} = 0$ for k not equal to k' or $N_i - k'$. This can be accomplished by appropriately selecting the output map \mathbf{a} , for this subchain. Recall that $\beta_{i,k} = |A_i(k)|^2 / N_i^2$. Thus, to make $\beta_{i,k} = 0$ for $k \neq k'$, one need only make $A(k) = 0$ for $k \neq k'$. The solution for this is to let $a_{i,n} = g_i \cos(2\pi k'n/N_i)$ where g_i is a constant which can be adjusted to make $\beta_{i,k'}$ as large or small as desired. If this is done, then

$$\beta_{i,k} = \begin{cases} \frac{g_i^2}{4} & k = k', N_i - k' \\ 0 & \text{else} \end{cases}$$

Consequently, the only eigenvalues of T which are derived from T_i and which have non-zero spectral coefficients are $\lambda'_{i,k'} \approx r e^{-j\theta}$ and $\lambda'_{i,N_i-k'} \approx r e^{j\theta}$. The spectral coefficients for these eigenvalues will be

$$\begin{aligned} \beta'_{i,k'} &= \frac{\beta_{i,k'}}{\alpha_i \sum_j \frac{1}{\alpha_j}} \\ &= \frac{g_i^2/4}{\alpha_i \sum_j \frac{1}{\alpha_j}} \end{aligned} \quad (\text{III.17a})$$

$$\begin{aligned} \text{and } \beta'_{i,N_i-k'} &= \frac{\beta_{i,N_i-k'}}{\alpha_i \sum_j \frac{1}{\alpha_j}} \\ &= \frac{g_i^2/4}{\alpha_i \sum_j \frac{1}{\alpha_j}} \end{aligned} \quad (\text{III.17b})$$

The constant g_i can be adjusted to achieve the desired magnitude.

Next suppose that $(-r)$ is a desired eigenvalue. As discussed above, for this eigenvalue one would choose $T_i = R_2$. The eigenvalues of R_2 are $\lambda_{i,0} = 1$ and $\lambda_{i,1} = -1$. By

selecting $\alpha_i = 1 - r$, T will have eigenvalue $\lambda'_{i,1} = (1 - \alpha_i)\lambda_{i,1} = -r$. Let us choose $a_i = (g \quad -g)^T$. In this case, it can be shown that $\beta_{i,1} = g_i^2$. Thus, T has an eigenvalue $\lambda'_{i,1} = -r$ with spectral coefficient

$$\begin{aligned}\beta'_{i,1} &= \frac{\beta_{i,1}}{\alpha_i \sum_j \frac{1}{\alpha_j}} \\ &= \frac{g_i^2}{\alpha_i \sum_j \frac{1}{\alpha_j}}.\end{aligned}\tag{III.18}$$

The constant g_i can be adjusted to achieve the desired magnitude for the spectral coefficient.

Finally suppose that r is a desired eigenvalue. As discussed above, for this eigenvalue one would choose $T_i = I_2$. The eigenvalues of R_2 are $\lambda_{i,0} = 1$ and $\lambda_{i,1} = 1$. By selecting $\alpha_i = 1 - r$, T will have eigenvalue $\lambda'_{i,1} = (1 - \alpha_i)\lambda_{i,1} = r$. Let us choose $a_i = (g \quad -g)^T$. In this case, it can be shown that $\beta_{i,1} = g_i^2$. Thus, T has an eigenvalue $\lambda'_{i,1} = r$ with spectral coefficient

$$\begin{aligned}\beta'_{i,1} &= \frac{\beta_{i,1}}{\alpha_i \sum_j \frac{1}{\alpha_j}} \\ &= \frac{g_i^2}{\alpha_i \sum_j \frac{1}{\alpha_j}}.\end{aligned}\tag{III.19}$$

Again, the constant g_i can be adjusted to achieve the desired magnitude for the spectral coefficient.

Using these techniques to define the subchains, let us work a simple example. Suppose the desired eigenvalues are $0.9e^{\pm 0.65\pi j}$, -0.8 , and 0.7 . Hence, three subchains are needed: one to produce eigenvalues $0.9e^{\pm 0.65\pi j}$; one to produce the eigenvalue -0.8 ; and one to produce the eigenvalue 0.7 .

First consider the complex conjugate eigenvalues $0.9e^{\pm 0.65\pi j}$. As detailed above, let $T_1 = R_{N_1}$ and $\alpha_1 = 1 - (0.9) = 0.1$. Next, a k' and N_1 need to be found such that $2\pi k'/N_1 \approx 0.65\pi$. As a simple approximation, let $k' = 1$ and $N_1 = 3$. Then $0.9e^{\pm 2j\pi/3}$ will be eigenvalues for the ring-structured MCRP. The output map $\mathbf{a}_1 = (a_{1,1} \ a_{1,2} \ a_{1,3})^T$ is defined by $a_{1,n} = g_1 \cos(2\pi k'n/N_1)$. Substituting $k' = 1$ and $N_1 = 3$, $a_{1,n} = g_1 \cos(2\pi n/3)$. The constant g_1 is a function of α_i for the other subchains, so the determination of g_1 will have to wait until all of the α_i terms have been defined.

Next consider the eigenvalue (-0.8). As detailed above, let $T_2 = R_2$ and $\alpha_2 = 1 - 0.8 = 0.2$. Then (-0.8) will be an eigenvalue for the ring-structured MCRP. The output map for this chain is $\mathbf{a}_2 = (g_2 \ -g_2)^T$. As above, determining g_2 will have to wait until the remaining α_i term is defined.

Finally, consider the eigenvalue 0.7. As detailed above, let $T_3 = I_2$ and $\alpha_3 = 1 - 0.7 = 0.3$. Then 0.7 will be an eigenvalue for the ring-structured MCRP. The output map for this chain is $\mathbf{a}_3 = (g_3 \ -g_3)^T$.

Now, according to Equations (III.17a) and (III.17b), the spectral coefficients for the complex-conjugate eigenvalues $0.9e^{\pm 2j\pi/3}$ are both given by

$$\frac{g_1^2/4}{\alpha_1 \sum_j \frac{1}{\alpha_j}}$$

As we want this constant to be equal to 2, as given above, $g_1 = 3.83$. Hence,

$$\mathbf{a}_1 = (-1.915 \ -1.915 \ 3.83)^T$$

From Equation (III.18), the spectral coefficient for the eigenvalue (-0.8) is given by

$$\frac{g_2^2}{\alpha_2 \sum_j \frac{1}{\alpha_j}}$$

As we wish this to be equal to 1, as given above, $g_2 = 1.915$. Hence,

$$a_2 = (1.915 \quad -1.915)^T$$

From Equation (III.19), the spectral coefficient for the eigenvalue 0.7 is given by

$$\frac{g_3^2}{\alpha_3 \sum_j \frac{1}{\alpha_j}}$$

As we wish this to be equal to 1.5, as given above, $g_2 = 2.345$. Hence,

$$a_3 = (2.345 \quad -2.345)^T$$

The transition probability matrix for the ring-structured MCRP is then given by

$$T = \begin{pmatrix} (1 - \alpha_1)T_1 & \alpha_1 \underline{1}_3 s_2^T & 0 \\ 0 & (1 - \alpha_2)T_2 & \alpha_2 \underline{1}_2 s_3^T \\ \alpha_3 \underline{1}_2 s_2^T & 0 & (1 - \alpha_3)T_3 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & .9 & 0 & .05 & .05 & 0 & 0 \\ 0 & 0 & .9 & .05 & .05 & 0 & 0 \\ .9 & 0 & 0 & .05 & .05 & 0 & 0 \\ 0 & 0 & 0 & 0 & .8 & .1 & .1 \\ 0 & 0 & 0 & .8 & 0 & .1 & .1 \\ .1 & .1 & .1 & 0 & 0 & .7 & 0 \\ .1 & .1 & .1 & 0 & 0 & 0 & .7 \end{pmatrix}$$

and the output map by

$$a = (a_1^T \quad a_2^T \quad a_3^T)^T$$

$$= (-0.7385 \quad 0.7385 \quad 1.477 \quad 1.915 \quad -1.915 \quad 2.345 \quad -2.345)^T$$

Note that $s_1 = (1/3 \quad 1/3 \quad 1/3)^T$ and that $s_2 = s_3 = (1/2 \quad 1/2)^T$. The power spectrum for this example is given in Figure 3.8 below.

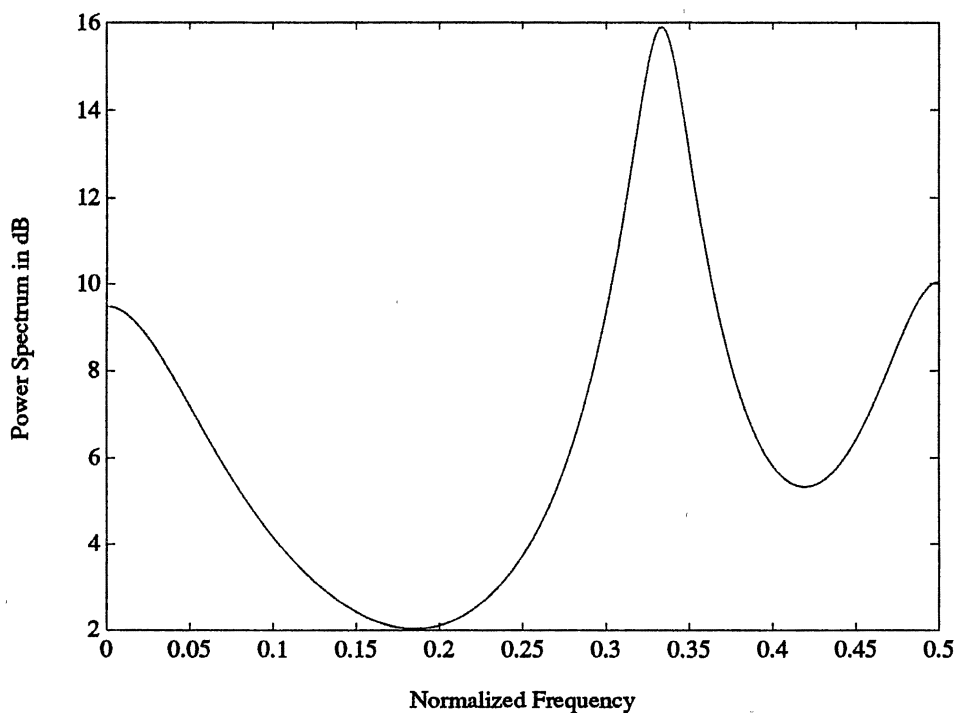


Figure 3.8. Power Spectrum of Synthesis Example.

Joint Probability Distribution - Power Spectral Density Synthesis

Now consider the more difficult problem of locating a random process which matches both a given probability distribution and power spectrum. Previous work done in this area is exemplified by the work of Sondhi [Son83] and more recently by Taori, for example [Tao90]. The traditional approach to this problem has been to pass a Gaussian white noise process through a linear filter and then through a memoryless nonlinearity. The nonlinearity achieves the desired distribution. Together, the linear filter and nonlinearity determine the spectral shaping of the resulting process. The crux of this problem is

to design the linear filter to produce the proper spectral shaping, taking into account the warping effects caused by the nonlinearity. In his paper, Sondhi reported successful results for many different combinations of spectra and distribution.

However, let us consider MCRPs for joint probability distribution/power spectral density synthesis. Having considered probability distribution synthesis and power spectral density synthesis, one can see that the solution is the intersection of the two solution sets. In particular, the solution set for this task is likely much smaller than either of the two sets. However, given the material from the previous sections, one can at least explore a few possibilities.

In particular, let us restrict our attention to the case when the desired probability distribution function is from a continuous-valued process. Recall from the first section of this chapter concerning probability distribution synthesis that T was chosen to be any doubly stochastic matrix and let \mathbf{a} was defined from centroids of the partitioned distribution function. If we adopt this as the basis for our joint synthesis problem, consider the restrictions.

As \mathbf{a} is fixed, the only item which can be modified to meet the spectral constraint is the transition probability matrix. However, T must be doubly stochastic. Of the structures considered in the previous section, only the circulant structure qualifies. Let us briefly review the results for circulant MCRPs. Recall that a circulant transition probability matrix T_C is defined by $[T_C]_{i,j} = c_{(j-i) \bmod N}$. For T_C , the eigenvalues are given by $\lambda_k = \sum_{i=0}^{N-1} c_i e^{-j2\pi i k/N}$. The spectral coefficients $\{\beta_k\}$ are found by taking the DFT of the output map: $\beta_k = |A(k)|^2 / N^2$ where $A(k) = \sum_{i=0}^{N-1} a_{i+1} e^{-j2\pi i k/N}$, and $\mathbf{a} = (a_1 \ a_2 \ \dots \ a_N)^T$. In our case, since \mathbf{a} is fixed, this means that $\{\beta_k\}$ are also fixed. Thus, only the eigenvalues can be modified.

We note here that we have a major problem as the spectral coefficients are fixed. Previously, we had asserted that we could add more eigenvalues than were strictly needed if the weights of these eigenvalues were identically zero. However, that approach will not work in this instance as the spectral coefficients will, in general, be other than zero. However, there is another way around this problem. Suppose that we can make $\lambda_k = 0$ for any spectral coefficient that is unwanted. As evidenced by the equation

$$\begin{aligned}\Phi(\omega) &= \sum_{k=1}^{N-1} \beta_k \left(\frac{1}{1 - \lambda_k e^{j\omega}} + \frac{1}{1 - \lambda_k e^{-j\omega}} - 1 \right) \\ &= \sum_{\lambda_k \neq 0} \beta_k \left(\frac{1}{1 - \lambda_k e^{j\omega}} + \frac{1}{1 - \lambda_k e^{-j\omega}} - 1 \right) + \sum_{\lambda_k = 0} \beta_k\end{aligned}$$

the result will be a bias in the spectral power equal to the sum of the spectral coefficients of all eigenvalues equal to zero. Hence, at least these terms will not influence the curvature of the power spectrum.

As our purpose here is only to perform a brief investigation, let us consider a simple example. Suppose that the process is to have a peak in the power spectrum at some specified radian frequency θ . As T_C is defined in terms of $\{c_i\}$, suppose

$$c_i = \frac{r}{N} \cos\left(\frac{2\pi i k'}{N} - \theta\right)$$

Then, the eigenvalues of T_C would be

$$\begin{aligned}\lambda_k &= \sum_{i=0}^{N-1} c_i e^{-j2\pi k i / N} \\ &= \begin{cases} r e^{-j\theta} & \text{for } k = k' \\ r e^{j\theta} & \text{for } k = N - k' \\ 0 & \text{else} \end{cases}\end{aligned}$$

This meets the criteria above wherein all eigenvalues are zero except for the desired eigenvalue. However, there is one major problem: a transition probability matrix based on $\{c_i\}$ defined above would not be stochastic as $\sum_i c_i = 0$ and $\min(c_i) < 0$. Hence, consider redefining $\{c_i\}$ as

$$c_i = \frac{1}{N} + \frac{2r}{N} \cos\left(\frac{2\pi i k'}{N} - \theta\right)$$

In this case, T_C based on $\{c_i\}$ is stochastic for $|r| \leq .5$, and the eigenvalues are

$$\begin{aligned} \lambda_k &= \sum_{i=0}^{N-1} c_i e^{-j2\pi k i / N} \\ &= \begin{cases} 1 & \text{for } k = 0 \\ r e^{-j\theta} & \text{for } k = k' \\ r e^{j\theta} & \text{for } k = N - k' \\ 0 & \text{else} \end{cases} \end{aligned}$$

The only difference from the above is that the magnitude of the eigenvalue with angle θ is restricted to being less than .5 in magnitude. Also note that, in practice, k' would most likely be chosen to correspond to the spectral coefficient with the largest magnitude.

As a specific example suppose the process is to have Laplacian density with unit variance, $\theta = \pi/4$ and N is chosen to be 20. Then the output map is as defined in the first section of this chapter. Based upon this output map, the largest β_k is $\beta_1 = 0.1438$. Hence, let $k' = 1$ so that

$$c_i = \frac{1}{20} + \frac{1}{20} \cos\left(\frac{2\pi i}{20} - \frac{\pi}{4}\right).$$

The power spectrum for this process is shown below in Figure 3.9.

In general, we would like to have control over the magnitude of the eigenvalue. It may be possible to do this if the output map can be chosen more flexibly. For, if some of

the weights can be made small, then undesired, non-zero eigenvalues could be tolerated. However, it is not immediately clear how one would modify the output map to achieve this result without compromising the probability distribution constraints.

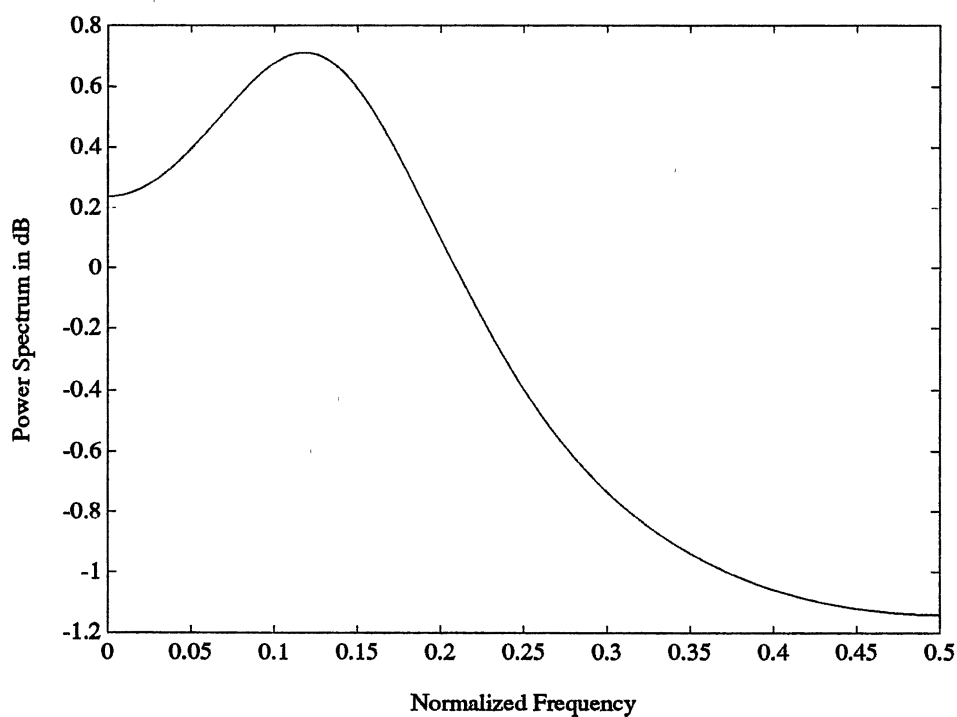


Figure 3.9. MCRP Power Spectrum from Joint Synthesis.

CHAPTER IV
STOCHASTIC MODELING WITH MARKOV
CHAIN RANDOM PROCESSES

In the previous chapter, a ring structure was introduced which allowed a MCRP to be found with a given set of eigenvalues. In this chapter, the flexibility of MCRPs, and of ring-structured MCRPs in particular, is tested by considering MCRPs as models for stochastic processes. Of course, the best random process which a MCRP could model would be another Markov chain random process. In general, however, it is worthwhile to consider MCRPs as models for arbitrary stochastic processes. In particular, the focus of this chapter is modeling stationary processes with MCRPs. The more difficult problem of modeling nonstationary processes is only discussed briefly.

Stationary Processes

In any modeling application, perhaps the first consideration should be the modeling criterion: i.e., those properties of the process which the model should try to match. Of course, this is application-dependent. However, for a large number of applications, the 2nd order statistics are of main importance [Pap84]. Hence, the autocorrelation, or equivalently the power spectrum, provides a good choice for modeling criterion: the closer the match between the power spectrum of the random process and the power spectrum of the model, the better the model.

Before considering MCRPs as stochastic models, it is useful to review other stationary random processes which have been used as stochastic models. In the following section, a class of linear processes known as ARMA is introduced and compared with MCRPs.

ARMA Processes

A linear single-input single-output discrete-time system can be expressed by the difference equation

$$y(n) = u(n) + \sum_{j=1}^L b_j u(n-j) - \sum_{i=1}^M a_i y(n-i) \quad (\text{IV.1})$$

where $u(n)$ is the input and $y(n)$ is the output of the process at time n . Such a system is commonly referred to as an *autoregressive moving-average* (ARMA) process. Two special cases of Equation (IV.1) are worth noting: if $L = 0$, the process is called *autoregressive* (AR), whereas if $M = 0$ the process is called a *moving average* process (MA). It is often more useful to interpret AR, MA, and ARMA processes as the output of a linear filter which has the transfer function

$$\begin{aligned} H(z) &= \frac{1 + \sum_{j=1}^L b_j z^{-j}}{1 + \sum_{i=1}^M a_i z^{-i}} \\ &= \frac{\prod_{j=1}^L (1 - z_j z^{-1})}{\prod_{i=1}^M (1 - p_i z^{-1})} \end{aligned} \quad (\text{IV.2})$$

where z_j are zeros of the system and p_i are poles of the system. Then, if the input $u(n)$ is a stationary process with power spectrum $\Phi_u(\omega)$, and all of the poles $\{p_i\}$ are less than unity in magnitude, the output $y(n)$ will be stationary and its power spectrum will be given by

$$\Phi_y(\omega) = |H(e^{j\omega})|^2 \Phi_u(\omega). \quad (\text{IV.3})$$

In many applications, $u(n)$ is chosen to be white noise so that $\Phi_u(\omega)$ is a constant. In this case, all of the spectral shaping is determined by the filter. Often, it is useful to model a random process which does not have a flat spectrum as the output of a linear filter driven by white noise.

ARMA Processes and MCRPs

In the time-domain, ARMA processes and MCRPs have little in common. This is chiefly because ARMA processes have continuous probability distribution functions whereas the probability distribution function of MCRPs is discrete. Still, both can be interpreted as being the output of a transfer function. In the case of ARMA processes, the transfer function is a linear, constant-coefficient filter. In contrast, from the discussion in Chapter II, the transfer function for MCRPs is nonlinear and state-dependent.

However, surprisingly enough, ARMA processes and MCRPs have quite similar spectral characteristics. This is clearly not apparent from comparing the ARMA power spectrum in Equation (IV.3) and the MCRP power spectrum given in Chapter II, Equation (II.34):

$$\Phi(\omega) = \sum_{k=1}^{N-1} \beta_k \left(\frac{1}{1 - \lambda_k e^{j\omega}} + \frac{1}{1 - \lambda_k e^{-j\omega}} - 1 \right) \quad (\text{II.34})$$

Yet, with a little manipulation, it is possible to discover a striking similarity between the two processes. Towards this end, consider the autocorrelation of an ARMA process. Suppose that the input to the ARMA process is white noise, there are no repeated poles, and the number of poles is greater than the number of zeroes. Under these restrictions, it is possible to express the autocorrelation of an ARMA process as

$$\phi_y(n) = \sum_{i=1}^N \gamma_i p_i^{|n|}. \quad (\text{IV.4})$$

where $\{p_i\}$ are the poles of the system and $\{\gamma_i\}$ are constants which are determined by $H(z)$ [Mul72]. Specifically,

$$\gamma_i = \sigma_v^2 A_i H(p_i^{-1}) \quad (\text{IV.5a})$$

$$\text{where } A_i = (1 - p_i z^{-1}) H(z) |_{z=p_i}. \quad (\text{IV.5b})$$

Now recall the MCRP autocorrelation expression from Chapter II, Equation (II.23):

$$\phi(n) = \sum_{k=1}^{N-1} \beta_k \lambda_k^{|n|} + \mu_y^2, \quad (\text{II.23})$$

Comparing Equations (IV.4) and (II.23) as given above, it is clear that there is a similarity between the autocorrelations, and therefore the power spectra, of ARMA processes and MCRPs. In fact, one researcher has gone so far as to conjecture that "any power spectrum obtained by shaping white noise with a finite order linear filter can also be obtained with a finite Markov chain" [Mul79]. From Equations (II.23) and (IV.4) it seems apparent that by setting $\lambda_k = p_i$, $\beta_k = \gamma_i$, and $\mu_y = 0$, the power spectra of the two processes would be identical.

In summary, then, while the first order properties of MCRPs and ARMA processes are dissimilar, their second order properties are remarkably similar. These comparisons are reiterated below in Table 4.1. Before continuing, it should be noted that it is possible to massage the power spectrum equation for MCRPs into a form similar to Equation (IV.3); however, as far as the author knows, there is no easy way of doing this. The representations of the autocorrelations given above appear to be the best grounds on which to compare the 2nd order properties of ARMA processes and MCRPs.

TABLE 4.1
COMPARISON OF ARMA PROCESSES AND MCRPS

	ARMA Processes	MCRPs
<u>Transfer Function:</u>	Linear, Constant-Coefficient	Nonlinear, State-Dependent
<u>Probability Distribution Function:</u>	Continuous	Discrete
<u>Autocorrelation:</u>	$\sum_i \gamma_i p_i^{ n }$	$\sum_k \beta_k \lambda_k^{ n }$

Parameter Solutions for ARMA Processes

In a real-world scenario of data modeling, one would be given a finite-length data sequence and be required to find the parameters of a model which best fit the data. For ARMA models, parameter estimation is based on the notion of prediction. To begin, recall the ARMA input-output equation:

$$y(n) = H(z)u(n)$$

$$= \frac{B(z)}{A(z)}u(n)$$

where $B(z) = 1 + b_1z^{-1} + \dots + b_Lz^{-L}$ and $A(z) = 1 + a_1z^{-1} + \dots + a_Mz^{-M}$. For the purposes of this research, we are most interested in the case for which $M > L$; i.e., when the denominator polynomial is of higher degree than the numerator polynomial. At any time n , a prediction of $y(n)$ given $\{y(n-1), y(n-2), \dots\}$ and based on the model parameters $\{a_i\}$ and $\{b_i\}$ can be made by $\hat{y}(n) = E[y(n) | y(n-1), y(n-2), \dots]$. The prediction error, $e(n) = y(n) - \hat{y}(n)$ is given by

$$\begin{aligned} e(n) &= H^{-1}(z)y(n) \\ &= \frac{A(z)}{B(z)}y(n). \end{aligned}$$

(See, for example, [Ros79] for details.) The model parameters $\{a_i\}$ and $\{b_i\}$ are often found by minimizing the error function

$$J = \sum_n e^2(n).$$

Unfortunately, since $e(n)$ is not a linear function of the model parameters, a linear solution for this minimization does not exist. Still, iterative approaches can be used in the solution of this problem.

However, suppose the process is restricted to being AR. In this case,

$$\begin{aligned} e(n) &= A(z)y(n) \\ &= y(n) + a_1y(n-1) + \cdots + a_My(n-M). \end{aligned}$$

It is clear that $e(n)$ is linear in the model parameters $\{a_i\}$, and therefore, a linear solution can be found which minimizes J .

Given the data set $\{y(1), y(2), \dots, y(N)\}$, J is minimized by [Men87]:

$$\Theta = (\Psi_N^T \Psi_N)^{-1} \Psi_N^T Y_N \quad (\text{IV.8})$$

where

$$\Psi_N = \begin{pmatrix} y(M-1) & y(M-2) & \cdots & y(1) \\ y(M) & y(M-1) & \cdots & y(2) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ y(N-1) & y(N-2) & \cdots & y(N-M) \end{pmatrix}, \Theta = \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_M \end{pmatrix}, \text{ and } Y_N = \begin{pmatrix} y(M+1) \\ y(M+2) \\ \cdot \\ \cdot \\ y(N) \end{pmatrix}$$

Calculating the model parameters by minimizing the sum of the squares of the prediction errors does not explicitly match data and model power spectrum. However, an information-theoretic argument can be used to justify that the resulting power spectrum

from the AR model is a valid power spectrum estimator of the data [Sch78]. Thus, this method of finding the model parameters is appropriate given that the modeling criteria is the power spectrum.

Parameter Solutions for MCRPs

For MCRPs, finding the model parameters is more difficult. The probabilistic nature of the model prevents a least-squares prediction-error formulation to solve for the model parameters. However, let us consider several other approaches.

One idea is to assume that the data is generated by an MCRP. In this case, the frequency of occurrence of values in the data provide a good estimate of the transition probabilities. However, the data to be modeled may have a wide range of values. Thus, each element of the time series must initially be quantized into one of N levels. The transition probabilities can then be estimated by averaging the number of state-to-state transitions:

$$\begin{aligned} \hat{t}_{i,j} &= \hat{P}[x(n+1) = j | x(n) = i] \\ &= \frac{\text{Number of transitions from level } i \text{ to } j}{\text{Number of occurrences of level } i}. \end{aligned} \quad (\text{IV.9})$$

This method does not inherently guarantee power spectrum matching, but it is quick, easy, and intuitively clear.

In fact, the above technique has been used in several recent studies. One study was in modeling solar radiation [Agu88]. In this study, the authors used a library of MCRPs to create a model of daily solar radiation at any place on earth. Multiple MCRPs were required to account for different radiation patterns around the globe. The authors found that this approach produced excellent results, and matched the probability distribution function and autocorrelation of known solar radiation data.

Another study used this approach to model wind speed [Jon86]. In this study, the wind speeds were categorized into eleven levels, and the transition probabilities were estimated as described above. The authors reported that the autocorrelation of the model and the data did not match exactly, though they were close for up to 30 lags.

While the above approach has merit, one can do better than this by explicitly matching the data and model power spectra. To begin, one must have an estimate of the data power spectra. One of the simplest estimates is the periodogram which is given by

$$\hat{\Phi}(\omega) = |Y(e^{j\omega})|^2 \quad (\text{IV.10})$$

where $Y(e^{j\omega})$ is the discrete-time Fourier transform of the data [Opp75]. However, this is not the most useful estimate for modeling purposes as it is difficult to determine the MCRP parameters from the estimate. It would be more advantageous to have an estimate of the data power spectrum from which the MCRP spectral parameters $\{\lambda_k\}$ and $\{\beta_k\}$ could be easily determined.

Towards this end, reconsider the AR model. As noted previously, the power spectrum from the AR model is a valid estimate of the data power spectrum. From the AR model for the data, the poles $\{p_i\}$ and spectral coefficients $\{\gamma_i\}$ can easily be found. Consequently, these values can be used to determine the MCRP spectral parameters $\{\lambda_k\}$ and $\{\beta_k\}$. Therefore, it appears possible to use an AR model as the basis for the MCRP model. This approach is illustrated below in Figure 4.1.

Before continuing, one might ask what advantage there is in using a MCRP as a model in this way. Especially since it appears that the MCRP model can at best be only as good as an AR model. One answer is that MCRPs might be able to better model certain types of processes. The modeling criteria has been chosen to be the power spectrum. However, for some processes, other statistics might be important as well. It is certainly possible that, for some types of random processes, the structure of MCRPs provides a better model than AR or for that matter, ARMA. A trivial example of this, of course, is if

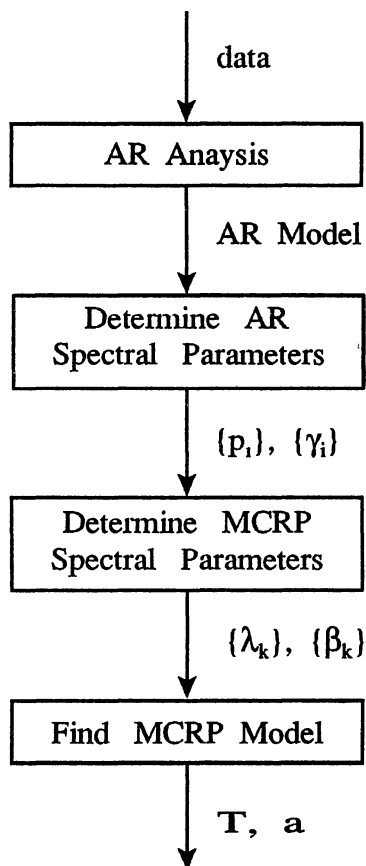


Figure 4.1. MCRP Modeling Strategy.

the data were generated from an MCRP. There may in fact be ways of finding MCRP model parameters from the data which do not rely on AR models. However using the AR model as a basis for the MCRP model seems a logical approach for now.

AR Processes and Ring-Structured MCRPs

Consider using the method illustrated above in Figure 4.1. Determining the AR model and its spectral parameters $\{p_i\}$ and $\{\gamma_i\}$ poses no problems. Given the AR spectral parameters, the MCRP parameters are immediately known also: the poles $\{p_i\}$ should

be equated to the eigenvalues $\{\lambda_k\}$ and the AR coefficients $\{\gamma_k\}$ to the MCRP coefficients $\{\beta_k\}$. One could then use a circulant or ring-structured MCRP to locate a transition probability matrix T for the MCRP which has the necessary eigenvalues $\{\lambda_k\}$. The only remaining issue is to determine the MCRP output map a so that $\{\beta_k\}$ equates to $\{\gamma_k\}$. However, herein lies the problem. For, the AR spectral coefficients γ_k are in general complex. While the MCRP pole weights β_k may also be complex, they must be real and positive for the only known MCRP structures. Consequently, there is a difference between the spectral properties of ARMA processes and known MCRP structures.

While this disparity creates problems, it does not mean that this modeling strategy must be discarded. However, this problem makes it necessary to investigate the differing spectra of MCRPs with real-valued spectral coefficients and AR processes which have complex-valued spectral coefficients. First consider the simple case of 2nd-order process which have complex pole-pairs. As an example, suppose the AR transfer function is

$$H(z) = \frac{1}{1 - .5z^{-1} + .25z^{-2}}$$

which has poles at $p_1 = .5e^{+j\pi/3}$, and $p_2 = p_1^*$. Assuming white noise with unity variance is the excitation signal, the spectral coefficients will be given by $\gamma_1 = 0.6349 - 0.2199j$ and $\gamma_2 = \gamma_1^*$. Using the ring structure described in the previous chapter, a MCRP can be found with poles $\lambda_1 = p_1$ and $\lambda_2 = p_2$ but its spectral coefficients, β_1 and β_2 , will be real and positive. Figure 4.2 shows the power spectra of the AR process and MCRP. For both curves, the spectral magnitudes have been normalized to 0 dB.

There are two main differences between the AR and MCRP curves. First, the shapes of the curves are different. The MCRP is more symmetric about its peak frequency than the AR spectra. The second and more important difference is that the power spectra do not experience their maxima at the same frequency. The AR peaks at a

normalized frequency of 0.14 whereas the MCRP peaks at 0.17. Differing peak frequencies is undesired because the frequencies near the peak contain most of the process's power. Thus, these are the frequencies which are most important to match.

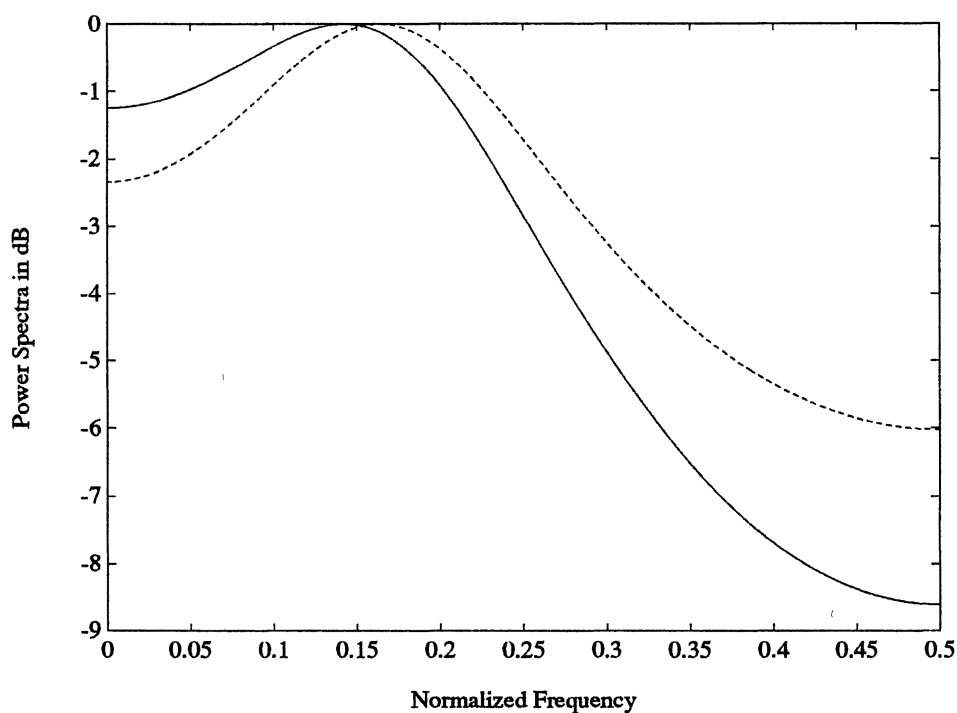


Figure 4.2. Power Spectra for AR Process (solid line) and MCRP (dashed line).

The spectra of AR processes and MCRPs with one pole-pair was calculated for many different pole-pairs. Figure 4.3 displays several AR and MCRP power spectra for various combinations of pole magnitude r and normalized pole angle f , where the pole-pairs are $re^{\pm j2\pi f}$. It was found that there is a symmetry about the imaginary axis for the

poles; that is, the plot for poles of angle $r e^{\pm j2\pi(0.5-f)}$ is symmetric to the plot for $r e^{\pm j2\pi f}$. Thus, only values of f from 0 to 0.5 are shown. Also, note that if the eigenvalues and poles are purely complex, the AR and MCRP power spectra agree exactly; in these cases, the AR spectral coefficients $\{\gamma_i\}$ are real, so exact matches are possible.

As mentioned previously, and as illustrated in Figure 4.3, the AR and MCRP processes experience maxima at different frequencies. The difference between these frequencies is plotted below in Figure 4.4. The figure plots the peak frequency of the AR process minus the peak frequency of the MCRP versus the pole/eigenvalue angle f where the AR poles and MCRP eigenvalues are $r e^{\pm j2\pi f}$. Each curve in the plot represents a constant r . From Figures 4.3 and 4.4, it is clear that the difference in peak frequency decreases as the pole magnitude increases. Also, the phase angle of the pole, f , is significant. The peak difference are less significant when the poles and eigenvalues are near the real or imaginary axis; i.e., when f is near 0, 0.25, or 0.5. In fact, as noted before, if the poles and eigenvalues are purely complex, i.e. $f=0.25$, there is no difference in maximal frequencies.

Of course, it might be possible to achieve a better spectral match by modifying the MCRP spectral parameters $\{\lambda_k\}$ and $\{\beta_k\}$. For example, for the case discussed above with one complex pole-pair, one could change the phase angle of the complex eigenvalue-pair so that peak frequencies would agree. Better still, one might want to minimize an error function like

$$J = \int_0^{\pi} [\Phi_{AR}(\omega) - \Phi_{MCRP}(\omega)]^2 d\omega$$

with respect to the MCRP spectral parameters. However, as J is not linear in the MCRP eigenvalues $\{\lambda_k\}$, a nonlinear solution would be required. Consequently, while the search for better a MCRP match is certainly worthwhile, it is also more difficult.

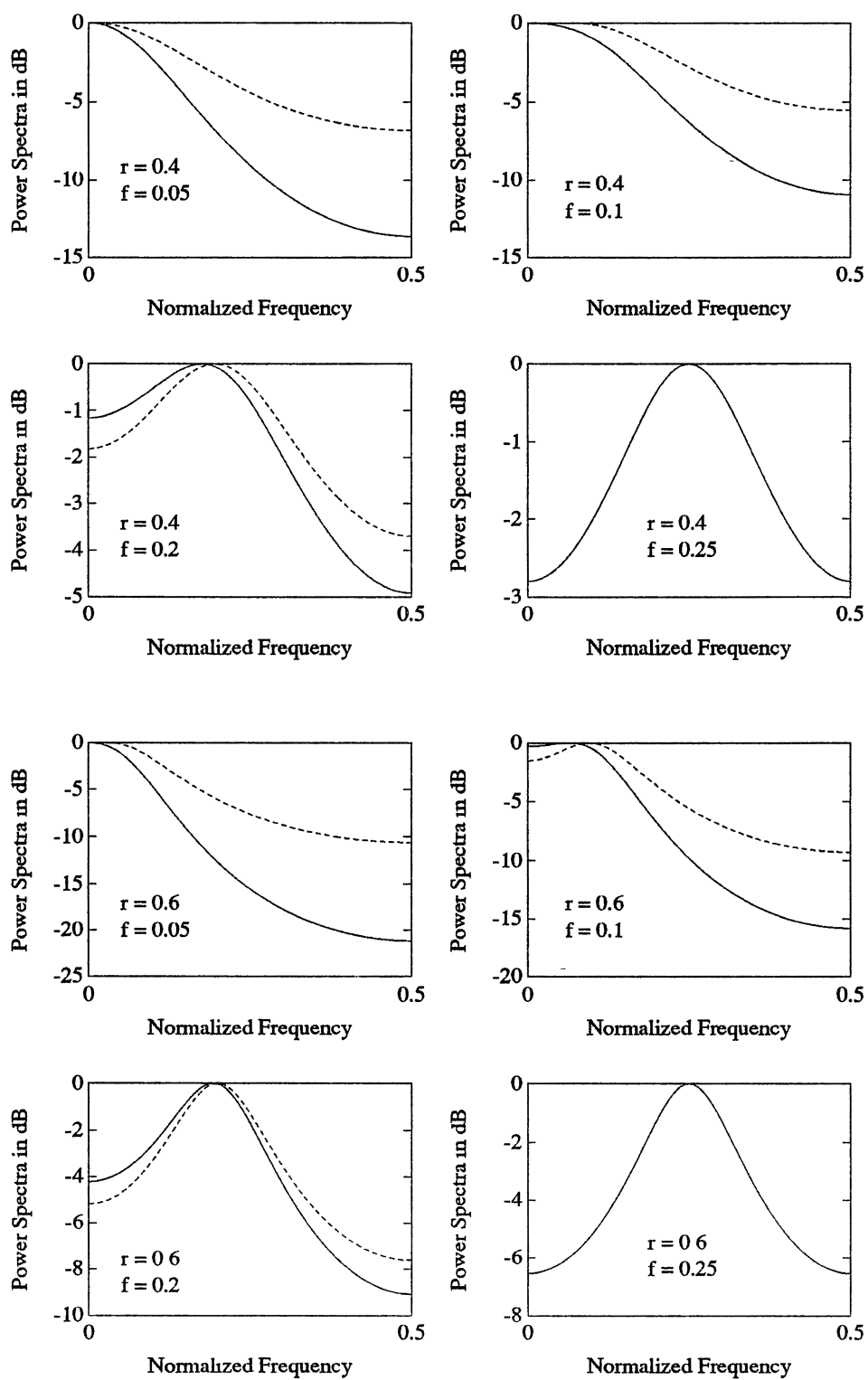


Figure 4.3. AR (solid) and MCRP (dashed) Power Spectra.

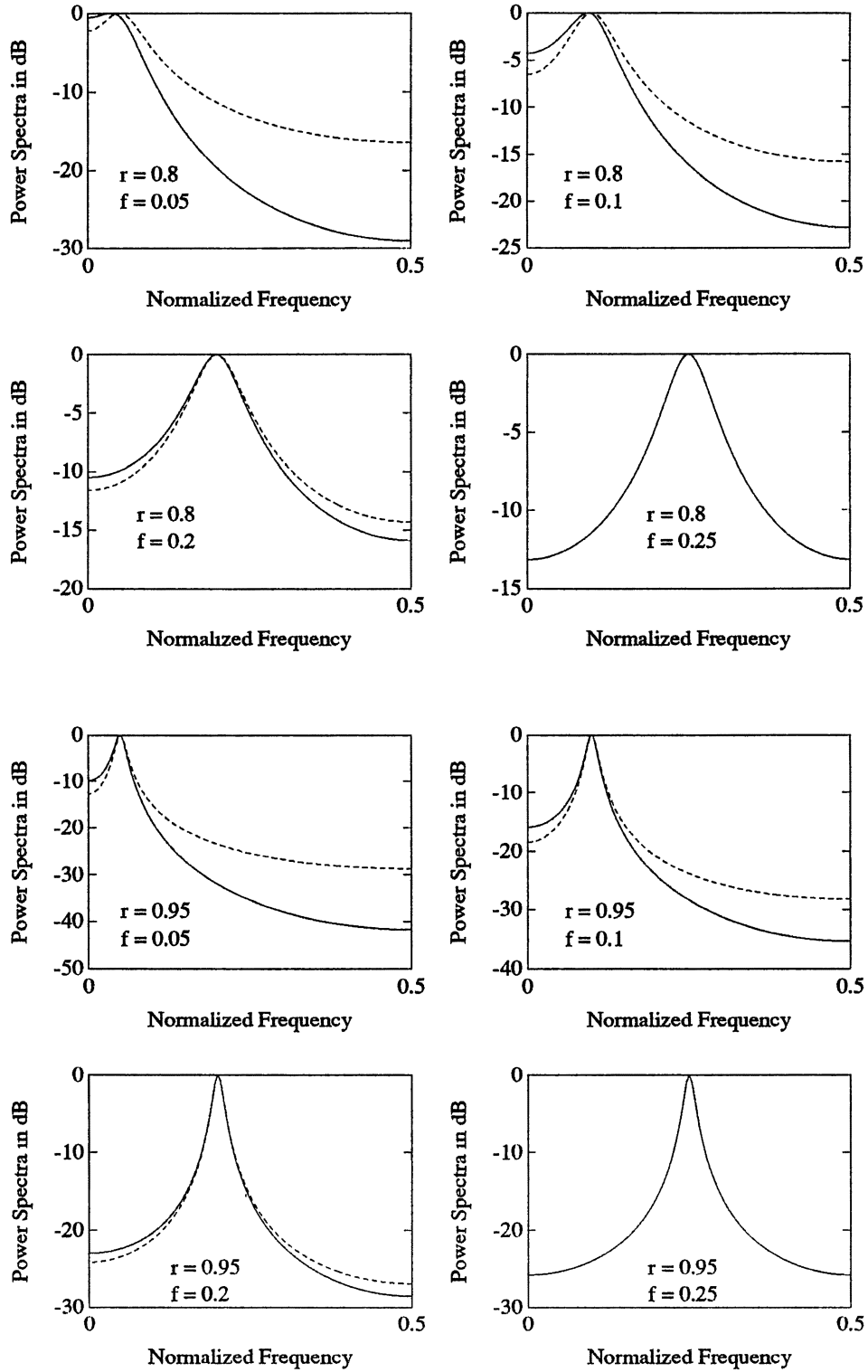


Figure 4.3. (Continued)

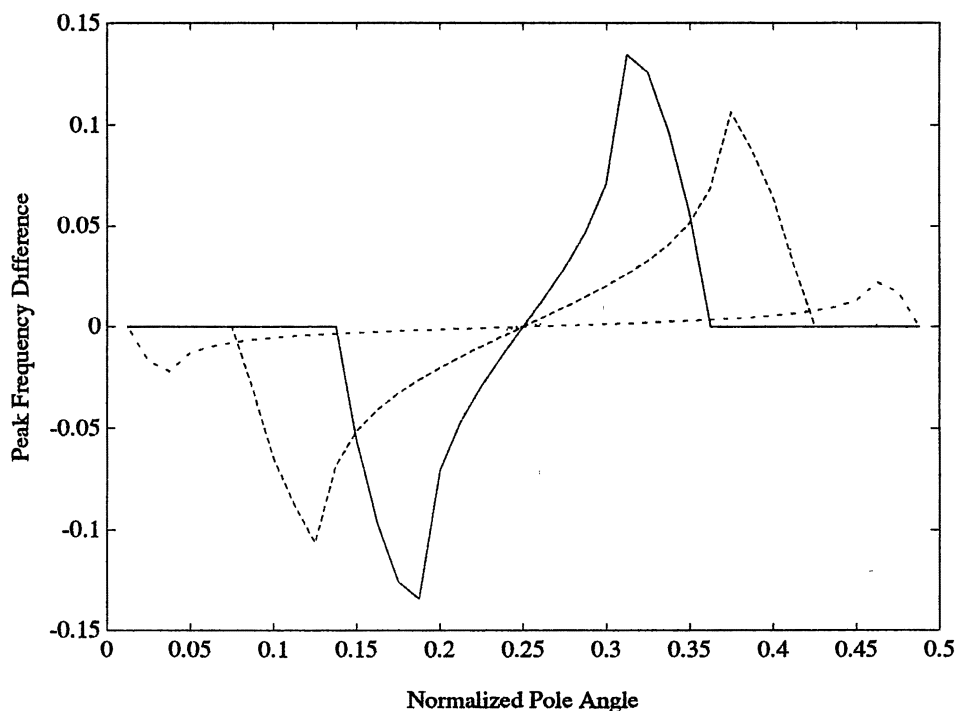


Figure 4.4. Difference in AR and MCRP Peak Frequencies. Pole magnitudes: 0.2 (solid), 0.4 (dash), 0.6 (dot), and 0.8 (dot-dash).

When there are multiple AR poles, the procedure to find the optimum MCRP becomes even more complex. Perhaps most important, the spectral coefficients $\{\beta_k\}$ for the MCRP poles have to be considered. For the one pole-pair case, the weights act only as a scaling factor, so they can be ignored. However, for multiple poles, the weights will in part determine the shape of the MCRP power spectrum.

Instead of pursuing an optimum match at this time, the author proposes the following procedure as a first approximation to match AR and MCRP power spectra. First, let the MCRP eigenvalues $\{\lambda_k\}$ be identical to the AR poles $\{p_i\}$. To see why this might work, consider that the AR poles with largest magnitudes will be the most predominant in the power spectrum. As in the single pole-pair case, as the pole magnitude increases, the

difference between the peaks of the AR and MCRP decreases. Thus, it would be reasonable to assume the same is true when multiple poles are considered. Of course, the poles with smaller magnitudes would likely not be matched very well under this assumption. However, these poles are not as significant to the power spectrum as are the poles with large magnitudes.

Assuming the necessary MCRP eigenvalues $\{\lambda_k\}$ have been determined, it remains to find the spectral coefficients $\{\beta_k\}$. Recall that the power spectrum for a MCRP is given by

$$\Phi(\omega) = \sum_{k=1}^{N-1} \beta_k \left(\frac{1}{1 - \lambda_k e^{j\omega}} + \frac{1}{1 - \lambda_k e^{-j\omega}} - 1 \right). \quad (\text{IV.11})$$

Thus, it is clear that $\Phi(\omega)$ is linear in $\{\beta_k\}$. This fact can be used to generate a set of linear equations in $\{\beta_k\}$ to insure that the MCRP and AR spectra agree at a given set of frequencies $\Omega = \{\omega_i\}$. Specifically, let the error function be

$$J = \sum_{\omega_i \in \Omega} e^2(\omega_i) \quad (\text{IV.12a})$$

$$\text{where } e(\omega_i) = \Phi_{\text{MCRP}}(\omega_i) - \Phi_{\text{AR}}(\omega_i). \quad (\text{IV.12b})$$

A good choice of frequencies Ω is the set of frequencies where the AR spectrum has an extrema. As this usually results in having more equations than unknowns, a least-squares solution is required.

As an example, consider the AR process with poles $.8e^{\pm 2\pi j}$ and $.6e^{\pm 7\pi j}$. Also, assume the white noise driving the linear filter has unity variance. Then, the MCRP poles are $\lambda_1 = .8e^{2\pi j}$, $\lambda_2 = \lambda_1^*$, $\lambda_3 = .6e^{7\pi j}$ and $\lambda_4 = \lambda_3^*$. Solving for the spectral coefficients by minimizing the sum of the squared difference between the power spectra at the extrema of the AR power spectrum yields $\beta_1 = .6971$, $\beta_2 = \beta_1$, $\beta_3 = .1055$ and $\beta_4 = \beta_3$. The power spectrum for the AR and MCRP processes are given in Figure 4.5. As can be seen, the spectra agree fairly well at the peaks.

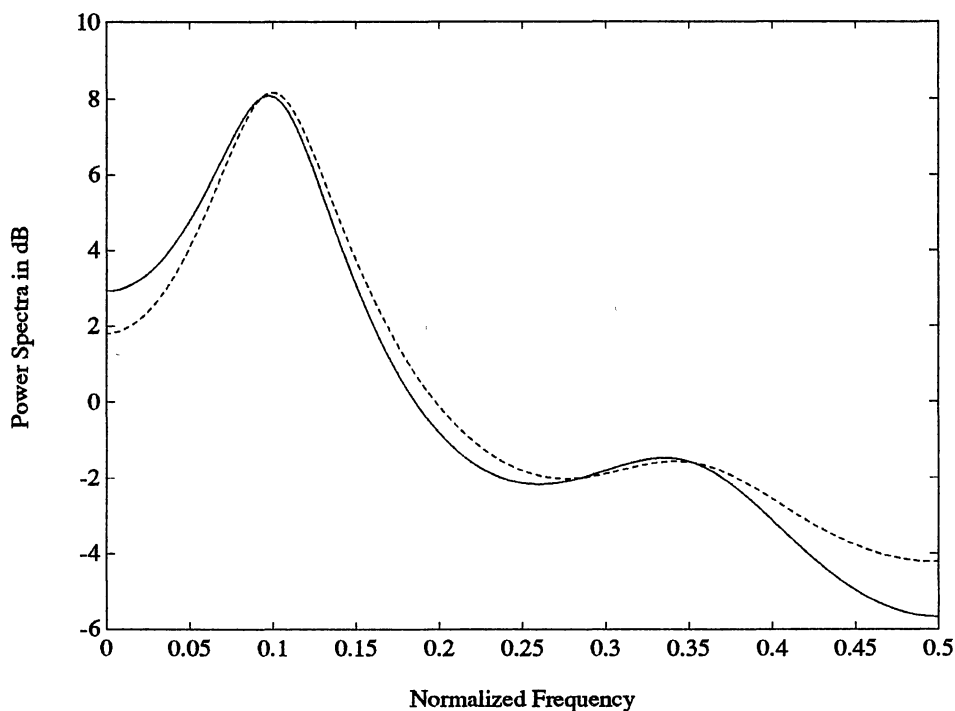


Figure 4.5. AR Power Spectrum (solid) and MCRP Power Spectrum (dashed) for Two Pole-Pair Example.

Although this procedure has not been investigated extensively, it appears to work fairly well. More examples will be given in the following chapter on speech modeling, which is based on much of the material from this chapter.

Nonstationary Processes

Before continuing to the next chapter, let us briefly consider the prospects for using MCRPs as models of nonstationary processes. There appear to be at least two paths which one could take on this endeavor. The first is to return to a discussion of Markov chains, and investigate chains which are nonstationary. However, while this may in fact lead to meaningful result, the author prefers the second approach.

The second approach is to try to make do with stationary MCRPs. Of course, the author does not mean to model a nonstationary process with a stationary process. Rather, the idea is to connect several dissimilar MCRPs together to model the nonstationary process. Recall from a previous discussion that many nonstationary processes can be treated as "short-time" stationary. That is, over small intervals of time, the process displays stationary characteristics.

In fact, it was this same discussion which arose from the review of Hidden Markov Models in Chapter II. For, HMMs provide a natural way to connect together the processes which model each short-time section. Typically, the short-time sections are modeled using linear, time-invariant systems (i.e., ARMA processes). However, there is no reason why MCRPs could not be used in this context. In fact, there is good reason why they should be used. Consider that the MCRP has the same Markov structure as an HMM. Consequently, the MCRP and the HMM could be joined together to form a complete Markov model for the nonstationary process.

A potential advantage of this approach lies in the transition region from one short-time section to another. For, if MCRPs and HMMs are used together, the transition could be specified more clearly by how the states in adjacent MCRPs are connected. It is the belief of the author that this approach can greatly enhance nonstationary modeling using HMMs.

However, the specifics of implementing the above idea are not immediately clear. While it was not the intention or goal of this thesis to explore nonstationary modeling, the author feels that this is a very natural and valid extension and deserves attention in the future.

CHAPTER V

SPEECH MODELING WITH MARKOV CHAIN

RANDOM PROCESSES

There are several reasons why one might wish to model speech. For one, speech signals contain much redundancy and digital transmission of speech signals can be accomplished at lower bit rates by transmitting the model parameters instead of the actual speech signal. Similarly, it is possible to reduce the storage requirements by storing the parameters of a mathematical model of the speech instead of the actual speech.

Other uses of speech modeling include speech recognition and speaker identification. In speech recognition, identification of spoken words is sought while in speaker identification the attempt is to identify the speaker. In these applications, the modeling transforms the speech signal into a domain whereby word or speaker recognition can be performed. Specifically, the model parameters are compared with sets of stored parameters to determine which word has been spoken or the speaker's identity.

Standard speech models are almost entirely based on autoregressive processes. The performance of these models is sufficient for many speech applications. However, there are problems with these models, such as poor performance in high noise environments [Sam76, Tea79], which provide motivation to consider other types of models.

Recently, there has been a good deal of interest in Markov chains within the speech processing community. This has come in the form of the hidden Markov model (HMM), which can be used to model the structure of speech on a phonemic level. Phonemes have been called the building blocks of speech, and can be thought of as the distinctive sounds

of a language. As speech is well structured, some phonemes are more likely to follow than others. Consequently, by letting each phoneme be a state in a Markov chain, the structure of speech can be modeled. However, this is complicated by the fact that, at any given time, there is no direct evidence to which phonemic state the speech is occupying. Thus, the current state of the Markov model is hidden, and can only be measured indirectly from the statistics of the speech. Hidden Markov models of speech have been employed widely in speech recognition [Rab89] and have recently been used in digital coding of speech signals [Far87, Dea89].

Motivated by the success of HMMs and by the problems with conventional speech models, this chapter explores the use of Markov chain random processes as a model for speech signals. Potential application of this model include all of the applications of conventional models: storage and transmission of speech signal, and speech and speaker recognition. Additionally, it is conceivable that a "complete" Markov model of speech might result by combining an MCRP model of speech at the sub-phonemic level and a HMM model of speech at the phonemic level. However, these applications are beyond the scope of this research; the goal here is merely to set forth the elements of MCRPs as models of speech and to discover whether they hold any promise.

Additionally, this application provides a good opportunity to gauge the success of both the spectral synthesis solutions set forth in Chapter III and the stochastic modeling strategy from Chapter IV. For, these techniques will play a part in the MCRP model of speech in this chapter. However, before this scheme is examined, it is necessary to discuss speech and AR models of speech.

Speech and Standard Speech Models

As speech is a continuous-time signal, the speech must first be sampled before any digital processing can begin. For most applications, speech frequencies above 4 kHz do

not play a significant role. Consequently, it is common to low-pass filter speech at 4 kHz and sample at 8 to 10 kHz. The result is a discrete-time sequence which is suitable for modeling. For the modeling criteria, consider that the human ear is more sensitive to the magnitude than the phase of the frequency response [Sch75]. Consequently, a natural criteria for speech modeling is the power spectrum.

Because speech is not stationary, it is necessary to model speech on a short-time basis by segmenting the signal into sections which are approximately stationary. For speech, it is common to use sections of 10-20 milliseconds [Rab78]. Thus, for each 10-20 millisecond section of speech, a stationary model is sought.

Short-time sections of speech are usually lumped into one of two categories: voiced or unvoiced. Voiced signals are periodic in nature on a short-time basis. Examples of voiced speech include the vowels. Conversely, those sounds with a distinctly aperiodic nature on a short-time basis are termed unvoiced. Unvoiced waveforms have little structure and appear random in nature. An example of an unvoiced sound is the /sh/ sound in the word 'show', while an example of voiced sound is the /o/ sound in 'show'.

Waveforms of voiced and unvoiced speech are shown in Figure 5.1.

Currently, the dominant mathematical models for discrete-time speech signals are based on autoregressive processes. As discussed in the previous chapter, autoregressive processes are characterized by a transfer function of the form

$$H(z) = \frac{1}{1 - \alpha_1 z^{-1} - \alpha_2 z^{-2} - \dots - \alpha_M z^{-M}}. \quad (\text{V.1})$$

From Chapter IV, the parameters $\{\alpha_i\}$ can be found by minimizing the prediction error between the model and the data. In this case, the model's power spectrum is a valid estimator of the data's power spectrum.

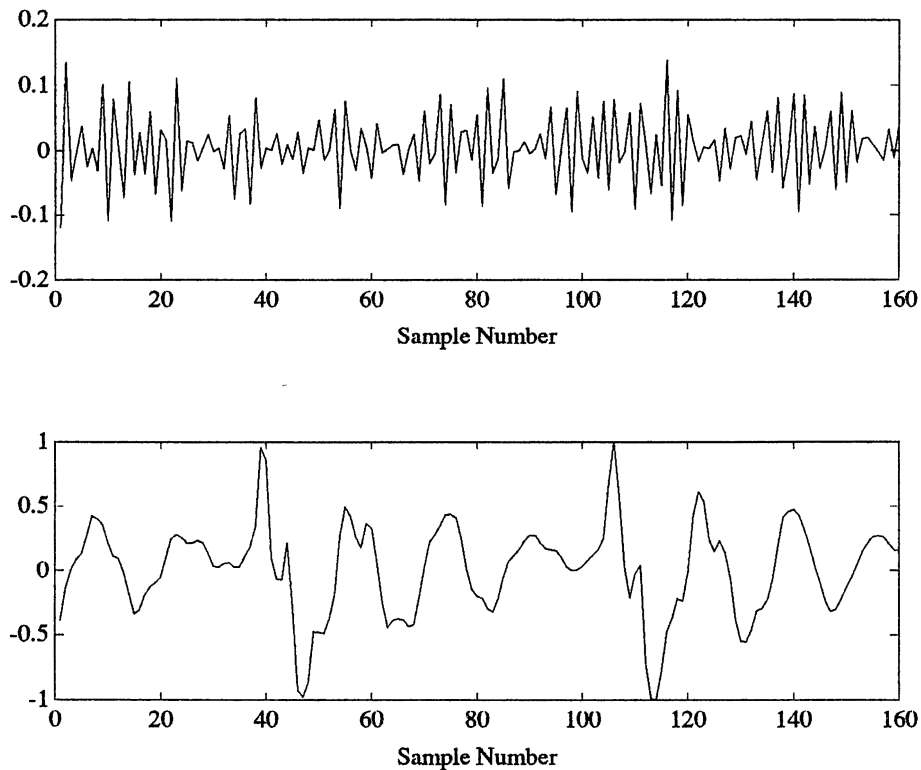


Figure 5.1. Speech Waveforms: Unvoiced /sh/ Sound from the Word 'Show' (top), and Voiced /o/ Sound from the Word 'Show' (bottom).

There are many different ways to implement an AR model of speech. Some of these are simple and some exotic, but they mainly differ by the choice of excitation signal. Here, the LPC model of speech shall be singled out for purposes of comparison with the MCRP speech model. For the LPC model, the differences in voiced and unvoiced sections of speech are reflected in the choice of input to the AR filter: for unvoiced sections, the excitation is white noise; for sections of speech which are voiced, the fundamental frequency is estimated from the waveform and a corresponding periodic impulse train provides the excitation.

MCRP Model of Speech

To consider MCRPs as a model for speech signals, much of the material from the previous chapter can be used. The previous chapter showed that AR processes (as well as ARMA processes) and MCRPs have similar power spectra. Unfortunately, an exact power spectra match for AR processes was not possible using known MCRP structures. However, an approximation was made so that the MCRP had a power spectrum which was close to the AR power spectrum. Thus, as in the previous chapter, an AR model can be used as the basis for the MCRP model.

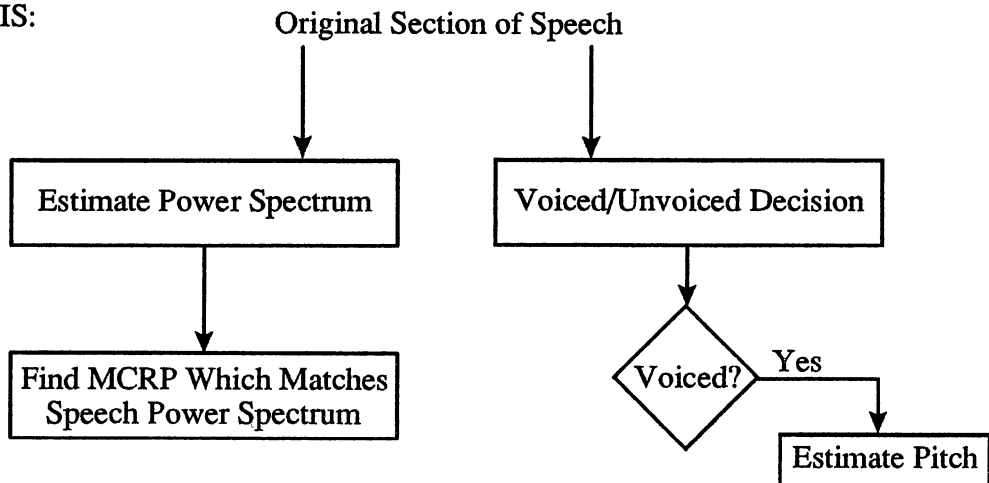
Because of the probabilistic nature of MCRPs, the output will appear random. Therefore, an MCRP seems well-suited to model unvoiced speech. By the same logic, however, there is a fundamental problem in using an MCRP to model voiced speech. One suggestion has been to filter the output of the MCRP with a comb filter [Sie76], which has the form

$$H_c(z) = \frac{1}{1 - rz^{-L}} \quad (\text{V.2})$$

where L would correspond to the pitch period. Essentially, this establishes correlation between samples $y(n)$ and $y(n - L)$, thereby inducing a periodic nature into the waveform. Thus, voiced sections of speech would be modeled by an MCRP which has been passed through a comb filter.

Based on this idea, Figure 5.2 shows a block diagram for analysis and synthesis of speech using an MCRP model. Note that, as in the AR model, analysis and synthesis are performed on a short-time basis. The resulting synthesized sections are concatenated to form an approximation of the original speech sequence. The voiced/unvoiced decision and pitch estimation can be made using any of the techniques in use with AR models. A comparison of LPC and MCRP speech models is given in Table 5.1.

ANALYSIS:



SYNTHESIS:

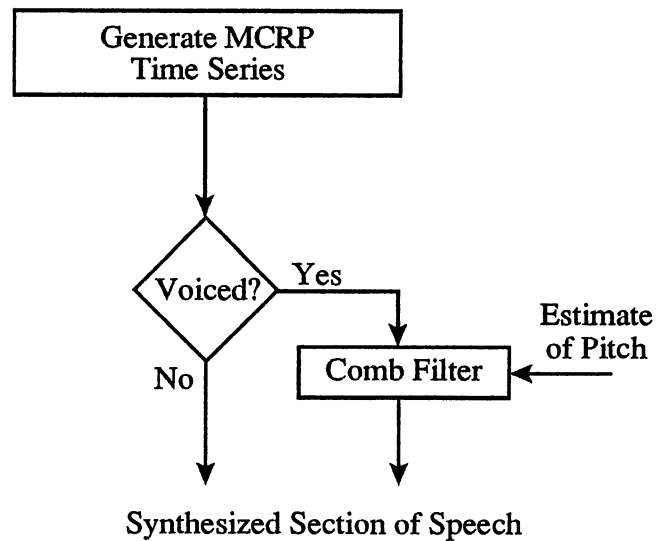


Figure 5.2. Block Diagram of MCRP Model of Speech.

TABLE 5.1
COMPARISON OF LPC AND MCRP MODELS

	LPC Model	MCRP Model
<u>Analysis:</u>	Minimize Prediction Error	Match Power Spectra
<u>Synthesis:</u>		
Unvoiced:	White Noise Excitation	No Processing Necessary
Voiced:	Periodic Impulse Train Excitation	Pass MCRP through Comb Filter

Examples

In this section, several examples of speech modeling using MCRPs are evaluated. The goal is to evaluate the performance of MCRPs as models for speech: in short, can an MCRP model generate speech-like sounds? At this point, it should be noted that, even though the concepts of an AR-based speech model such as LPC are quite simple, the actual implementation which achieves useable results is fairly complicated. To do something similar for MCRPs is beyond the scope of this document. However, it is still possible to test the modeling scheme on a scaled down basis by considering a few phonemes.

Additionally, we shall steer away from some of the more problematic phonemes such as diphthongs and voiced fricatives. These represent special challenges which would only confuse the issue. Also, instead of dealing with phonemes directly, we shall look instead at several simple words which can be adequately modeled with only a couple of phonemes. For, it is difficult to determine speech quality from unconnected phonemes.

Towards this goal, several simple words have been chosen. These include the words 'off', 'us', 'ash', and 'eat'. Each of these words contain exactly two phonemes: one voiced and one unvoiced phoneme. Digital samples of each of these words were recorded and the words were then segmented into sections representing silence, voiced phoneme, or unvoiced phoneme. For each section of phoneme, a representative frame was chosen for analysis purposes. For each representative frame, the AR parameters were determined. MCRP parameters were then determined as described previously in Chapter IV. Also, for voiced phonemes, estimates of the pitch periods were made.

The resulting spectra for each phoneme are given below in Figure 5.3. Each plot shows the AR and MCRP power spectra for a specified phoneme. The MCRP curves do a fairly good job of estimating the LPC power spectra. Some are a bit poor, and some are exceptional. The match in the /f/ phoneme deserves special recognition. Overall, the most significant peaks appear to be matched very well. However, the MCRP power spectra do not match the AR spectra at frequencies of low magnitude. But then, these frequencies are less important.

Based upon the model parameters, synthetic speech was generated. This was undertaken in three phases. In the first phase, the speech was synthesized from the AR parameters as in LPC. The purpose of this phase was to confirm the validity of the analysis. In each case, the resulting synthetic speech sounded very similar to the original speech.

The next phase was an intermediate step between the LPC and MCRP models. In this step, the voiced sections were generated in the same fashion proscribed for the MCRP model; namely, white noise was passed through the LPC filter to perform spectral shaping and then through a comb filter to induce periodicity. Essentially, the results of this test act as an upper limit on the performance of the MCRP model. That is, if exact AR-MCRP spectral matching is achieved, the sound quality of the MCRP speech should

be equivalent to the speech for this intermediate step. However, since exact AR-MCRP spectral matching was not achieved, the MCRP speech is not expected to sound quite as good as the synthesis speech for this intermediate step. Informal listening tests showed that synthetic speech from this process sounded much like the LPC speech. However, the voiced sections were not of as good quality. In particular, there was a noticeable ringing present, although it was not severe.

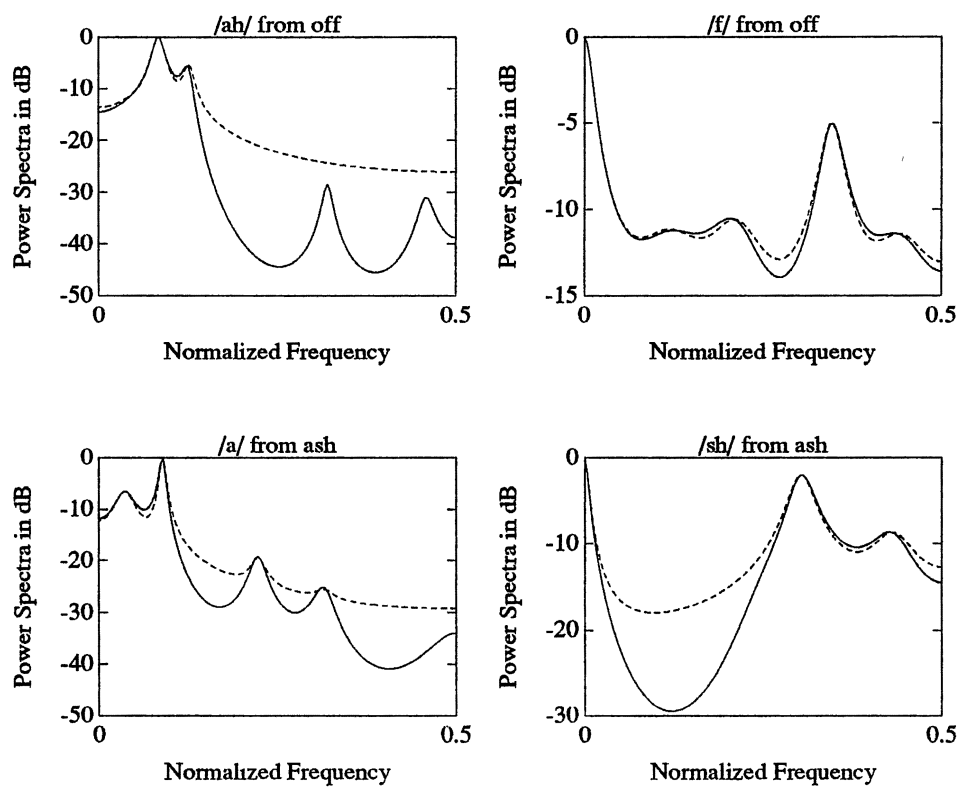


Figure 5.3. Power Spectra of AR (solid) and MCRP (dashed) for Various Phonemes.

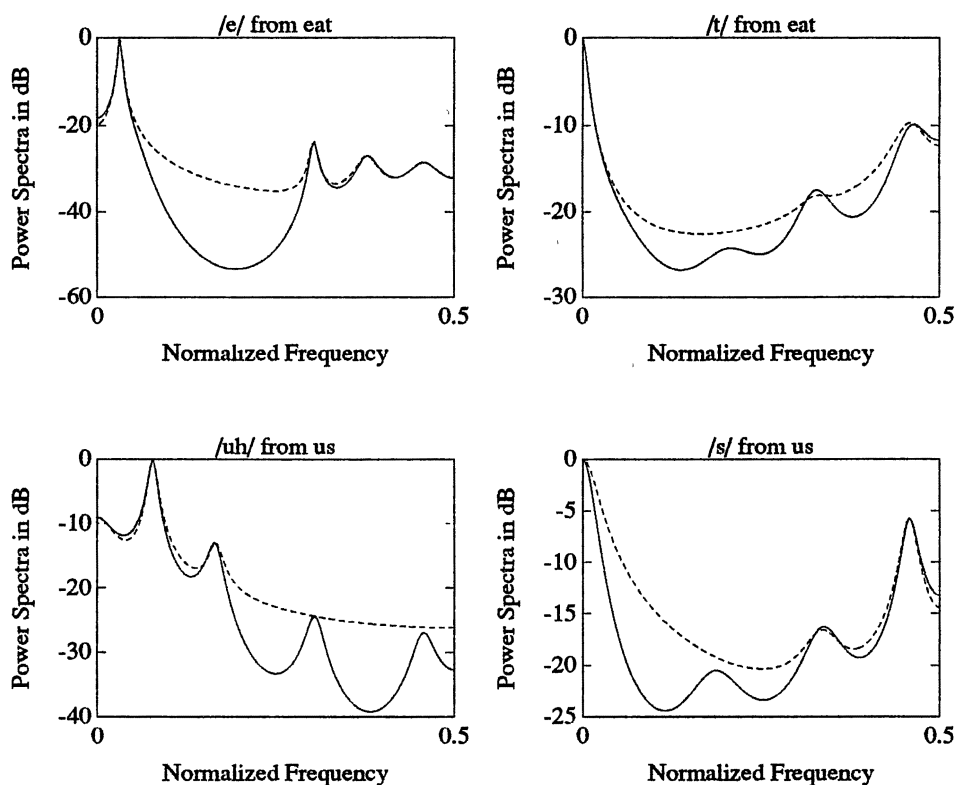


Figure 5.3. (Continued.)

Finally, the MCRP synthetic speech was produced. As expected, since exact power spectral matching was not achieved, the synthetic speech was not quite as good as the speech from the intermediate phase. However, the results were not significantly worse. In fact, the most discernable and objectionable problem with the speech is the ringing which was also present in the intermediate-phase speech. Plots of the original and synthetic speech are given below in Figures 5.4 through 5.7. In each plot, the top figure is the original speech and the bottom plot is the MCRP speech.

To summarize, an MCRP model of speech can be based upon the stochastic modeling strategy of Chapter IV. However, a modification had to be made to account for voiced speech. Although there are details which would have to be worked out in any

practical implementation of this model, the results of this section have shown that the model is capable of generating speech-like sounds. Thus, this model appears to be a valid model of speech. Additionally, the positive results of this chapter confirm the validity of the MCRP modeling techniques introduced in the previous chapter.

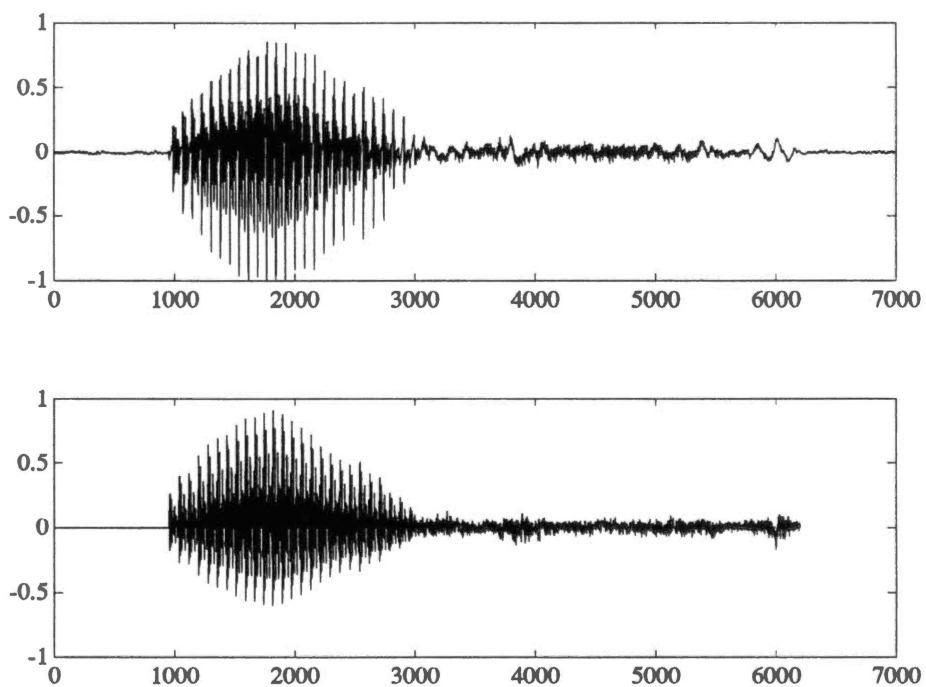


Figure 5.4. Original (top) and Synthetic (bottom) Waveforms for 'Off'.

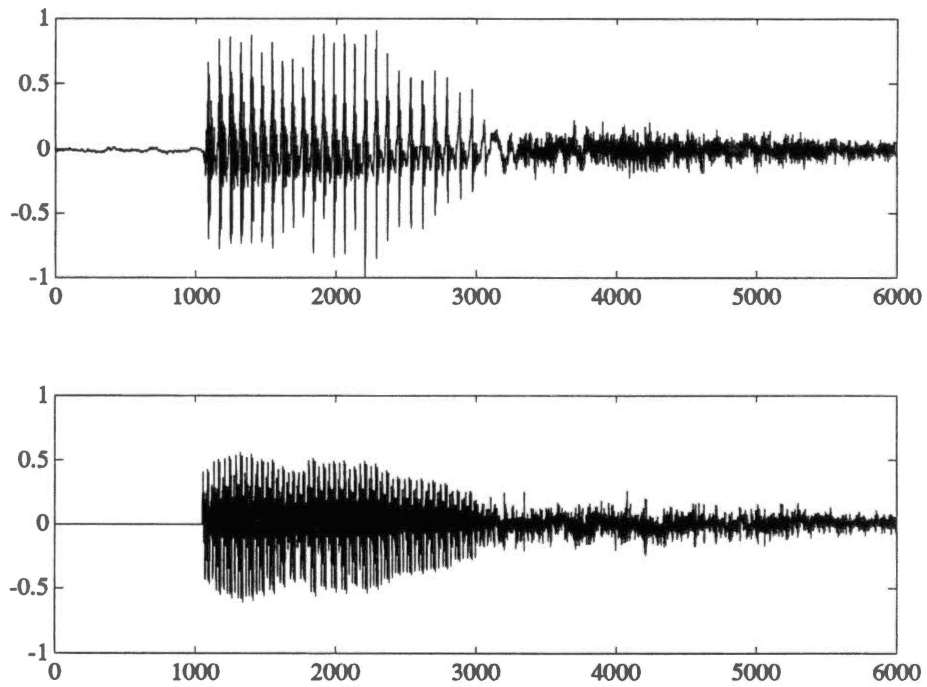


Figure 5.5. Original (top) and Synthetic (bottom) Waveforms for 'Ash'.

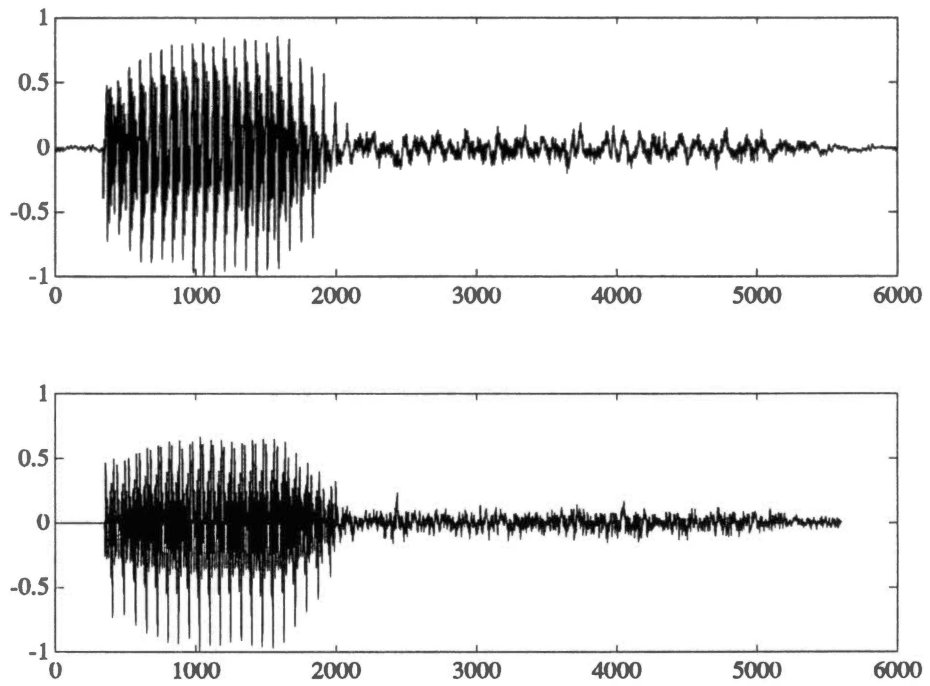


Figure 5.6. Original (top) and Synthetic (bottom) Waveforms for 'Us'.

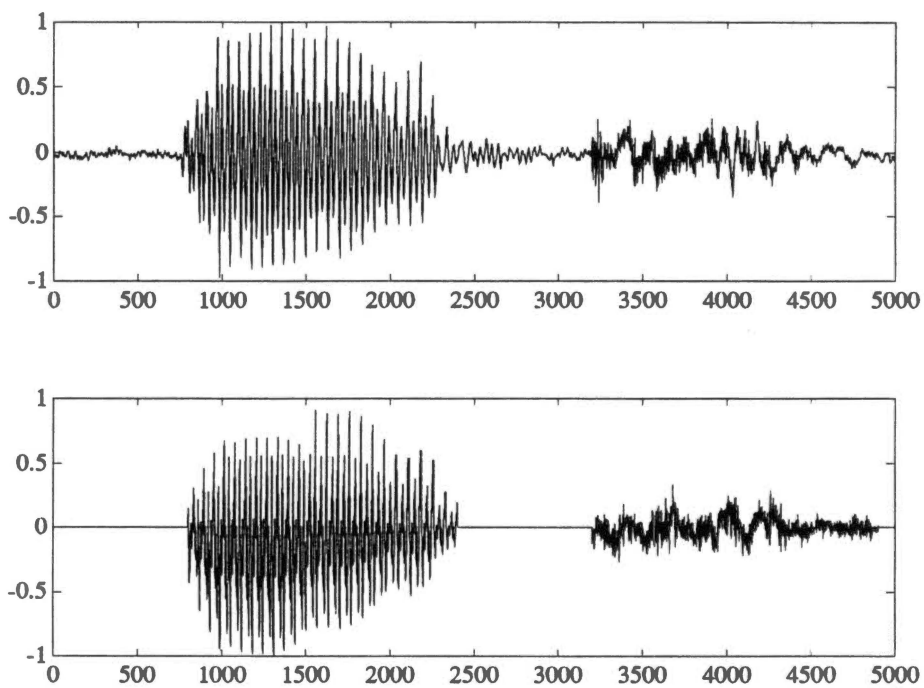


Figure 5.7. Original (top) and Synthetic (bottom) Waveforms for 'Eat'.

CHAPTER VI

SPREAD SPECTRUM COMMUNICATIONS

Since the advent of the telegraph over a hundred years ago, we have been able to communicate quickly and efficiently by electronic means. Since then, there have been many developments in communications. In recent years, the thrust has been toward digital communications. Of course, this is clearly well-suited for some signals, such as those generated by computers. However, even such traditionally analog domains as telephone channels have been converting to digital transmissions. There are several reasons for this. Perhaps foremost, digital communications offers performance and flexibility unattainable by analog communications [Zie85].

There are many different digital communications schemes, and a specific implementation depends on the requirements of the application. Influencing factors include bandwidth, power, desired error performance, cost, complexity, data security, and interference rejection among others. Some sophisticated techniques have been developed to deal with these problems. Among these tools is spread spectrum communications.

As its name implies, spread spectrum is a technique which spreads the signal's spectrum. There are several advantages of this. Among these are rejection of jamming and other interfering signals, low probability of detection (LPD) and low probability of interception (LPI) from undesired parties, and multiple user access in the same frequency band [Pic82, Skl88]. While spread spectrum has traditionally been a tool for military communications, it has seen recent growth in commercial applications [Sch90].

Spread spectrum systems are usually classified as either frequency hopping, direct sequence, or hybrid. (For frequency hopping, the signal's center frequency is constantly changed, while for direct sequence, the signal is spread by direct modulation with a high-bandwidth signal. Hybrid systems are combinations of frequency hopping and direct sequence. In this chapter, MCRPs are considered for use as the spreading signal in direct sequence spread spectrum.

Direct Sequence Spread Spectrum

A block diagram of a direct-sequence spread spectrum (DS-SS) system is shown below in Figure 6.1. On the transmitter side, the ± 1 -valued data signal $d(t)$ is first modulated. For this discussion, binary phase-shift keying (BPSK) modulation shall be used, although other modulation schemes can be employed. The modulated signal is

$$s_m(t) = \sqrt{\frac{2E_b}{T_d}} \cos[2\pi f_0 t + \theta_d(t)] \quad (\text{VI.1})$$

where f_0 is the carrier frequency, E_b is the energy per data bit of the transmitted signal, T_d is the data bit duration, and $\theta_d(t)$ is a function of the data sequence $d(t)$. The simplest choice for $\theta_d(t)$ here is $\theta_d(t) = 0$ for $d(t) = 1$ and $\theta_d(t) = \pi$ for $d(t) = -1$. In this case, the modulated signal can be expressed as

$$s_m(t) = \sqrt{\frac{2E_b}{T_d}} d(t) \cos[2\pi f_0 t]. \quad (\text{VI.2})$$

The modulated signal $s_m(t)$ is then spread by multiplication with the spreading or code signal $c(t)$. The transmitted signal is

$$s_t(t) = \sqrt{\frac{2E_b}{T_d}} c(t) d(t) \cos[2\pi f_0 t]. \quad (\text{VI.3})$$

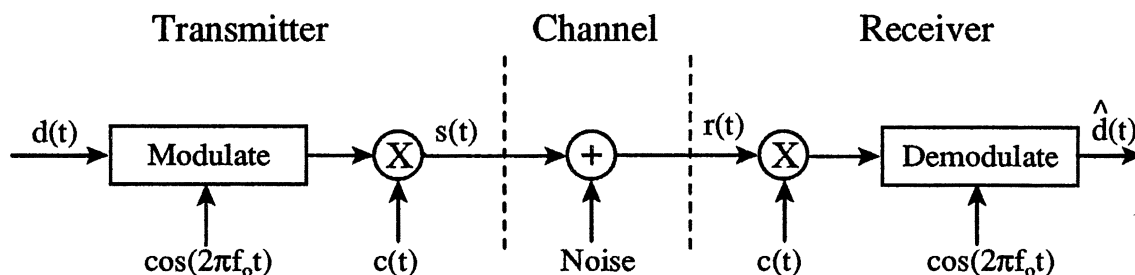


Figure 6.1. Block Diagram of DS-SS System.

Desirable properties for $c(t)$ include a large bandwidth relative to the data signal, ease of generation, and being deterministic but possessing random-like qualities. Such a signal can be obtained from pseudo-noise (PN) sequences termed m-sequences [Her86, Coo86, Pic82]. These are special binary sequences which can be implemented easily and efficiently using shift registers.

At the receiver, the signal is first despread by the code signal. Next comes demodulating. For a BPSK modulation scheme, this is accomplished by correlating the input signal with a coherent reference signal. Specifically, the input to the demodulator is multiplied by a coherent reference signal and then integrated over the data bit duration T_d . Finally, an estimate $\hat{d}(t)$ of $d(t)$ is made based upon the output of the correlator.

To illustrate the effects of spreading, consider the power spectra of the transmitted signal with and without spreading. First, consider the power spectrum of the data signal. The continuous-time data signal is obtained by passing the discrete-time data signal through a zero order hold. That is, if d_n is the discrete-time signal, the continuous-time signal is given by $d(t) = d_n$ for $nT_d \leq t < (n+1)T_d$. Then, if $\Phi_d(\omega)$ is the power spectrum of $\{d_n\}$, it can be shown that the power spectrum of $d(t)$ is given by [Zie85]

$$S_d(f) = T_d \text{sinc}^2(fT_d) \Phi_d(2\pi fT_d). \quad (\text{VI.4})$$

Now, assume that the data sequence is white. That is, $E[d_n d_m] = \delta(n - m)$, or equivalently $\Phi_d(\omega) = 1$. This is a valid assumption as the source is usually passed through a source encoder whose purpose is to remove redundancy from the signal. In this case, the power spectrum of the continuous-time data signal is

$$S_d(f) = T_d \text{sinc}^2(fT_d). \quad (\text{VI.5})$$

When the data signal is modulated by a sinusoidal signal, the power spectrum is translated in the frequency domain to plus and minus the frequency of the carrier. Thus, including the gain term, the power spectrum of the transmitted signal is [Sch90]

$$S_m(f) = E_b \{ \text{sinc}^2[(f - f_0)T_d] + \text{sinc}^2[(f + f_0)T_d] \}. \quad (\text{VI.6})$$

Now consider the power spectrum after spreading. First, the power spectrum of the spreading signal must be found. The spreading signal is generated in a similar manner to the data signal by passing a sequence through a zero-order hold. Hence, as in Equation (VI.4), the power spectrum of the spreading signal is given by

$$S_c(f) = T_c \text{sinc}^2(fT_c) \Phi_c(2\pi fT_c) \quad (\text{VI.7})$$

where T_c is the code bit duration and $\Phi_c(\omega)$ is the power spectrum of the sequence from which $c(t)$ is obtained. In this case, $c(t)$ is based upon an m-sequence which has a flat spectrum; i.e., the power spectrum of the m-sequence approximates $\Phi_c(\omega) = 1$. Hence, $S_c(f) = T_c \text{sinc}^2(fT_c)$.

The spectrum of the spread signal may now be found. As multiplication in the time domain corresponds to convolution in the frequency domain, the power spectrum of the spread signal is $S_s(f) = S_m(f) * S_c(f)$. Let $R_d = 1/T_d$ be the bit-rate for the data and $R_c = 1/T_c$ be the bit-rate for the code sequence. The ratio $R_c/R_d = T_d/T_c$ is denoted G_p , and referred to as the processing gain. Typically, $G_p \gg 1$, so that the spreading signal has a much

larger bandwidth than the data signal. In this case, $S_m(f)$ has such a much smaller bandwidth than $S_c(f)$, so $S_m(f)$ can be approximated as impulses at $\pm f_0$ for the purposes of the convolution. Specifically, $S_m(f) \approx (E_b/T_d)\delta(f - f_0) + (E_b/T_d)\delta(f + f_0)$. Thus,

$$\begin{aligned}
 S_s(f) &= S_m(f) * S_c(f) \\
 &\approx \frac{E_b}{T_d} S_c(f - f_0) + \frac{E_b}{T_d} S_c(f + f_0) \\
 &= \frac{E_b T_c}{T_d} \{ \text{sinc}^2[(f - f_0)T_c] + \text{sinc}^2[(f + f_0)T_c] \} \\
 &= \frac{E_b}{G_p} \{ \text{sinc}^2[(f - f_0)T_c] + \text{sinc}^2[(f + f_0)T_c] \} \quad (\text{VI.8})
 \end{aligned}$$

Although G_p would normally be much larger, for purposes of illustration, a plot of $S_m(f)$ and $S_s(f)$ is given in Figure 6.2 for $G_p = 4$.

The effect of the spreading is to reduce the magnitude of the power spectrum by $1/G_p$ while increasing the bandwidth by G_p . This makes it more difficult for the signal to be detected by an undesirable party. Even if the signal is detected, it is difficult for the data to be intercepted as the m-sequence would have to be known. However, it is possible to determine the m-sequence from the transmitted signal [Tor81, Her86], though this is computationally expensive. Consequently, if secure communications are required, the data should be encrypted.

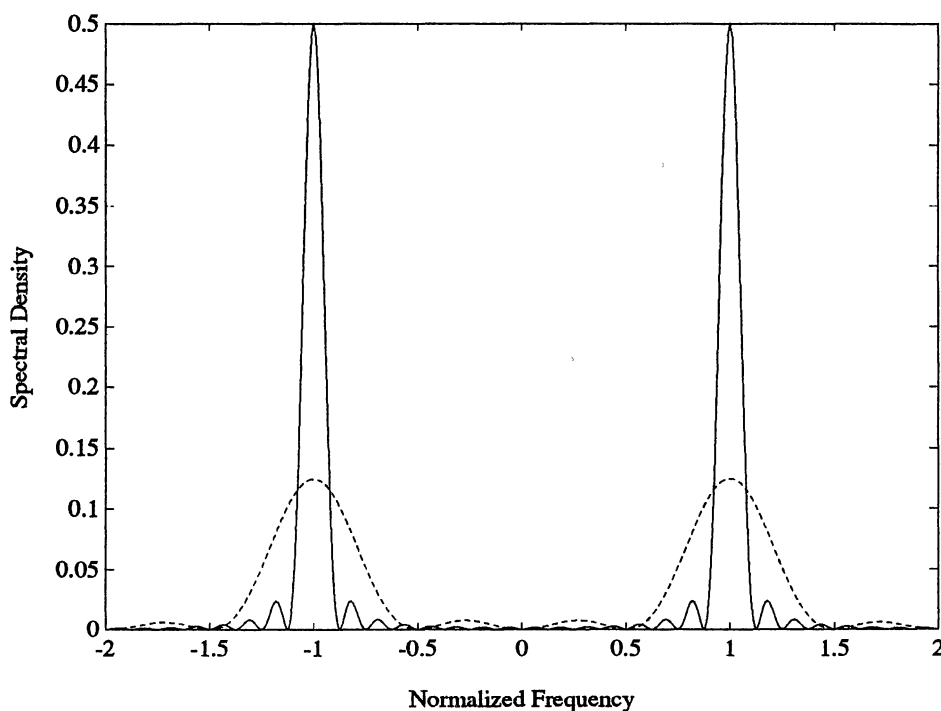


Figure 6.2. Power Spectrum of Signal without Spreading (solid) and Signal with Spreading (dotted), for $G_p = 4$.
(The Frequency Axis is Normalized about f_0)

MCRP Spread Spectrum

Consider using an MCRP as the code sequence. The potential advantage of this approach is that an MCRP is much more flexible than an m-sequence. Whereas the spectrum of an m-sequence is always white, the spectrum of an MCRP can be colored. Of course, an MCRP is by definition a random process. This seems to be a problem as the spreading sequence should be deterministic. However, in any implementation of an MCRP, a random number generator will have to be used to determine the state trans-

itions. If this random number generator is portable, it can be used at the transmitter and receiver with identical results. The consequence is that identical MCRP time series can be generated at both the transmitter and receiver.

The idea of using MCRPs for direct sequence spread spectrum was proposed by John Hershey [Her81a]. His idea centered around colored spreading sequences which would spread the data unevenly across the transmission bandwidth. To "fill out" the spectrum, noise would be added to the signal before transmission. Thus, the resulting spectrum would be the same as in direct sequence spread spectrum. If a jammer were to jam only a portion of the signal's bandwidth, which is a common technique, the frequencies in which the data is most heavily represented might be missed. If this were not the case, then perhaps the data and added noise frequencies could be reallocated.

A block diagram of a spread spectrum system based on this idea is shown in Figure 6.3. This scheme shall be referred to as MCRP-SS for short. As in DS-SS, the data signal is first modulated and then multiplied by a spreading signal $c_1(t)$. The result is the signal

$$s_1(t) = \sqrt{\frac{E_b}{T_d}} c_1(t) d(t) \cos[2\pi f_0 t]. \quad (\text{VI.9a})$$

At this point, a constant signal is modulated with the same carrier frequency but not necessarily the same phase. This modulated signal is multiplied by the spreading signal $c_2(t)$ which has the same bit rate as $c_1(t)$. In this way, an additive noise source is generated which is given by

$$s_2(t) = \sqrt{\frac{E_b}{T_d}} c_2(t) \cos[2\pi f_0 t + \phi]. \quad (\text{VI.9b})$$

The resulting transmitted signal is then given by

$$s_t(t) = \sqrt{\frac{E_b}{T_d}} c_1(t) d(t) \cos[2\pi f_0 t] + \sqrt{\frac{E_b}{T_d}} c_2(t) \cos[2\pi f_0 t + \phi]. \quad (\text{VI.10})$$

Note that the transmission power has been split between the data and additive noise signals.

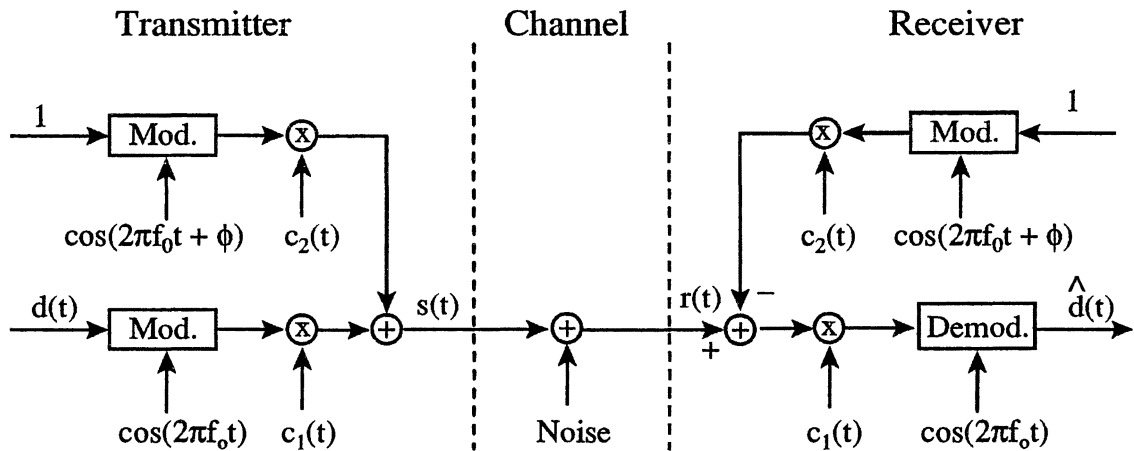


Figure 6.3. Block Diagram of MCRP-SS System.

Let us look at the spectra of these signals. Let $\Phi_1(\omega)$ and $\Phi_2(\omega)$ be the power spectra of the sequences from which $c_1(t)$ and $c_2(t)$ are obtained, respectively. Using the same procedure as in the previous section, the power spectra of $s_1(t)$ and $s_2(t)$ are

$$S_1(f) = \frac{E_b}{2G_p} \{ \Phi_1(2\pi(f-f_0)T_c) \text{sinc}^2[(f-f_0)T_c] + \Phi_1(2\pi(f+f_0)T_c) \text{sinc}^2[(f+f_0)T_c] \} \quad (\text{VI.11a})$$

$$\text{and } S_2(f) = \frac{E_b}{2G_p} \{ \Phi_2(2\pi(f-f_0)T_c) \text{sinc}^2[(f-f_0)T_c] + \Phi_2(2\pi(f+f_0)T_c) \text{sinc}^2[(f+f_0)T_c] \}, \quad (\text{VI.11b})$$

respectively. Assuming the spreading sequences are independent, the power spectrum of $s_i(t)$ equals the sum of the component's power spectra:

$$S_i(f) = \frac{E_b}{2G_p} \{ [\Phi_1(2\pi(f-f_0)T_c) + \Phi_2(2\pi(f-f_0)T_c)] \text{sinc}^2[(f-f_0)T_c] + [\Phi_1(2\pi(f+f_0)T_c) + \Phi_2(2\pi(f+f_0)T_c)] \text{sinc}^2[(f+f_0)T_c] \} \quad (\text{VI.12})$$

Next, suppose that $\Phi_1(\omega)$ and $\Phi_2(\omega)$ are such that $\Phi_1(\omega) + \Phi_2(\omega)$ is constant for all ω .

Pairs of spreading sequences which have this property shall be referred to as being *complementary*. Further, if the code sequences are complementary and ± 1 -valued, then it can be shown that $\Phi_1(\omega) + \Phi_2(\omega) = 2$ for all ω . In this case,

$$S_i(f) = \frac{E_b}{G_p} \{ \text{sinc}^2[(f-f_0)T_c] + \text{sinc}^2[(f+f_0)T_c] \} \quad (\text{VI.13})$$

which is identical to the power spectrum for DS-SS. Hence, if the spreading sequences are independent with complementary power spectra, the resulting power spectrum for the transmitted signal will be identical to that for DS-SS.

The potential advantage of this scheme is that the communication scheme appears to be standard DS-SS, at least in the frequency domain. Consequently, the jammer will likely employ a jamming noise which is detrimental to standard DS-SS. Thus, the jammer is fooled and it is likely that the jamming noise will be spectrally disjoint with the data signal. If it is not, then it may be possible to alter the spreading signals $c_1(t)$ and $c_2(t)$ so that this is the case.

Complementary MCRPs

Of course, for this scheme to work complementary sequences must be found. Thus, the first step must be to locate complementary MCRPs. But what other properties should these sequences possess? The answer is that they should be as spectrally disjoint as pos-

sible. As an illustrative example, suppose that $\Phi_1(\omega)$ and $\Phi_2(\omega)$ are as shown below in Figure 6.4. Note that the power spectra are disjoint and that $\Phi_1(\omega) + \Phi_2(\omega) = 2$ for all ω . Thus, two processes which had power spectra $\Phi_1(\omega)$ and $\Phi_2(\omega)$, respectively, would be complementary, spectrally-disjoint processes.

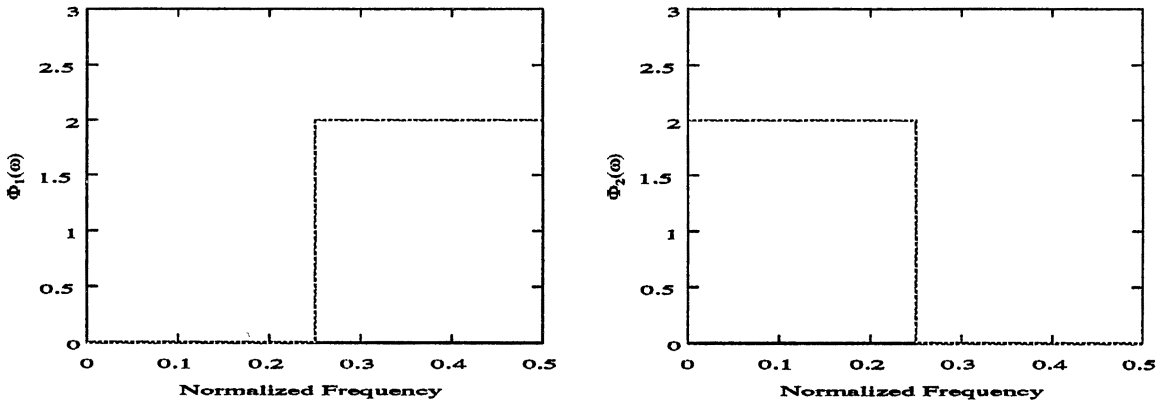


Figure 6.4. Power Spectra of Discrete-time Complementary Processes.

To see why being spectrally disjoint is desirable, recall the modulated and spread data power spectrum $S_1(f)$ and the complementary noise power spectra $S_2(f)$ as given in Equations (VI.11a) and (VI.11b), respectively:

$$S_1(f) = \frac{E_b}{2G_p} \{ \Phi_1(2\pi(f-f_0)T_c) \operatorname{sinc}^2[(f-f_0)T_c] + \Phi_1(2\pi(f+f_0)T_c) \operatorname{sinc}^2[(f+f_0)T_c] \} \quad (\text{VI.11a})$$

$$\text{and } S_2(f) = \frac{E_b}{2G_p} \{ \Phi_2(2\pi(f-f_0)T_c) \operatorname{sinc}^2[(f-f_0)T_c] + \Phi_2(2\pi(f+f_0)T_c) \operatorname{sinc}^2[(f+f_0)T_c] \}, \quad (\text{VI.11b})$$

If $\Phi_1(\omega)$ and $\Phi_2(\omega)$ from Figure 6.4 are used, the resulting power spectra $S_1(f)$ and $S_2(f)$ would be as illustrated in Figure 6.5. For clarity, the figure shows only those frequencies about f_0 . Again, the power spectra are disjoint. Further, if the power spectra were added together, the result would have a sinc^2 shape identical to the spectral shape of the transmitted signal for DS-SS. However, suppose a jammer employed a jamming signal of relatively narrow bandwidth about f_0 . In this case, the jamming signal and the data signal occupy completely different frequencies since the data spectrum, $S_1(f)$, occupies none of the frequencies about f_0 . It is precisely this reason why the spreading sequences need to be spectrally disjoint.

Now, the power spectra in Figure 6.4 were examples of ideal cases. In practice it is not possible to generate binary MCRPs which achieve these power spectra. However, the following seems to be a good approximation. First, recall that a random sequence is white if its autocorrelation can be expressed as $\phi(n) = K\delta(n)$; i.e., $\phi(n)$ is zero for all $n \neq 0$, and is equal to some constant K for $n = 0$. Recall also that, for MCRPs, the autocorrelation is given by $\phi(n) = \sum_k \beta_k \lambda_k^{|n|}$. Suppose that the two MCRPs have eigenvalues $\lambda_{1,k} = r e^{j2\pi k/K}$ and $\lambda_{2,k} = r e^{j2\pi k/K}$, and spectral coefficients $\beta_{1,k}$ and $\beta_{2,k}$, respectively. Then, the autocorrelations of the two processes can be expressed as

$$\phi_1(n) = \sum_{k=0}^{K-1} \beta_{1,k} (r e^{j2\pi k/K})^{|n|} \quad (\text{VI.14a})$$

$$\text{and } \phi_2(n) = \sum_{k=0}^{K-1} \beta_{2,k} (r e^{j2\pi k/K})^{|n|} \quad (\text{VI.14b})$$

respectively. Suppose further that the two processes are independent and

$\beta_{1,k} + \beta_{2,k} = B$ for all k . Then, as the autocorrelation of the sum of two independent processes is the sum of their autocorrelations,

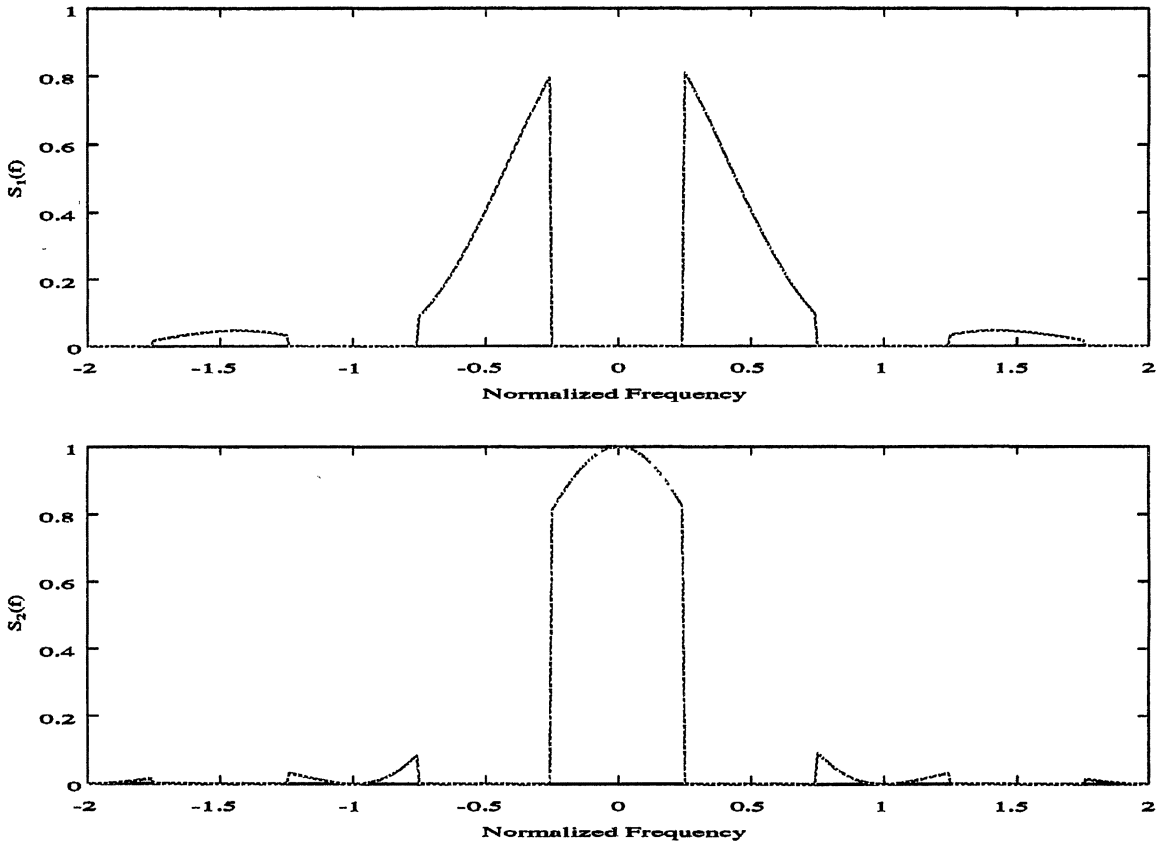


Figure 6.5. Resulting Signal Spectrum (top) and Added-Noise Spectrum (bottom).

$$\begin{aligned}
 \phi_{\text{sum}} &= \phi_1(n) + \phi_2(n) \\
 &= \sum_{k=0}^{K-1} \mathbf{B}(r e^{j2\pi k/K})^{|n|} \\
 &= \begin{cases} \mathbf{BK}r^{|n|} & \text{for } n = 0, \pm K, \pm 2K, \dots \\ 0 & \text{else} \end{cases} \quad (\text{VI.15})
 \end{aligned}$$

While the resulting power spectrum would not be exactly white, it would be very close for large K or small r .

For binary-valued MCRPs, the following approximately complementary MCRPs have been identified. The first MCRP is given by

$$\mathbf{T}_a = \begin{pmatrix} \frac{(1-r)}{4} & r + \frac{(1-r)}{4} & \frac{(1-r)}{4} & \frac{(1-r)}{4} \\ \frac{(1-r)}{4} & \frac{(1-r)}{4} & r + \frac{(1-r)}{4} & \frac{(1-r)}{4} \\ \frac{(1-r)}{4} & \frac{(1-r)}{4} & \frac{(1-r)}{4} & r + \frac{(1-r)}{4} \\ r + \frac{(1-r)}{4} & \frac{(1-r)}{4} & \frac{(1-r)}{4} & \frac{(1-r)}{4} \end{pmatrix}$$

$$\text{and } \mathbf{a}_a = (1 \quad 1 \quad -1 \quad -1).$$

Here, r has been introduced as a variable. Any value of r satisfying $0 < r < 1$ will work.

The choice of a particular value for r will be discussed shortly. For this MCRP, the eigenvalues which have non-zero weights are $re^{j\pi/2}$ and $re^{-j\pi/2}$ with $\beta_k = .5$ for both eigenvalues. The second MCRP uses the ring structure developed in Chapter III with two subchains:

$$\mathbf{T}_b = \begin{pmatrix} r & 0 & \frac{(1-r)}{4} & \frac{(1-r)}{4} & \frac{(1-r)}{4} & \frac{(1-r)}{4} \\ 0 & r & \frac{(1-r)}{4} & \frac{(1-r)}{4} & \frac{(1-r)}{4} & \frac{(1-r)}{4} \\ \frac{(1-r)}{2} & \frac{(1-r)}{2} & 0 & r & 0 & 0 \\ \frac{(1-r)}{2} & \frac{(1-r)}{2} & 0 & 0 & r & 0 \\ \frac{(1-r)}{2} & \frac{(1-r)}{2} & 0 & 0 & 0 & r \\ \frac{(1-r)}{2} & \frac{(1-r)}{2} & r & 0 & 0 & 0 \end{pmatrix}$$

$$\text{and } \mathbf{a}_b = (1 \quad -1 \quad 1 \quad -1 \quad 1 \quad -1).$$

Again, this matrix also uses the same variable r as in \mathbf{T}_a . The eigenvalues for this MCRP

which have non-zero weights are r and $re^{j\pi}$ with $\beta_k = .5$ for both eigenvalues.

Thus, in this case,

$$\phi_1(n) = .5[(re^{j\pi/2})^{|n|} + (re^{-j\pi/2})^{|n|}] \quad (\text{VI.16a})$$

$$\phi_2(n) = .5[(r)^{|n|} + (re^{j\pi})^{|n|}] \quad (\text{VI.16b})$$

so that

$$\begin{aligned} \phi_{\text{sum}}(n) &= \sum_{n=0}^{4-1} (.5)(re^{j2\pi k/4})^{|n|} \\ &= \begin{cases} 4r^{|n|} & \text{for } n = 0, \pm 4, \pm 8, \dots \\ 0 & \text{else} \end{cases} \end{aligned} \quad (\text{VI.17})$$

The choice of r is a compromise. As r increases, the power spectra $\Phi_1(\omega)$ and $\Phi_2(\omega)$ become more separated. However, it is also true that as r increases, the two processes become less complementary. Thus, there is a tradeoff as the processes need to be both spectrally disjoint and complementary. The choice of r shall be discussed later in conjunction with the probability of bit error. However, for now, suppose $r = 0.77$. The power spectrum for the resulting MCRPs are plotted in Figure 6.6. As can be seen, the two MCRPs are nearly complements with 7 dB between the largest and smallest magnitudes in the summed spectra. Again, it would be nice to have more separation, but this can be gained only at a loss of the processes' complementary status.

Now, let us see how the transmitted signal's power spectrum would look if these MCRPs were used. Recall that the power spectrum of the transmitted signal for DS-SS is given by

$$S_i(f) = \frac{E_b}{G_p} \{ \text{sinc}^2[(f - f_0)T_c] + \text{sinc}^2[(f + f_0)T_c] \}$$

whereas

$$\begin{aligned} S_i(f) &= \frac{E_b}{2G_p} \{ [\Phi_1(2\pi(f - f_0)T_c) + \Phi_2(2\pi(f - f_0)T_c)] \text{sinc}^2[(f - f_0)T_c] + \\ &\quad [\Phi_1(2\pi(f + f_0)T_c) + \Phi_2(2\pi(f + f_0)T_c)] \text{sinc}^2[(f + f_0)T_c] \} \end{aligned}$$

is the power spectrum for MCRP-SS. Both spectra are plotted in Figure 6.7. As can be seen, the MCRP-SS spectrum follows the DS-SS spectrum but contains spectral oscillations.

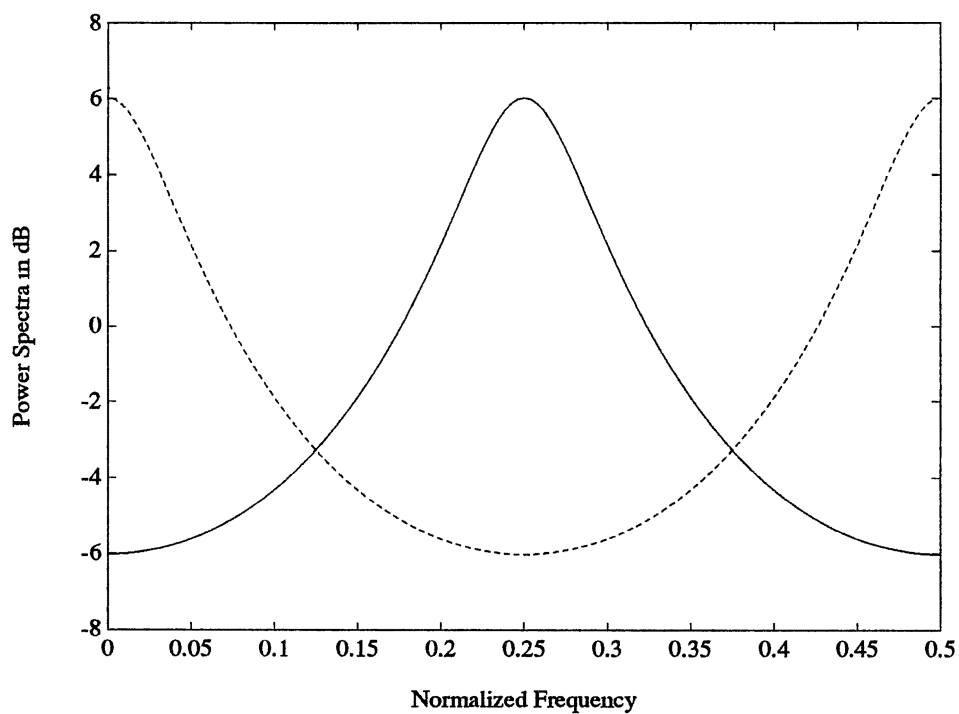


Figure 6.6. Power Spectra of Complementary MCRPs: $\Phi_1(\omega)$ (solid); $\Phi_2(\omega)$ (dashed); and $\Phi_{sum}(\omega)$ (dotted).

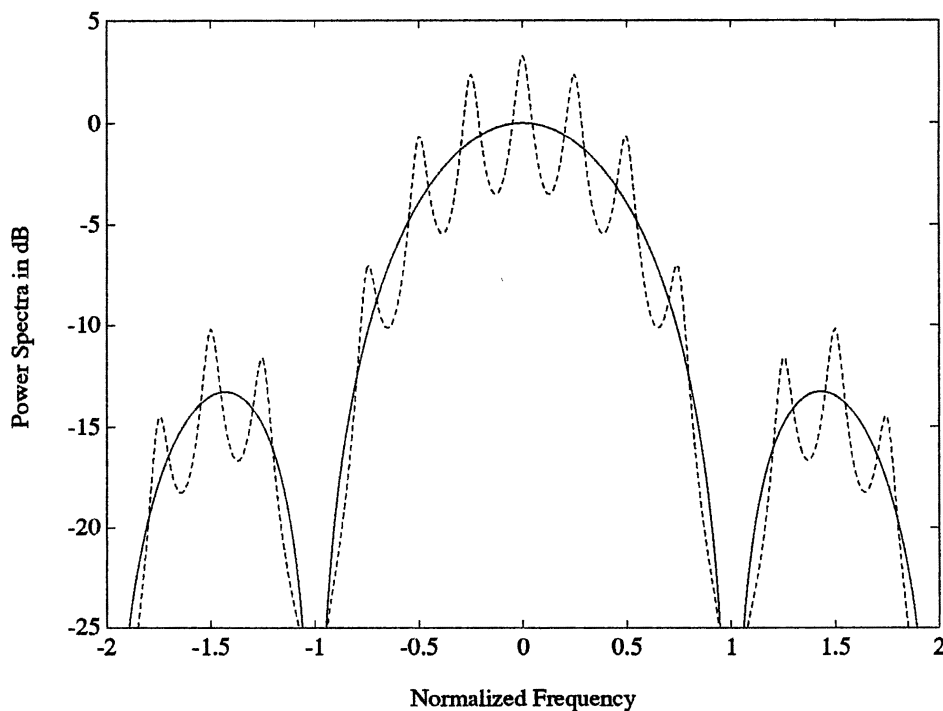


Figure 6.7. DS-SS (solid) and MCRP-SS (dashed) Power Spectra.

Effects of Partial-band Jamming

As depicted in Figures 6.1 and 6.3, it is common to model the communications channel between the transmitter and receiver as introducing additive noise to the transmitted signal. This noise can be categorized into two classes. The first is unintentional noise. This includes noises such as thermal noise and other types of nonhostile interference such as multipath noise and noise from other users in a multiple access environment. The second class of noise is intentional jamming.

There are several types of intentional jamming signals. These include barrage noise, partial-band, single tone, and pulsed noise [Zie85, Tor81]. Barrage noise is band-limited white noise over the bandwidth of the transmitted signal. Partial-band is similar

to barrage noise, but only covers a portion of the bandwidth. Single tone noise is just a signal at a fixed frequency, usually the center frequency of the transmission. Pulsed noise is noise that is "on" only a portion of the time.

The degradation of the communication channel depends on the communication scheme as well as the type of jamming noise. The detrimental effects of jamming include increased probability of bit error and degradation of synchronization between the transmitter and receiver. For the purpose of this research, the main interest is the effects of jamming on the probability of a bit error, P_b . In particular, the focus of this section is the effect of partial-band jamming. In the previous section, it was shown that the MCRP-SS scheme seems well-suited to deal with this type of jamming signal. Hence, the performance of this system in partial-band noise should provide a measure of the merits of this scheme.

Probability of Bit Error for DS-SS

For comparison and derivational purposes, let us first consider P_b for DS-SS with partial-band jamming. To begin, recall that the transmitted signal is $s_t(t) = \sqrt{E_b/T_d} d(t)c(t) \cos(2\pi f_0 t)$. The partial-band jamming noise shall be represented by $n_j(t)$. This noise shall be assumed to have power spectra density given by

$$S_{n_j}(f) = \begin{cases} \frac{N'_j}{2} & |f \pm f_0| < \frac{\rho W_{SS}}{2} \\ 0 & \text{else} \end{cases} \quad (\text{VI.18})$$

Here, W_{SS} is the bandwidth of the spread spectrum signal, which is given by $2/T_c$, and ρ is a fraction which determines how much of the bandwidth of the spread spectrum signal the jamming signal occupies. It shall be assumed that $\rho \ll 1$ since the jamming signal is partial-band. Additionally, assume $n_j(t)$ is zero-mean and has a Gaussian density. In

addition to the intentional jamming noise, it is common to also include a broadband noise term. Let $n_o(t)$ represent a white, Gaussian, zero-mean process with power spectrum $S_{n_o}(f) = N_o/2$.

At the receiver, the signal is despread and passed through a correlator. Based upon the output of the correlator, an estimate is made of $d(t)$. Of course, due to noise, $s_r(t)$ is not the only signal to reach the receiver: the background noise signal, $n_o(t)$, and the partial-band jamming signal, $n_j(t)$, are also received. Let us consider the effects of each of these in turn on the correlator.

The output of the correlator due to $s_r(t)$ will be

$$\begin{aligned} u_{s_r} &= \int_0^{T_d} s_r(t) c(t) \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) dt \\ &= \int_0^{T_d} \left\{ d(t) c(t) \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) \right\} c(t) \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) dt \\ &= \sqrt{E_b} d(t) \int_0^{T_d} c^2(t) \left\{ \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) \right\}^2 dt \end{aligned}$$

The data signal $d(t)$ is constant, being either 1 or -1 throughout the limits of integration.

The spreading signal is also ± 1 -valued, so $c^2(t) = 1$. Finally, note that the modulating and demodulating signal $\sqrt{2/T_d} \cos(2\pi f_0 t)$ has been normalized so that

$\int_0^{T_d} \left\{ \sqrt{2/T_d} \cos(2\pi f_0 t) \right\}^2 dt = 1$. Hence,

$$u_{s_r} = \sqrt{E_b} d(t). \quad (\text{VI.19})$$

Now consider the output of the correlator due to $n_o(t)$. First, consider the mean

$E[u_{n_o}]$:

$$\begin{aligned}
E[u_{n_o}] &= E\left[\int_0^{T_d} n_o(t)c(t)\sqrt{\frac{2}{T_d}}\cos(2\pi f_0 t)dt\right] \\
&= \int_0^{T_d} E[n_o(t)]c(t)\sqrt{\frac{2}{T_d}}\cos(2\pi f_0 t)dt \\
&= 0
\end{aligned}$$

since $u_{n_o}(t)$ is a zero-mean process. Next, consider the variance:

$$\begin{aligned}
E[u_{n_o}^2] &= E\left[\int_0^{T_d} n_o(t)c(t)\sqrt{\frac{2}{T_d}}\cos(2\pi f_0 t)dt \int_0^{T_d} n_o(\tau)c(\tau)\sqrt{\frac{2}{T_d}}\cos(2\pi f_0 \tau)d\tau\right] \\
&= \int_0^{T_d} \int_0^{T_d} E[n_o(t)n_o(\tau)]c(t)c(\tau)\sqrt{\frac{2}{T_d}}\cos(2\pi f_0 t)\sqrt{\frac{2}{T_d}}\cos(2\pi f_0 \tau)dt d\tau.
\end{aligned}$$

Since $n_o(t)$ is white, $E[n_o(t)n_o(\tau)] = (N_o/2)\delta(t - \tau)$. Thus,

$$\begin{aligned}
E[u_{n_o}^2] &= \frac{N_o}{2} \int_0^{T_d} c^2(t) \left\{ \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) \right\}^2 dt \\
&= \frac{N_o}{2}.
\end{aligned} \tag{VI.20}$$

As before, $c^2(t) = 1$ and the integral equals 1.

The jamming noise $n_j(t)$ may now be considered. Since this is also a zero-mean process, $E[u_{n_j}] = 0$. Next consider the variance:

$$\begin{aligned}
E[u_{n_j}^2] &= E\left[\int_0^{T_d} n_j(t)c(t)\sqrt{\frac{2}{T_d}}\cos(2\pi f_0 t)dt \int_0^{T_d} n_j(\tau)c(\tau)\sqrt{\frac{2}{T_d}}\cos(2\pi f_0 \tau)d\tau\right] \\
&= \int_0^{T_d} \int_0^{T_d} E[n_j(t)c(t)n_j(\tau)c(\tau)]\sqrt{\frac{2}{T_d}}\cos(2\pi f_0 t)\sqrt{\frac{2}{T_d}}\cos(2\pi f_0 \tau)dt d\tau.
\end{aligned} \tag{VI.21}$$

Now, since $n_j(t)$ and $c(t)$ are independent, $E[n_j(t)c(t)n_j(\tau)c(\tau)]$ equals

$E[n_j(t)n_j(\tau)]E[c(t)c(\tau)]$, which is just the product of the autocorrelations of $n_j(t)$ and $c(t)$.

Now, let us use the fact that multiplication in the time domain is convolution in the frequency domain. Thus, the power spectrum of $n_j(t)c(t)$ is given by $S_{n_j}(f)*S_c(f)$. Recall that the power spectra of $n_j(t)$ has been defined as $S_{n_j}(f) = N'_j/2$ for $|f \pm f_0| < \rho W_{SS}/2$. If $\rho \ll 1$, the bandwidth of $S_{n_j}(f)$ is much less than the bandwidth of $S_c(f)$, so $S_{n_j}(f)$ can be approximated by impulses at $\pm f_0$. Specifically,

$$S_{n_j}(f) \approx \frac{\rho W_{SS} N'_j}{2} \delta(f - f_0) + \frac{\rho W_{SS} N'_j}{2} \delta(f + f_0) \quad (\text{VI.22})$$

Recall that the power spectra of $c(t)$ is given by $T_c \text{sinc}^2(fT_c)$. Hence,

$$S_{n_j}(f)*(T_c \text{sinc}^2(fT_c)) = \frac{T_c \rho W_{SS} N'_j}{2} \text{sinc}^2((f - f_0)T_c) + \frac{T_c \rho W_{SS} N'_j}{2} \text{sinc}^2((f + f_0)T_c)$$

Before proceeding, let us simplify this expression a bit. First, it is common to define $N_j/2$ as the spectral density had the jammer applied his power evenly across the transmission bandwidth. The jammer's power is $P_{\text{jammer}} = \rho W_{SS} N'_j$ for partial-band jamming or $P_{\text{jammer}} = W_{SS} N_j$ for wideband jamming. Hence, $N_j = \rho N'_j$. Also, recall that $W_{SS} = 2/T_c$. Making these substitutions,

$$S_{n_j}(f)*(T_c \text{sinc}^2(fT_c)) = N_j \text{sinc}^2((f - f_0)T_c) + N_j \text{sinc}^2((f + f_0)T_c). \quad (\text{VI.23})$$

Now, we note that the resulting power spectrum of $n_j(t)c(t)$ as given above is relatively flat about $\pm f_0$. Hence, for the correlator, $n_j(t)c(t)$ can be approximated as white noise with a power spectral density of N_j . Therefore, $E[n_o(t)c(t)n_o(\tau)c(\tau)] \approx N_j \delta(t - \tau)$. Making this substitution in Equation (VI.21),

$$\begin{aligned}
E[u_{n_j}^2] &= \int_0^{T_d} \int_0^{T_d} N_j \delta(t - \tau) \sqrt{\frac{2}{T_d}} \cos(2\pi f_o t) \sqrt{\frac{2}{T_d}} \cos(2\pi f_o \tau) dt d\tau \\
&= N_j \int_0^{T_d} \left\{ \sqrt{\frac{2}{T_d}} \cos(2\pi f_o t) \right\}^2 dt \\
&= N_j.
\end{aligned} \tag{VI.24}$$

Again, the result follows from the fact that the integral evaluates to unity.

We may now combine these results. The mean of the output of the correlator is given by $E[u] = \sqrt{E_b} d(t)$ and the variance by $E[(u - E[u])^2] = N_o/2 + N_j$. Since the noises were assumed to have Gaussian densities, the output of the correlator will also have a Gaussian density and its mean will be $\sqrt{E_b} d(t)$ and variance will be $N_o/2 + N_j$. If $d(t) = 1$, the distribution is centered at $\sqrt{E_b}$ whereas if $d(t) = -1$, the distribution is centered at $-\sqrt{E_b}$.

To minimize the probability of error, estimate $d(t)$ by $\hat{d}(t) = 1$ if $u > 0$ and $\hat{d}(t) = -1$ if $u < 0$. In this case, the probability of error is given by

$$\begin{aligned}
P_b &= P[u > 0 \mid d(t) = -1] \\
&= \int_0^{\infty} \frac{1}{\sqrt{2\pi(N_o/2 + N_j)}} e^{-\frac{1}{2} \left(\frac{x + \sqrt{E_b}}{\sqrt{N_o/2 + N_j}} \right)^2} dx
\end{aligned}$$

If a substitution of variables is made, namely $v = (x + \sqrt{E_b})/\sqrt{N_o/2 + N_j}$, the result can be expressed using the co-error function $Q(x)$ as

$$P_b = Q\left(\sqrt{\frac{2E_b}{N_o + 2N_j}}\right) \tag{VI.25}$$

$$\text{where } Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} dv \tag{VI.26}$$

As the argument of $Q(x)$ increases, $Q(x)$ decreases. Therefore, P_b becomes smaller as E_b increases and as N_o and N_j decrease, which makes sense.

Probability of Bit Error for MCRP-SS

Now consider P_b for MCRP-SS. Recall that the transmitted signal is given by

$$s_i(t) = \sqrt{\frac{E_b}{T_d}} d(t) c_1(t) \cos(2\pi f_0 t) + \sqrt{\frac{E_b}{T_d}} c_2(t) \cos(2\pi f_0 t + \phi)$$

As before, two types of noises are added in the communication channel: a wideband noise $n_o(t)$ with spectral density $N_o/2$; and a partial-band noise $n_j(t)$ with spectral density $N'_j/2$ for $|f \pm f_0| < \rho W_{SS}/2$, where $\rho \ll 1$ and $W_{SS} = 2/T_c$ is the bandwidth of the transmitted signal $s_i(t)$.

From Figure 6.3, the first function of the receiver is to subtract the noise which was added at the transmitter. It shall be assumed that the receiver is able to perform this function so that the noise added at the transmitter can be disregarded for this analysis. There are several reasons why this is a valid assumption. While it must be conceded that the receiver is not likely to be able to subtract this signal exactly, it should be able to do so fairly well. Additionally, the effects of the added noise can be reduced by having its modulation out of phase with the modulation of the data; that is, let $\phi = 90^\circ$. In this way, any portion of this noise which is not subtracted out will have a minimal impact on the correlator. Additionally, it is felt that the effects of the jamming noise will be much more significant than the effects of this complementary noise.

Ignoring this added noise, the signal at the receiver is $r(t) = s_1(t) + n_o(t) + n_j(t)$

where $s_1(t) = \sqrt{E_b/T_d} d(t) c_1(t) \cos(2\pi f_0 t)$. Again, the effects of each component of $r(t)$ shall be handled in turn. To begin, consider the correlator output to $s_1(t)$:

$$\begin{aligned}
u_{s_1} &= \int_0^{T_d} s_1(t) c_1(t) \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) dt \\
&= \int_0^{T_d} \left\{ \sqrt{\frac{E_b}{T_d}} d(t) c_1(t) \cos(2\pi f_0 t) \right\} c_1(t) \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) dt \\
&= \sqrt{\frac{E_b}{2}} d(t) \int_0^{T_d} c_1^2(t) \left\{ \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) \right\}^2 dt \\
&= \sqrt{\frac{E_b}{2}} d(t). \tag{VI.27}
\end{aligned}$$

As in the previous section, note that $c_1^2(t) = 1$ since it is ± 1 -valued, and integral in the last step equals one.

Now consider the effects of $n_o(t)$ on the correlator. First, consider the mean $E[u_{n_o}]$:

$$\begin{aligned}
E[u_{n_o}] &= E \left[\int_0^{T_d} n_o(t) c_1(t) \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) dt \right] \\
&= \int_0^{T_d} E[n_o(t)] c_1(t) \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) dt \\
&= 0
\end{aligned}$$

since $u_{n_o}(t)$ is assumed to be a zero-mean process. Next, consider the variance:

$$\begin{aligned}
E[u_{n_o}^2] &= E \left[\int_0^{T_d} n_o(t) c_1(t) \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) dt \int_0^{T_d} n_o(\tau) c_1(\tau) \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 \tau) d\tau \right] \\
&= \int_0^{T_d} \int_0^{T_d} E[n_o(t) n_o(\tau)] c_1(t) c_1(\tau) \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 \tau) dt d\tau.
\end{aligned}$$

Since $n_o(t)$ is white, $E[n_o(t) n_o(\tau)] = (N_o/2) \delta(t - \tau)$. Thus,

$$\begin{aligned}
E[u_{n_o}^2] &= \frac{N_o}{2} \int_0^{T_d} c_1^2(t) \left\{ \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) \right\}^2 dt \\
&= \frac{N_o}{2}.
\end{aligned} \tag{VI.28}$$

Once again, note that $c_1^2(t) = 1$ and the integral equals 1.

The jamming noise $n_j(t)$ may now be considered. Since this is also a zero-mean process, $E[u_{n_j}] = 0$. Next consider the variance:

$$\begin{aligned}
E[u_{n_j}^2] &= E \left[\int_0^{T_d} n_j(t) c_1(t) \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) dt \int_0^{T_d} n_j(\tau) c_1(\tau) \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 \tau) d\tau \right] \\
&= \int_0^{T_d} \int_0^{T_d} E[n_j(t) c_1(t) n_j(\tau) c_1(\tau)] \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 \tau) dt d\tau.
\end{aligned} \tag{VI.29}$$

Now, since $n_j(t)$ and $c_1(t)$ are independent, $E[n_j(t) c_1(t) n_j(\tau) c_1(\tau)]$ equals

$E[n_j(t) n_j(\tau)] E[c_1(t) c_1(\tau)]$, which is just the product of the autocorrelations of $n_j(t)$ and $c_1(t)$. Now, let us use the fact that multiplication in the time domain is convolution in the frequency domain. Thus, the power spectrum of $n_j(t) c_1(t)$ is given by $S_{n_j}(f) * S_{c_1}(f)$.

Recall that the power spectra of $n_j(t)$ has been defined as $S_{n_j}(f) = N'_j/2$ for

$|f \pm f_0| < \rho W_{SS}/2$. If $\rho \ll 1$, the bandwidth of $S_{n_j}(f)$ is much less than the bandwidth of $S_{c_1}(f)$, so $S_{n_j}(f)$ can be approximated by impulses at $\pm f_0$. Specifically,

$$S_{n_j}(f) \approx \frac{\rho W_{SS} N'_j}{2} \delta(f - f_0) + \frac{\rho W_{SS} N'_j}{2} \delta(f + f_0)$$

Recall that the power spectra of $c_1(t)$ is given by $T_c \text{sinc}^2(f T_c) \Phi_1(2\pi f T_c)$ where $\Phi_1(\omega)$ is

the power spectrum of the discrete-time process from which $c_1(t)$ is obtained. Then,

$$S_{n_j}(f)*[T_c \text{sinc}^2(fT_c)\Phi_1(2\pi fT_c)] = \frac{T_c \rho W_{SS} N'_j}{2} \text{sinc}^2((f-f_0)T_c)\Phi_1(2\pi[f-f_0]T_c) + \\ \frac{T_c \rho W_{SS} N'_j}{2} \text{sinc}^2((f+f_0)T_c)\Phi_1(2\pi[f+f_0]T_c)$$

As in the previous section, this expression can be simplified by substituting $N_j = \rho N'_j$, and $W_{SS} = 2/T_c$. Making these substitutions,

$$S_{n_j}(f)*(T_c \text{sinc}^2(fT_c)T_c)\Phi_1(2\pi fT_c) = N_j \text{sinc}^2((f-f_0)T_c)\Phi_1(2\pi[f-f_0]T_c) + \\ N_j \text{sinc}^2((f+f_0)T_c)\Phi_1(2\pi[f+f_0]T_c).$$

Again, note that the resulting psd of $n_j(t)c_1(t)$ as given above is flat about $\pm f_0$, at which points the power spectral density is $N_j\Phi_1(0)$. Hence, for the correlator, $n_j(t)c_1(t)$ can be approximated as white noise with a power spectral density of $N_j\Phi_1(0)$. Therefore, $E[n_o(t)c(t)n_o(\tau)c(\tau)] \approx N_j\Phi_1(0)\delta(t-\tau)$. Making this substitution in Equation (VI.29),

$$E[u_{n_j}^2] = \int_0^{T_d} \int_0^{T_d} N_j\Phi_1(0)\delta(t-\tau) \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 \tau) dt d\tau \\ = N_j\Phi_1(0) \int_0^{T_d} \left\{ \sqrt{\frac{2}{T_d}} \cos(2\pi f_0 t) \right\}^2 dt \\ = N_j\Phi_1(0). \tag{VI.30}$$

Again, the integral evaluates to unity. In this case, the choice of spreading signal $c_1(t)$, and more importantly the value of its power spectrum at $\omega = 0$, clearly influences the effect of the jamming noise on the correlator.

The above results may now be combined. The mean of the output of the correlator is given by $E[u] = \sqrt{E_b/2} d(t)$ and the variance by $E[(u - E[u])^2] = N_o/2 + N_j\Phi_1(0)$. Since the noises were assumed to have Gaussian densities, the output of the correlator will also have a Gaussian density and will have mean $\sqrt{E_b/2} d(t)$ and variance $N_o/2 + N_j\Phi_1(0)$.

As in the previous section, to minimize the probability of error, estimate $d(t)$ by $\hat{d}(t) = 1$ if $u > 0$ and $\hat{d}(t) = -1$ if $u < 0$. For this case, the probability of error is given by

$$P_b = P[u > 0 \mid d(t) = -1]$$

$$= \int_0^{\infty} \frac{1}{\sqrt{2\pi(N_o/2 + N_j\Phi_1(0))}} e^{-\frac{1}{2}\left(\frac{x + \sqrt{E_b/2}}{\sqrt{N_o/2 + N_j\Phi_1(0)}}\right)^2} dx$$

If a substitution of variables is made, namely $v = (x + \sqrt{E_b/2})/\sqrt{N_o/2 + N_j\Phi_1(0)}$, the result can be expressed using the co-error function $Q(x)$ as

$$P_b = Q\left(\sqrt{\frac{E_b}{N_o + 2N_j\Phi_1(0)}}\right) \quad (\text{VI.31})$$

Again, note that the argument of $Q(x)$ increases, $Q(x)$ decreases. Therefore, P_b becomes smaller as E_b increases and as N_o and N_j decrease, which makes sense. Also, P_b becomes smaller as $\Phi_1(0)$ decreases.

Comparative Results

Let us now compare the resulting probabilities of bit error. These are given by

$$P_b = Q\left(\sqrt{\frac{2E_b}{N_o + 2N_j}}\right)$$

for DS-SS, and

$$P_b = Q\left(\sqrt{\frac{E_b}{N_o + 2N_j\Phi_1(0)}}\right)$$

for MCRP-SS. There are clearly two differences in the two expressions. The first is that the DS-SS expression has an extra factor of 2 in the numerator of the radical. The lack of this term in the MCRP-SS expression follows from the fact that only half of the power has been devoted to the data signal; the remaining power was dedicated to the additive noise. Clearly, this factor favors the DS-SS expression over the MCRP-SS one. The

other difference is the inclusion of the $\Phi_1(0)$ term in the MCRP-SS expression. If $\Phi_1(0)$ can be made small enough, then the effects of the jamming noise can be minimized. To learn more about $\Phi_1(0)$, let us return to the discussion of complementary MCRPs.

For the MCRP which produces $c_1(t)$ the autocorrelation was given by

$$\phi_1(n) = 0.5[(re^{j\pi/2})^{|n|} + (re^{-j\pi/2})^{|n|}]$$

from which it is possible to express the power spectrum as

$$\Phi_1(\omega) = \frac{1 - r^4}{1 + 2r^2 \cos(2\omega) + r^4} \quad (\text{VI.32})$$

Hence,

$$\Phi_1(0) = \frac{1 - r^2}{1 + r^2} \quad (\text{VI.33})$$

Thus, we see that $\Phi_1(0)$ depends upon the choice of r : as r increases, $\Phi_1(0)$ decreases.

From the MCRP-SS equation for P_b it is clear that as $\Phi_1(0)$ decreases, P_b decreases. It follows that as r increases, P_b decreases. Additionally, it has already been noted that r influences the separation and degree of complement of $\Phi_1(\omega)$ and $\Phi_2(\omega)$. To summarize, as r increases, the separation and probability of bit error improve; however, at the same time, $\Phi_1(\omega)$ and $\Phi_2(\omega)$ become less complementary.

This tradeoff in the choice of r is illustrated below in Figure 6.8. The figure plots the spectral magnitude variation in $\Phi_{\text{sum}}(\omega) = \Phi_1(\omega) + \Phi_2(\omega)$ versus $\Phi_1(0)$. The spectral magnitude variation shows how well $\Phi_1(\omega)$ and $\Phi_2(\omega)$ are complementary, and should be as small as possible. However, $\Phi_1(0)$ is a measure of P_b , and should be as small as possible, also. Hence, one must choose a value of $\Phi_1(0)$ which falls near the knee of the curve. A reasonable choice is $\Phi_1(0) = 0.25$, for which the spectral oscillations in $\Phi_{\text{sum}}(\omega)$ are about 7 dB.

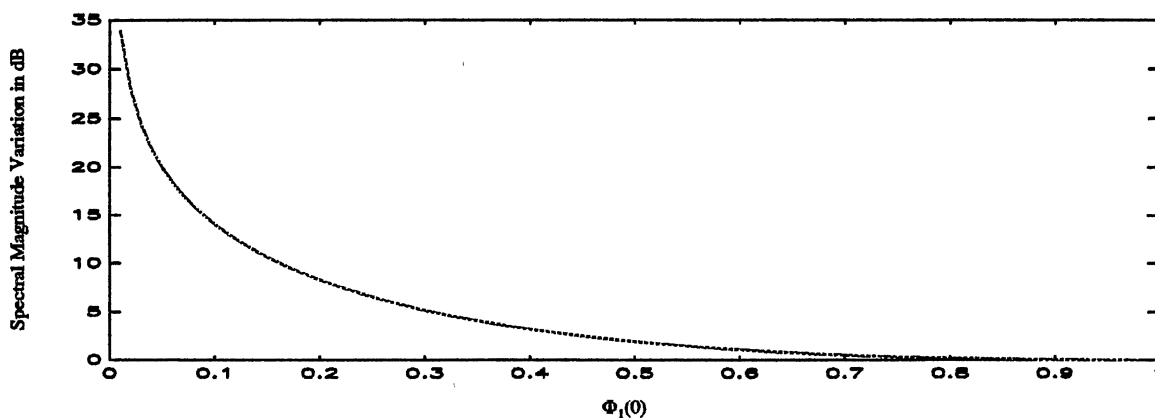


Figure 6.8. Spectral Magnitude Variation Versus $\Phi_1(0)$.

Plots comparing P_b for DS-SS and MCRP-SS with partial-band jamming are given below in Figures 6.9 through 6.11. Each plot represents a different ratio of E_b/N_o , which is a measure of the signal-to-noise ratio in the absence of jamming. In each graph, P_b is plotted versus E_b/N_j on logarithmic scales. Each plot shows four different curves. The first of these is P_b for DS-SS, which is the solid line. There are three P_b curves for MCRP-SS, representing $\Phi_1(0) = 0.3$, $\Phi_1(0) = 0.25$, and $\Phi_1(0) = 0.2$; these curves are dashed, dotted, and dot-dashed, respectively.

From the plots, it is clear that MCRP-SS has a lower probability of bit error than DS-SS when E_b/N_j is sufficiently small. Thus, when the jamming noise is sufficiently large in comparison to the signal power, the MCRP-SS scheme improves the bit error performance of the system. When there is relatively less jamming noise, the MCRP-SS does not improve the performance of the system, and in fact degrades it; in this case, the power devoted to the added noise at the transmitter is wasted.

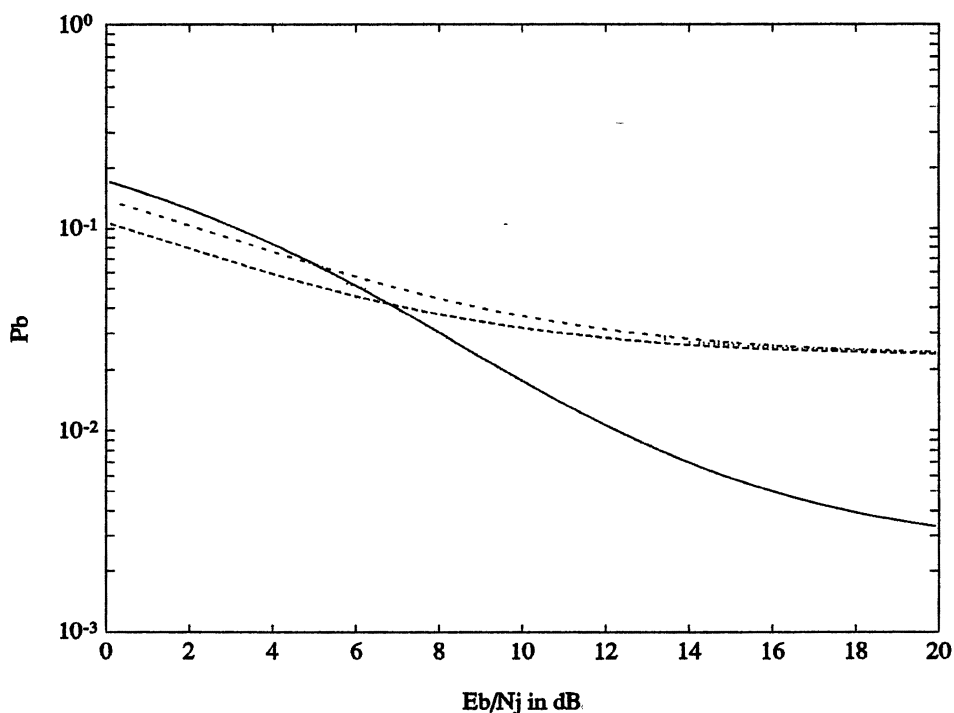


Figure 6.9. P_b for DS-SS and MCRP-SS with $E_b/N_o = 6\text{dB}$.

The results for different values of $\Phi_1(0)$ once again illustrate the tradeoff previously mentioned. A smaller value of $\Phi_1(0)$ decreases P_b but also increases the spectral variation of the transmitted signal, which is undesirable. However, the plots indicate that P_b does not vary greatly with values of $\Phi_1(0)$ near the knee of the tradeoff curve in Figure 6.8. Overall, $\Phi_1(0) = 0.25$ appears to be a good choice.

To summarize, this chapter has shown that MCRPs can be useful in a spread-spectrum communications scheme which reduces the effects of jamming. By unevenly spreading the data signal and then adding complementary noise to "fill out" the spectrum, it is possible for the jamming signal to miss the frequencies in which the data is most heavily represented. For the case of partial-band jamming it was shown that, under cer-

tain conditions, this scheme delivers improved bit error probability performance over DS-SS. Further improvements to this communication scheme rely on locating MCRPs which have a higher degree of separability while retaining a complementary nature.

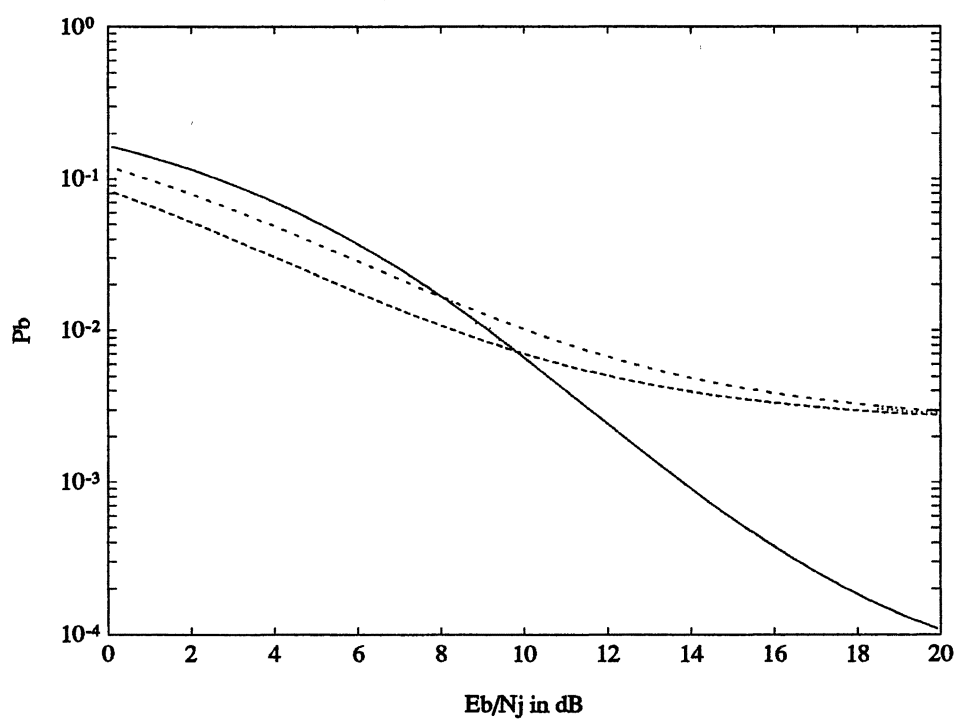


Figure 6.10. P_b for DS-SS and MCRP-SS with $E_b/N_o = 9$ dB.

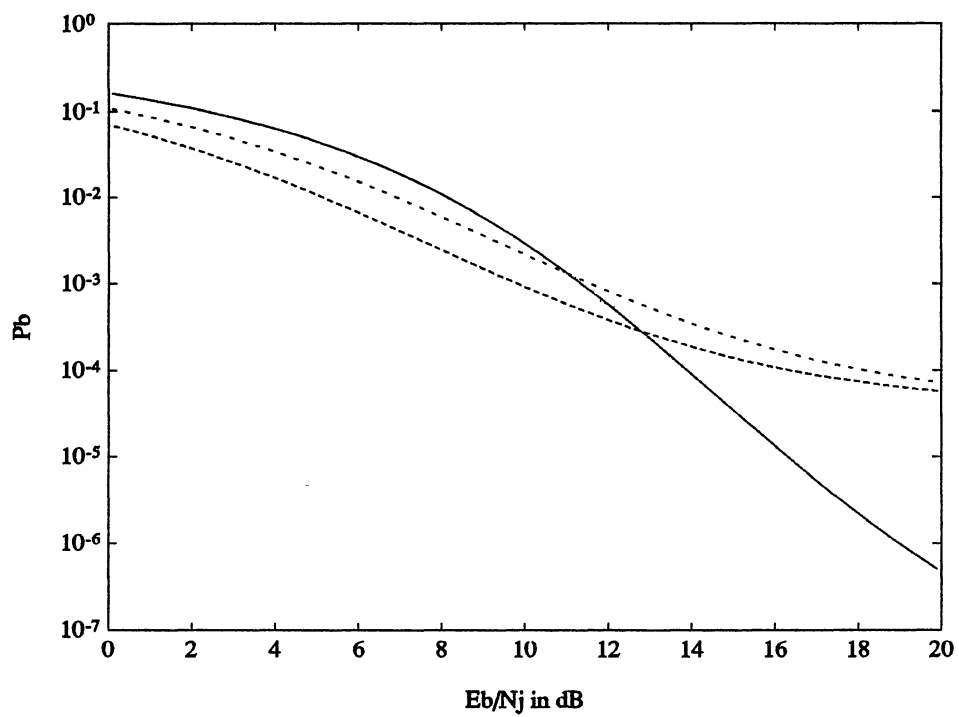


Figure 6.11. P_b for DS-SS and MCRP-SS with $E_b/N_o = 12$ dB.

CHAPTER VII

MARKOV RANDOM FIELDS

Although it is common to consider the Markov property in terms of random processes, other interpretations are possible. For example, suppose instead of a time index one has a spatial index. In addition, the problem under consideration could be a multidimensional case. As a specific example, one could think of a sampled image; i.e., a set of pixels. Here, the Markov property could be used to indicate that each pixel is directly influenced only by the pixels closest to it.

However, there is an important distinction between these spatial-domain and time-domain processes. The concept of causality makes sense for a time-domain process; the future cannot influence the present. However, a similar interpretation does not usually hold in the spatial-domain. There is no notion of which pixel came first and which preceded or followed it; any ordering of the pixels is inherently arbitrary and does not express a causal relationship. Instead, the pixels must be thought of as a whole, as a group. The necessarily different basis of spatial processes must be handled somewhat differently, giving rise to what has been called Markov random fields (MRFs).

Markov random fields have been applied to a range of spatial processes [Kin80]. For images, Markov random fields have been used as models [Kas81, Hac87]. As such, they have proven useful in image segmentation, image restoration, and synthesis of textured images. In addition, MRFs have been applied in other areas such as formant tracking in speech applications [Wil90].

The objective of this chapter is to briefly explore the extension of MCRPs to two dimensions. In fact, there are several ways to approach such an extension. One thought is to use the MCRP time series to construct an image on a row-by-row or column-by-column basis. A similar technique was used by Hershey, who used m-sequences to construct a library of two-tone images for the purpose of image compression [Her81b]. However, at this time, the author has chosen to pursue a somewhat different direction which is based on the concepts of Markov random fields. It should be stated that the goals of this chapter are quite modest. We wish only to show that is possible, and may be advantageous, to use MCRPs in image processing applications.

To begin, a little notation is necessary. Let L be a finite lattice; i.e., $L = \{(i, j): 1 \leq i \leq N_1, 1 \leq j \leq N_2\}$. These are the points over which the image is defined. It should be noted that it is possible to consider infinite lattices; however, the author shall not do so here. A *random field* (RF) X is a collection of spatially-oriented random variables and a probability measure which specifies the probability of occurrence of each member in the collection.

A *Markov random field* is a random field where the probability measure is determined by a set of probabilities over neighborhoods of the lattice. To explain further, let $\eta_{i,j}$ denote a neighborhood about the point (i, j) . Neighborhoods must obey certain properties, among them that pixel (k, l) is in the neighborhood of pixel (i, j) , $\eta_{i,j}$, if and only if pixel (i, j) is in the neighborhood of pixel (k, l) , $\eta_{k,l}$. Also, by convention, the neighborhood of (i, j) , $\eta_{i,j}$, does not contain the pixel (i, j) . The two most common neighborhoods are

$$\eta_{i,j}^1 = \{(k, l): 0 < (i - k)^2 + (j - l)^2 \leq 1\} \quad (\text{VII.1})$$

$$\eta_{i,j}^2 = \{(k, l): 0 < (i - k)^2 + (j - l)^2 \leq 2\} \quad (\text{VII.2})$$

which are illustrated below in Figure 7.1.

The Markov property holds over the random field by the conditional probability equation

$$P[X_{k,l} = x_{k,l} | X_{i,j} = x_{i,j}, (i,j) \in L] = P[X_{k,l} = x_{k,l} | X_{i,j} = x_{i,j}, (i,j) \in \eta_{k,l}] \quad (\text{VII.3})$$

Simply stated, the value of pixel (i, j) depends only upon the pixels in its neighborhood.

Of course, similar definitions hold for higher dimensions and even for one dimension; however, MRFs are most commonly used in two-dimensional, image processing applications.

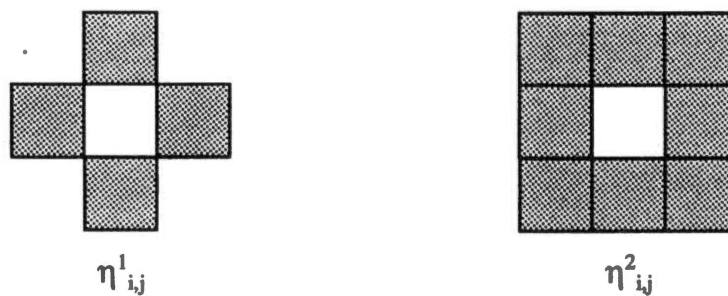


Figure 7.1. Neighborhoods $\eta_{i,j}^1$ and $\eta_{i,j}^2$.

Traditionally, pixel values in MRFs have been defined in terms of neighboring pixels by

$$x_{i,j} = \sum_{(k,l) \in \eta_{i,j}} \theta_{k,l} x_{k,l} + u_{i,j} \quad (\text{VII.4})$$

where $\{u_{i,j}\}$ is a collection of independent, identically distributed random variables. The above equation is, in fact, very similar to that for an autoregressive process. However, the above equation is not causal. For, $x_{i,j}$ is a function of the pixels in its neighborhood,

which are in turn functions of $x_{i,j}$. Because of this, Equation (VII.4) has been called simultaneous autoregressive (SAR), noncausal autoregressive (NCAR), and spatial autoregressive by various authors [Der89].

While MRFs have traditionally been defined by Equation (VII.4), the author sees no reason why MCRPs could not be used. To do so, one would have to extend MCRPs to higher dimensions much in the same manner that the autoregressive equation was extended. The advantage of using MCRPs is derived from the fact that MCRPs are by definition finite-valued processes. This seems ideally suited to work with images which are quantized to a relatively few number of pixel bits. In particular, each state of the MCRP could represent one level of quantization.

To see how this could be done, consider an MRF based on an MCRP which has N states. Thus, each pixel in the image can take on any one of the N values specified by the output map of the MCRP. The required "transition" probabilities would be of the form $P[x_{i,j} = a_k | \eta_{i,j}]$ where a_k is the output map of the MCRP associated with state k . If the neighborhood $\eta_{i,j}$ consists of p pixels, then there would be N^p possible configurations of neighborhoods for each k . Hence, there are $N \cdot N^p = N^{p+1}$ probabilities which must be specified. For example, if the chosen neighborhood is η^1 , then each pixel has four neighbors, so $p = 4$. For this situation, if N is 2, then $N^{p+1} = 32$. Of course, if there is any symmetry in the MRF then the number of unique probabilities will be much smaller than this.

As a simple example, suppose, as above, that η^1 is the neighborhood and $N = 2$. In this case, each pixel can occupy only one of two states, namely black or white. For this example, let us state that, if all of the neighbors about a given pixel are black (white), then there the probability is 1.0 that the pixel is also black (white). If three of the four neighbors are black (white), then there is a probability of 0.9 that the pixel is also black (white). The only other case is when two neighbors are black and two are white. In this

case, the probability is equal that the pixel is white or black. Note that introducing symmetry has greatly reduced the number of probabilities which needed to be defined. Also, based upon these probabilities, the pixels should have a tendency to clump into areas which are all black or all white.

The actual generation of an image based on these probabilities presents a bit of a problem. As noted previously, the pixel values have been defined as being noncausal. Hence, the image must be created as a whole; at the very least, pixel values cannot be determined one at a time. However, the author knows of no way to generate an MCRP-based MRF as a whole in a single step. Despite this, consider the following approach. Begin with a randomly generated image in which each pixel value is independent and generated from identically distributed random variables. Also, each pixel value of the image should belong to the output map of the MCRP. Now proceed through the image row by row, one pixel at a time. At each pixel, based on the neighbors and the transition probabilities, determine the new value of the pixel. For example, as defined above, if 3 of the 4 neighbors are black, then there is an 90% probability that the pixel will be changed to black; 10% of the time the pixel will be changed to white. After the pixel value has been updated, the procedure is repeated on the next pixel. This process is continued until every pixel has been updated.

For the example described above, an image of 100 pixels by 100 pixels was generated. This image and the original image on which it was based, are shown below in Figure 7.2. As can be seen, the black and white pixels are clumped together much more than in the original image. This is to be expected because of the way the transition probabilities were defined.

As a measure of the validity of this approach, let us estimate the transition probabilities based upon the resulting image. The relevant data is given below in Table 7.1. The estimates of the transition probabilities based on the resulting image are close to the probabilities given above, but could be better.

In an attempt to obtain more desirable results, the procedure was repeated. The resulting second-pass image is shown in Figure 7.3. Again, the black and white pixels have clustered together, even more so than before. Estimates of the transition probabilities, given in Table 7.2, are very close to the true probabilities. Thus, this MCRP MRF approach appears to be valid.

To summarize, because of their discrete-valued nature, MCRPs appear to be appropriate to model images, which are very often quantized to only a few bits. However, to do so involved extending MCRPs to two dimensions. As well, the inherently non-casual relationship of pixels needed to be accounted for. By adopting the key elements of Markov random fields, it was shown that an MCRP could be useful in image processing.



Figure 7.2. Original and First Pass MRF Images.

TABLE 7.1

ESTIMATE OF TRANSITION PROBABILITIES

Neighborhood	Number of White Pixels	Number of Black Pixels	Percentage of White Pixels	Percentage of Black Pixels
All White	1382	77	95	5
3 White	1714	444	80	20
2 of Each	1272	1216	51	49
3 Black	452	1859	20	80
All Black	83	1501	5	95

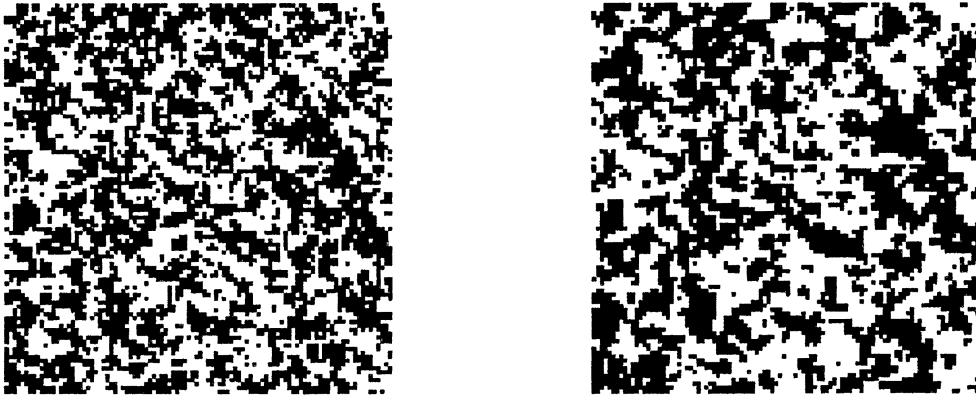


Figure 7.3. First and Second Pass MRF Images.

TABLE 7.2

ESTIMATE OF SECOND PASS TRANSITION PROBABILITIES

Neighborhood	Number of White Pixels	Number of Black Pixels	Percentage of White Pixels	Percentage of Black Pixels
All White	2018	38	98	2
3 White	1715	262	87	13
2 of Each	979	922	51	49
3 Black	235	1702	12	88
All Black	32	2097	2	98

CHAPTER VIII

CONCLUSIONS

Summary

By associating a real number with each state of a Markov chain, a new but related process, termed a Markov chain random process (MCRP), is created. These processes have some interesting properties: they are easy to implement, inherently stable, and can be made stationary and ergodic. As discrete-valued processes, they are of course ideally suited to model processes with discrete distributions. However, these discrete-valued processes also bear a strong relationship to the continuous-valued class of ARMA processes. Despite all of the interesting properties and potentially substantial applications of these processes, few researchers have studied them.

The objective of this research was to investigate this class of random processes in several stages. The first stage was to study the statistics of these processes. Next came an exploration of the synthesis behavior of these processes, i.e., how one can determine the MCRP parameters which result in desired statistics. Once synthesis had been studied, it was possible to investigate MCRPs as stochastic models. Finally, several applications of MCRPs were considered including speech modeling, spread spectrum communications, and a brief look at the extension of these processes to two dimensions.

As MCRPs are defined in terms of Markov chains, the first task of this research was to study Markov chains. The transition probabilities $\{t_{i,j}\}$, transition probability matrix T , and state probabilities $s_i(n)$ were examined. There are many classes of Markov chains,

and some are better suited for use in MCRPs than others. Since the main interest was stationary processes, stationary Markov chains were desired. Better yet, the class of regular Markov chains produce processes which are ergodic as well as stationary.

After reviewing Markov chains, a few statistics of MCRPs were derived. These included the mean, variance, autocorrelation, and power spectral density. Eigen decompositions were used to express the autocorrelation and power spectrum in terms of the eigenvalues of the transition probability matrix. Namely, if T is diagonalizable and the process has zero mean, from Equation (II.23) the autocorrelation is given by

$$\phi(n) = \sum_{k=1}^N \beta_k \lambda_k^{|n|} + \mu_y^2$$

and from Equation (II.34) the power spectrum is

$$\Phi(\omega) = \sum_{k=1}^N \beta_k \left(\frac{1}{1 - \lambda_k e^{j\omega}} + \frac{1}{1 - \lambda_k e^{-j\omega}} - 1 \right)$$

where $\{\lambda_k\}$ are the eigenvalues of T , and $\{\beta_k\}$ are constants which depend upon the eigenvectors of T and the real numbers associated with the states. From these equations, it was clear that an understanding of the eigenvalues of T was crucial to working with MCRPs.

Having found the statistics of MCRPs, the next question was how one could determine MCRP parameters to achieve given statistics. Three cases were considered: probability distribution synthesis; power spectrum synthesis; and synthesis given both probability distribution and power spectrum. For probability distribution synthesis, two cases were considered: continuous and discrete distributions. An exact match was given for processes with discrete probability distributions. For continuous distributions, a procedure was detailed in which an MCRP could approximate the continuous distribution.

Spectral synthesis was a much more difficult problem, and in fact has as yet no general solution. However, two new classes of MCRPs were proposed which offer partial solutions: a summed MCRP structure, and a ring MCRP structure. In both cases, processes with a few number of states were combined to form a larger process. Ring-structured MCRPs were thoroughly examined and a technique was outlined to determine an MCRP with a given set of eigenvalues and spectral coefficients; however, the limitation of this approach was that the spectral coefficients had to be real and positive.

The third synthesis case was when both probability distribution and power spectrum were specified. This, of course, is an even more difficult problem. However, a simple example was considered for a process with a Laplacian distribution and single complex eigenvalue-pair. For this case, a solution was found, though there was a restriction on the eigenvalue magnitude.

Chapter IV established a connection between ARMA processes and MCRPs. Although the two classes of processes are dissimilar in the time-domain, they nevertheless exhibit very similar spectral behavior. In particular, the poles of an ARMA process and the eigenvalues of a transition probability matrix play exactly the same role in the autocorrelation and power spectrum expressions.

To study this connection closer, the special situation of AR processes with a single complex pole-pair and MCRPs with a single complex eigenvalue-pair was studied. However, though the AR poles and MCRP eigenvalues were equated, there was a difference between the two processes' power spectra. The explanation was that, for known MCRP structures, the spectral coefficients must be real and positive; in contrast, for AR processes the corresponding coefficients are in general complex. The most noticeable consequence of this disparity was that the power spectra of AR processes and MCRPs experienced maxima at slightly different frequencies.

Despite this discrepancy, it was clear that an AR process could act as the basis for an MCRP stochastic model. The approach was to first establish an AR model of a given data set using known techniques. Next, the MCRP eigenvalues were equated to the poles of the AR filter. Finally, the MCRP spectral coefficients $\{\beta_k\}$ were determined by minimizing the difference between the AR and MCRP power spectra at certain frequencies; namely, the frequencies where the AR power spectra experienced extrema.

In Chapter V, this modeling approach was tested by applying it to speech modeling. There were several reasons why this application was chosen. First, there are many uses for models of speech, including storage and transmission of speech signals, speech recognition, and speaker identification. Although there has been much work in these areas, current solutions could use improvement. The use of MCRPs for speech was partially motivated by the success of hidden Markov models in speech processing. Additionally, it was felt that this application would provide a good test of the stochastic modeling techniques proposed in Chapter IV.

For comparison purposes with the MCRP speech model, a standard speech model known as LPC was chosen as a reference. In LPC, unvoiced speech is modeled as being the output of a AR filter driven by white noise; voiced speech is modeled as the output of an AR filter driven by an impulse train with a period corresponding to the pitch period of the voiced speech. In the MCRP model, unvoiced speech was modeled by an MCRP. For voiced speech, the model was an MCRP that had been passed through a comb filter. To test the MCRP model, several simple words were chosen. For each word, an MCRP model was created and synthesized speech was generated. Although there were differences in quality between the original and MCRP speech, it was clear that the modeling strategy was valid and that MCRPs could generate speech-like sounds.

The second application was to spread-spectrum communication. In this utilization, MCRPs were employed as spreading sequences in direct-sequence spread-spectrum communications. The specific approach involved unevenly spreading the data signal and then adding noise to "fill out" the spectrum. The idea was that a jamming signal might miss the frequencies where the data was most heavily represented.

One obstacle to this application was locating a pair of binary, ± 1 -valued MCRPs which were spectrally complementary and disjoint. That is, the sum of the two spectra is constant, and the two processes occupy different areas of the spectrum. Using the ring-structure of Chapter III, MCRPs were located which approximated this condition.

The probability of bit error for the MCRP spread spectrum system was derived for partial band jamming. It was shown that, under certain jamming conditions, this system had a lower probability of bit error than conventional direct-sequence spread spectrum.

The final topic of this research was an extension of MCRPs to two dimensions, and to image processing in particular. The reasoning behind this investigation stems from the fact that MCRPs are inherently discrete-valued and can assume only a few values. This seems particularly well-suited to model images, which are very often quantized to only a few bits. As the pixels in images are non-causally related, two-dimensional processes that are non-causal are required to properly model images. This realization led to an examination of Markov random fields. The key elements of MRFs were adopted to MCRPs, and a simple example was given which showed that MCRPs could be used in image processing applications.

Considerations for Future Research

There is still much research which can be performed on MCRPs, including areas directly related to this research. Although the statistics of MCRPs appears to be well understood, there remain unsolved problems in synthesis. For example, in spectral syn-

thesis, structures need to be found which can have complex spectral coefficients, β_k . While the author was able to accomplish much using real spectral coefficients, exact power spectra matches would have resulted if structures with complex spectral coefficients could have been found.

For stochastic modeling with MCRPs, it is worthwhile to pursue a modeling strategy which does not rely on the in-between step of finding an AR model; i.e., so that the MCRP model parameters could be found directly from the data. The only other strategy described was an estimate of the transition probabilities directly from the data. Barring a completely new method, perhaps a hybrid between these two approaches could be investigated.

For the MCRP speech model, voicing is a concern. There appeared to be some undesirable qualities present in the synthesized speech for voiced phonemes. Perhaps there is a better way to model voiced speech or a method whereby the offending artifacts can be eliminated. Also, for any practical implementation of an MCRP model, numerous details would have to be worked out. For example, phonemes that are difficult to model, such as voiced fricatives, would need to be studied. Additionally, the possibility of combining hidden Markov models and MCRP models of speech would perhaps be fruitful.

The MCRP spread spectrum approach too could benefit from some further research. In particular, better complementary MCRPs need to be found. If MCRPs could be located which achieve a better tradeoff between being complementary and being spectrally-disjoint, the performance of the system would improve. Also, it may be worthwhile to consider using an unbalanced system wherein more power is devoted to the data than to the added noise.

There is much research which could be done for MCRPs as applied to two dimensions. For one, statistics of non-causal processes derived from MCRPs need to be determined. Also, a procedure to synthesis images should be researched. As well, specific applications where such processes could be useful need to be identified.

REFERENCES

- [Agu88] Aguiar, R.J., M. Collares-Pereira, and J.P. Conde, "Simple Procedure for Generating Sequences of Daily Radiation Values Using a Library of Markov Transition Matrices," *Solar Energy*, Vol. 40, No. 3, pp. 269-279, 1988.
- [Bar64] Barnard, R.D., "On the Discrete Spectral Densities of Markov Pulse Trains," *The Bell System Technical Journal*, Vol. 43, pp. 233-259, January, 1964.
- [Bha60] Bharucha-Reid, A.T., *Elements of the Theory of Markov Processes and Their Applications*, McGraw-Hill. 1960.
- [Bro85] Brogan, W.L., *Modern Control Theory*, 2nd Edition, Prentice Hall, Englewood Cliffs, N.J., 1985.
- [Buc78] Buck, R.C., *Advanced Calculus*, McGraw-Hill, 1978.
- [Coo86] Cooper, G.R., and C.D. McGillem, *Modern Communications and Spread Spectrum*, McGraw-Hill, New York, 1986.
- [Cor87] Cormack, G.V., and R.N.S. Horspool, "Data Compression Using Dynamic Markov Modeling," *The Computer Journal*, Vol. 30, No. 6, 1987, pp. 541-550.
- [Dav79] Davis, P.J., *Circulant Matrices*, John Wiley & Sons, New York, 1979.
- [Dea89] Dean, R., *Global Maximum-Likelihood Decoding using Hidden Markov Models*, Doctoral Dissertation, Oklahoma State University, 1989.
- [Der89] Derin, H., and P.A. Kelly, "Discrete-Index Markov-Type Random Processes," *Proceedings of the I.E.E.E.*, Vol. 77, No. 10, October, 1989, pp. 1485-1510.
- [Deu71] Deutsch, E., and C. Zenger, "Inclusion Domains for the Eigenvalues of Stochastic Matrices," *Numerical Math.*, Volume 18, pp. 182-192, 1971.
- [ElJ88] El-Jaroudi, A., *Discrete Spectral Modeling with Application to Speech Analysis*, Doctoral Dissertation, Northeastern University, 1988.
- [Far87] Farges, E.P., *An Analysis-by-Synthesis Hidden Markov Model of Speech*, Doctoral Dissertation, Georgia Institute of Technology, November, 1987.
- [Gan64] Gantmacher, F.R., *The Theory of Matrices*, Chelsea Publishing Co., New York, 1964.
- [Hau87] Hacipasaoglu, K., *Segmentation of Noisy Images Using Nonstationary Markov Fields*, Master's Thesis, Naval Postgraduate School, December, 1987.

- [Her81a] Hershey, J.E., "Markov Filtering, Psychology and Spread Spectrum," Unpublished paper, 1981.
- [Her81b] Hershey, J.E., *Representation of Two-Tone Images Using Pseudorandomly Derived Square Matrices*, Doctoral Dissertation, Oklahoma State University, 1981.
- [Her86] Hershey, J.E. and R. Yarlagadda, *Data Transportation and Protection*, Plenum Press, 1986.
- [Hug57] Huggins, W.H., "Signal-Flow Graphs and Random Signal," *Proceedings of the I.R.E.*, Vol. 45, January, 1957, pp. 74-86.
- [Ios80] Iosifescu, M., *Finite Markov Processes and Their Applications*, John Wiley & Sons, 1980.
- [Jon86] Jones, D.I., and M.H. Lorenz, "An Application of a Markov Chain Noise Model to Wind Generator Simulation," *Mathematics and Computers in Simulation*, 28 (1986), pp. 391-402.
- [Kar51] Karpelevic, F.I., "On the Characteristic Roots of Matrices With Non-negative Elements," *Akademiia Nauk Izvestiia, Seriia Matematicheskaiia*, Vol. 15, 1951. pp. 361-383 (Russian).
- [Kas81] Kashyap, R.L., "Analysis and Synthesis of Image Patterns by Spatial Interaction Models," in *Progress in Pattern Recognition*, L.N. Kanal and A. Rosenfeld, Eds., North-Holland, New York, 1981, pp. 149-186.
- [Kaz82] Kazakos, D., "Statistical Discrimination Using Inaccurate Modeling," *IEEE Transactions on Information Theory*, September, 1982, pp. 720-728.
- [Kem76] Kemeny, J.G., and J.L. Snell, *Finite Markov Chains*, Springer-Verlag, 1976.
- [Kin80] Kinderman, R., and J.L. Snell, *Markov Random Fields and Their Applications*, American Mathematical Society, Providence, R.I., 1980.
- [Lan85] Lancaster, P., and M. Tismenetsky, *Theory of Matrices*, 2nd Edition, Academic Press, New York, 1985.
- [Mir64] Mirsky, L., "Inequalities and Existence Theorems in the Theory of Matrices," *Journal of Mathematical Analysis and Applications*, August 1964, pp. 99-118.
- [Mul72] Mullis, C.T., and K. Steiglitz, "Circulant Markov Chains as Digital Signal Sources," *IEEE Transactions on Audio and Electroacoustics*, October, 1972, pp. 246-248.
- [Mul79] Mullis, C.T., and R.A. Roberts, "On Weak Equivalence of Linear Systems and Finite State Systems," *SIAM Journal on Mathematical Analysis*, May, 1979, pp. 498-511.
- [Oli698] de Oliveira, G.N., "Sobre Matrices Estocasticas e Duplamente Estocasticas," *Rev. Fac. Ci. Univ. Coimbra*, Vol. 41 (1968), pp. 15-221 (Spanish).

- [Opp75] Oppenheim, A.V. and R.W. Schaffer, *Digital Signal Processing*, Prentice Hall, 1975.
- [Pap84] Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1984.
- [Per64] Perfect, H., and L. Mirsky, "Spectral Properties of Doubly-Stochastic Matrices," *Monatshefte Für Mathematik*, November, 1964, pp. 35-57.
- [Pic82] Pickholtz, R.L., D.L. Schilling, and L.B. Milstein, "Theory of Spread Spectrum Communications - A Tutorial," *IEEE Transactions on Communications*, May, 1982, pp. 855-884.
- [Rab78] Rabiner, L.R., and R.W. Schaffer, *Digital Processing of Speech Signals*, Prentice Hall, 1978.
- [Rab86] Rabiner, L.R., and B.H. Juang, "An Introduction to Hidden Markov Models," *IEEE ASSP Magazine*, January, 1989, pp. 4-16.
- [Rab89] Rabiner, L.R., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, Vol. 77, pp. 257-286, February, 1989.
- [Rob79] Robinson, E.A., and M.T. Silva, *Digital Foundations of Time-Series Analysis*, Vol. 1., Holden-Day, 1979.
- [Sam76] Sambur, M.R., and N.S. Jayant, "LPC Analysis/Synthesis from Speech Inputs Containing Quantizing Noise or Additive White Noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*. Vol., ASSP-24, pp. 488-494, December, 1976.
- [Sch90] Schilling, D.L., R.L. Pickholtz, and L.B. Milstein, "Spread Spectrum Goes Commercial," *IEEE Spectrum*, August, 1990, pp. 40-45.
- [Sch75] Schroeder, M.R., "Models of Hearing," *Proceedings of the IEEE*, September, 1975. pp. 1332-1350.
- [Sch84] Schroeder, M.R., "Linear Prediction, Entropy and Signal Analysis," *IEEE ASSP Magazine*, July 1984, pp. 3-11.
- [Sen81] Seneta, E., *Non-negative Matrices and Markov Chains*, Springer-Verlag, New York, 1981.
- [Sie76] Siegel, L.J., K. Steiglitz, and M. Zuckerman, "The Design of Markov Chains for Waveform Generation," *IEEE Transactions of Acoustics, Speech, and Signal Processing*, Vol. 24, December, 1976, pp. 558-562.
- [Sit54] Sittler, R.W., *Analysis and Design of Simple Nonlinear Noise Filters*, Doctoral Dissertation, Massachusetts Institute of Technology, 1954.
- [Sit56] Sittler, R.W., "Systems Analysis of Discrete Markov Processes," *IRE Transactions on Circuit Theory*, December, 1956, pp. 257-266.

- [Skl88] Sklar, B., *Digital Communications*, Prentice Hall, Englewood Cliffs, New Jersey, 1988.
- [Ste83] Steinhardt, A.O., *An Optimization Theoretic Framework for Spectral Estimation*, Doctoral Dissertation, University of Colorado at Boulder, 1983.
- [Son83] Sondhi, M.M., "Random Processes with Specified Spectral Density and First-Order Probability Density," *The Bell System Technical Journal*, Vol. 62, No. 3, March 1983, pp. 679-701.
- [Sul49] Suleimanova, H.R., "Stochastic Matrices with Real Characteristic Numbers," *Doklady Akad. Nauk SSSR N.S.*, Vol. 66, 1949, pp. 343-345 (Russian).
- [Tao90] Taori, R., G.D. Cain, and K.V. Lever, "Correlation Windowing and Spectral Biasing in the Design of Noise Generators Having Specified Probability Densities and Power Spectra," *Electronic Letters*, Vol. 26, No. 14, July 5, 1990, pp. 1041-1043.
- [Tea79] Teacher, C.F., "Performance of LPC Vocoder in a Noisy Environment," *Proc. Int. Conf. on Acoustics, Speech, and Signal Proc.*, 1979, pp. 216-219.
- [Tor81] Torrieri, D.J., *Principles of Military Communication Systems*, Artech House, Inc., Dedham, Massachusetts, 1981.
- [Wil90] Wilcox, L.D., and F.R. Chen, "Application of Markov Random Fields to Formant Extraction," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Albuquerque, NM, April, 1990. pp. 349-352.
- [Yos73] Yoshida, Y, "Method of Power Spectrum Calculation of Markov Signals," *Electronics and Communications in Japan*, Vol. 56-1, No. 9, September, 1973, pp. 64-73.
- [Zad57] Zadeh, L.A., "Signal Flow Graphs and Random Signal," Letters to the Editor, *Proceeding of the I.R.E.*, Vol. 45, January, 1957, pp. 44-86.
- [Zen72] Zenger, C., "A Comparison of Some Bounds for the Nontrivial Eigenvalues of Stochastic Matrices," *Numerical Math*, Vol. 18 (1971), pp. 209-211.
- [Zie85] Ziemer, R.E., and R.L. Peterson, *Digital Communications and Spread Spectrum Systems*, MacMillan Publishing Co., New York, 1985.

APPENDIX

DERIVATION OF SPECTRAL PARAMETERS FOR RING STRUCTURED MARKOV CHAIN RANDOM PROCESSES

In this appendix, the spectral parameters are derived for MCRPs with the ring structure described in Chapter III. Using block matrix notation, the transition probability matrix for this process can be expressed as

$$T = \begin{pmatrix} (1-\alpha_1)T_1 & \alpha_1 \mathbf{1}_{N_1} s_2^T & 0 & \cdot & \cdot & 0 \\ 0 & (1-\alpha_2)T_2 & \alpha_2 \mathbf{1}_{N_2} s_3^T & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & (1-\alpha_{M-1})T_{M-1} & \alpha_{M-1} \mathbf{1}_{N_{M-1}} s_M^T \\ \alpha_M \mathbf{1}_{N_M} s_1^T & 0 & 0 & \cdot & 0 & (1-\alpha_M)T_M \end{pmatrix} \quad (\text{III.12})$$

where $\mathbf{1}_{N_i}$ is a vector of length N_i , where every element is 1, T_i is the $N_i \times N_i$ transition probability matrix for the i^{th} subchain, s_i is the stationary distribution for T_i , and $0 < \alpha_i < 1$ for all i . The total number of states in T shall be designated by N : $N = \sum_i N_i$. The only other restriction is that each of the T_i stochastic matrices must be diagonalizable. The restriction that these matrices be irreducible is relaxed in this instance since T can be shown to be irreducible whether the T_i matrices are irreducible or not. However, if a given T_i is reducible, s_i will not be unique, and must be chosen with some care. Specifically, every element of s_i must be nonzero. Every element of s_i will be always be nonzero for irreducible T_i .

To continue, some notation from Chapter III needs to be reiterated. Let $\mathbf{a}_i = (a_{i,1} \ a_{i,2} \ \cdots \ a_{i,N_i})^T$ be the output map for the i^{th} subchain, so that $\mathbf{a} = (\mathbf{a}_1^T \ \mathbf{a}_2^T \ \cdots \ \mathbf{a}_M^T)^T$ for the entire chain. Let $\{\lambda_{i,k}\}$ for $0 \leq k \leq N_i - 1$ be the eigenvalues of T_i with associated right and left eigenvectors $\{\mathbf{v}_{i,k}\}$ and $\{\mathbf{u}_{i,k}\}$, respectively. The eigenvectors should be normalized so that $\mathbf{u}_{i,k}^T \mathbf{v}_{i,k} = 1$ for all i,k . By convention, $\lambda_{i,0} = 1$ for every T_i , and s_i is the left eigenvector of T_i associated with this eigenvalue: $s_i = \mathbf{u}_{i,0}$. The spectral coefficients for the eigenvalues $\{\lambda_{i,k}\}$ are given by $\beta_{i,k} = \mathbf{a}^T S_i \mathbf{v}_{i,k} \mathbf{u}_{i,k}^T \mathbf{a}^T$ where S_i is a diagonal matrix composed of the elements of s_i .

Using this notation, it will be shown that the stationary state probabilities for this structure can be expressed as

$$\mathbf{s} = \frac{1}{\sum_j \frac{1}{\alpha_j}} \left(\frac{1}{\alpha_1} s_1^T \quad \frac{1}{\alpha_2} s_2^T \quad \cdots \quad \frac{1}{\alpha_M} s_M^T \right)^T \quad (\text{A.2})$$

Also, assuming $\mu_i = s_i^T \mathbf{a} = 0$ for all i , the eigenvalues of T which can have non-zero spectral coefficients are $\{\lambda'_{i,k}\}$ where

$$\lambda'_{i,k} = (1 - \alpha_i) \lambda_{i,k} \quad \text{for } 1 \leq i \leq M, 0 < k < N_i \quad (\text{A.3})$$

and their associated coefficients are $\{\beta'_{i,k}\}$

$$\beta'_{i,k} = \frac{\beta_{i,k}}{\alpha_i \sum_j \frac{1}{\alpha_j}} \quad \text{for } 1 \leq i \leq M, 0 < k < N_i. \quad (\text{A.4})$$

Now consider the proof of equation (A.2). First note that \mathbf{s} in (A.2) is a valid stationary distribution as all of its elements are positive. Also, the scale factor $1/(\sum_j [1/\alpha_j])$ is included to insure that the elements of \mathbf{s} sum to unity. To show that \mathbf{s} is the stationary state probability vector of T , it must be shown that $\mathbf{s}^T T = \mathbf{s}^T$. Consider the product $\mathbf{s}^T T$:

$$\frac{1}{\sum_j \frac{1}{\alpha_j}} \begin{pmatrix} \frac{1}{\alpha_1} s_1^T & \frac{1}{\alpha_2} s_2^T & \cdots & \frac{1}{\alpha_M} s_M^T \end{pmatrix} \begin{pmatrix} (1-\alpha_1)T_1 & \alpha_1 \mathbf{1}_{N_1} s_2^T & & \\ & (1-\alpha_2)T_2 & \alpha_2 \mathbf{1}_{N_2} s_3^T & \\ & & \ddots & \\ \alpha_M \mathbf{1}_{N_M} s_1^T & & & (1-\alpha_M)T_M \end{pmatrix}$$

The first N_1 elements of the resulting row vector are given by

$$\begin{aligned} & \frac{1}{\sum_j \frac{1}{\alpha_j}} \left(\frac{1-\alpha_1}{\alpha_1} s_1^T T_1 + \frac{\alpha_M}{\alpha_M} s_M^T \mathbf{1}_{N_M} s_1^T \right) \\ &= \frac{1}{\sum_j \frac{1}{\alpha_j}} \left(\frac{1-\alpha_1}{\alpha_1} s_1^T + s_1^T \right) \\ &= \frac{1}{\sum_j \frac{1}{\alpha_j}} \left(\frac{1}{\alpha_1} s_1^T \right) \end{aligned}$$

since $s_1^T T_1 = s_1^T$ and $s_M^T \mathbf{1}_{N_M} = 1$. By proceeding to consider subsequent groups of N_i elements, the result is that

$$\begin{aligned} s^T T &= \frac{1}{\sum_j \frac{1}{\alpha_j}} \begin{pmatrix} \frac{1}{\alpha_1} s_1^T & \frac{1}{\alpha_2} s_2^T & \cdots & \frac{1}{\alpha_M} s_M^T \end{pmatrix} \\ &= s^T \end{aligned}$$

as was claimed. Also, recall from Chapter III that S was defined as the diagonal matrix of elements of s . Thus

$$S = \frac{1}{\sum_j \frac{1}{\alpha_j}} \begin{pmatrix} \frac{1}{\alpha_1} S_1 & & & & & & \\ & \frac{1}{\alpha_2} S_2 & & & & & \\ & & \ddots & & & & \\ & & & \ddots & & & \\ & & & & \ddots & & \\ & & & & & \ddots & \\ & & & & & & \frac{1}{\alpha_M} S_M \end{pmatrix}. \quad (\text{A.5})$$

Now consider the eigenvalues of T . Some of T 's eigenvalues are derived from the eigenvalues of the T_i matrices. For example, consider T_1 . Let $\{v_{1,k}\}$ be the right eigenvectors of T_1 associated with the eigenvalues $\{\lambda_{1,k}\}$ for $0 < k < N_1$. Next, let $\{v'_{1,k}\}$ be vectors obtained by appending $N - N_1$ zeros to $v_{1,k}$: $v'_{1,k} = (v_{1,k}^T \quad 0 \quad \cdots \quad 0)^T$ in block vector notation. Now, consider $v'_{1,k}$ for any $0 < k < N_1$ as an eigenvector of T :

$$T v'_{1,k} = \begin{pmatrix} (1 - \alpha_1) T_1 & \alpha_1 \mathbf{1}_{N_1} s_1^T & & & & & \\ & (1 - \alpha_2) T_2 & \alpha_2 \mathbf{1}_{N_2} s_2^T & & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & & & \ddots & \\ \alpha_M \mathbf{1}_{N_M} s_M^T & & & & & & (1 - \alpha_M) T_M \end{pmatrix} \begin{pmatrix} v_{1,k} \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}$$

The first N_1 elements of the resulting column vector can be represented as $(1 - \alpha_1) T_1 v_{1,k}$.

But $T_1 v_{1,k} = \lambda_{1,k} v_{1,k}$, so the first N_1 elements are $(1 - \alpha_1) \lambda_{1,k} v_{1,k}$. Now, recall that $s_1 = u_{1,0}$, where $u_{1,0}$ is the left eigenvector of T_1 associated with the eigenvalue $\lambda_{1,0} = 1$. In general, for any diagonalizable matrix A , the right eigenvectors $\{v_i\}$ and left eigenvectors $\{u_i\}$ must satisfy $u_i^T v_j = 0$ for $i \neq j$ [Lan85]. Hence, since $s_1 = u_{1,0}$, $s_1^T v_{1,k} = 0$ for $0 < k < N_1$. Thus, the last N_M elements of $T v'_{1,k}$ are zero. The remaining elements of $T v'_{1,k}$ are trivially zero.

Thus,

$$\begin{aligned}
T\mathbf{v}'_{1,k} &= (1 - \alpha_1)\lambda_{1,k} \begin{pmatrix} \mathbf{v}_{1,k} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \\
&= (1 - \alpha_1)\lambda_{1,k}\mathbf{v}'_{1,k}
\end{aligned}$$

proving that $\{\mathbf{v}'_{1,k}\}$ are eigenvectors of T associated with the eigenvalues $\{(1 - \alpha_1)\lambda_{1,k}\}$ for $0 < k < N_1$. Similarly, it can be shown that, if $\{\mathbf{u}_{1,k}\}$ are eigenvectors of T_1 , then the left eigenvectors of T associated with $\{(1 - \alpha_1)\lambda_{1,k}\}$ for $0 < k < N_1$ are $\{\mathbf{u}'_{1,k}\}$ where $\mathbf{u}'_{1,k} = (\mathbf{u}_{1,k}^T \ 0 \ \cdots \ 0)^T$ using block vector notation.

Next, note that the numbering of the subchains is arbitrary. That is, any subchain could be labeled as the first subchain as long as the remaining subchains are renumbered accordingly. The consequence of this observation is that, if a result can be shown for one subchain, a similar result is true for all of the other subchains. Thus, it follows that $\{(1 - \alpha_i)\lambda_{i,k}\}$ are eigenvalues of T for $0 < k < N_i$, $1 \leq i \leq M$ which confirms equation (A.3).

The spectral coefficients of these eigenvalues can now be found. Recall that, from Chapter II, for an eigenvalue λ of T with right (left) eigenvector \mathbf{v} (\mathbf{u}) such that $\mathbf{u}^T \mathbf{v} = 1$, the coefficient is $\beta = \tilde{\mathbf{a}}^T S \mathbf{v} \mathbf{u}^T \tilde{\mathbf{a}}$. For the ring structure, the mean of the process is $\boldsymbol{\mu} = \mathbf{s}^T \mathbf{a} = [1/(\sum_j 1/\alpha_j)] \sum_i (1/\alpha_i) s_i^T \mathbf{a}_i$. Assuming each subchain is a zero-mean process, $\boldsymbol{\mu}_i = s_i^T \mathbf{a}_i = 0$, and hence, $\boldsymbol{\mu} = 0$. Then, $\tilde{\mathbf{a}} = \mathbf{a}$. As before, consider the eigenvalues $\{(1 - \alpha_1)\lambda_{1,k}\}$ for $0 < k < N_1$. The associated right eigenvectors of T were shown to be given by $\mathbf{v}'_{1,k} = (\mathbf{v}_{1,k}^T \ 0 \ \cdots \ 0)^T$ and the left eigenvectors were given by $\mathbf{u}'_{1,k} = (\mathbf{u}_{1,k}^T \ 0 \ \cdots \ 0)^T$. Note that

$$\begin{aligned}
 (\mathbf{u}'_{1,k})^T \mathbf{v}'_{1,k} &= (\mathbf{u}_{1,k}^T \quad 0 \quad \cdots \quad 0) \begin{pmatrix} \mathbf{v}_{1,k} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \\
 &= \mathbf{u}_{1,k}^T \mathbf{v}_{1,k}.
 \end{aligned}$$

By a previous assumption, $\mathbf{u}_{1,k}^T \mathbf{v}_{1,k} = 1$ for all k so the condition $(\mathbf{u}'_{1,k})^T \mathbf{v}'_{1,k} = 1$ is met.

For the eigenvalues $\{\lambda'_{1,k}\}$ for $0 < k < N_1$, consider the spectral coefficients $\{\beta'_{1,k}\}$ which are given by $\beta'_{1,k} = \mathbf{a}^T \mathbf{S} \mathbf{v}'_{1,k} (\mathbf{u}'_{1,k})^T \mathbf{a}$. Let us split this expression into two parts and first consider $(\mathbf{u}'_{1,k})^T \mathbf{a}$:

$$\begin{aligned}
 (\mathbf{u}'_{1,k})^T \mathbf{a} &= (\mathbf{u}_{1,k}^T \quad 0 \quad \cdots \quad 0) \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \cdot \\ \cdot \\ \mathbf{a}_M \end{pmatrix} \\
 &= \mathbf{u}_{1,k}^T \mathbf{a}_1
 \end{aligned}$$

Next consider $\mathbf{a}^T \mathbf{S} \mathbf{v}'_{1,k}$:

$$\begin{aligned}
 \mathbf{a}^T \mathbf{S} \mathbf{v}'_{1,k} &= \frac{\mathbf{a}^T}{\sum_j \frac{1}{\alpha_j}} \begin{pmatrix} \frac{1}{\alpha_1} S_1 & & & & \\ & \frac{1}{\alpha_2} S_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \frac{1}{\alpha_M} S_M \end{pmatrix} \begin{pmatrix} v_{1,k} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \\
 &= \frac{\mathbf{a}^T}{\sum_j \frac{1}{\alpha_j}} \begin{pmatrix} \frac{1}{\alpha_1} S_1 v_{1,k} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}
 \end{aligned}$$

Expanding \mathbf{a} yields

$$\begin{aligned}
 \mathbf{a}^T \mathbf{S} \mathbf{v}'_{1,k} &= \frac{1}{\sum_j \frac{1}{\alpha_j}} (\mathbf{a}_1^T \quad \mathbf{a}_2^T \quad \cdots \quad \mathbf{a}_M^T) \begin{pmatrix} \frac{1}{\alpha_1} S_1 v_{1,k} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \\
 &= \frac{\mathbf{a}_1^T S_1 v_{1,k}}{\alpha_1 \sum_j \frac{1}{\alpha_j}}
 \end{aligned}$$

Combining these parts,

$$\begin{aligned}\beta'_{i,k} &= \mathbf{a}^T \mathbf{S} \mathbf{v}'_{1,k} (\mathbf{u}'_{1,k})^T \mathbf{a} \\ &= \frac{\mathbf{a}_1^T \mathbf{S}_1 \mathbf{v}_{1,k} \mathbf{u}_{1,k}^T \mathbf{a}_1}{\alpha_1 \sum_j \frac{1}{\alpha_j}}\end{aligned}$$

But $\beta_{i,k} = \mathbf{a}_i^T \mathbf{S}_i \mathbf{v}_{i,k} \mathbf{u}_{i,k}^T \mathbf{a}_i$. Thus

$$\beta'_{1,k} = \frac{\beta_{1,k}}{\alpha_1 \sum_j \frac{1}{\alpha_j}}. \quad (\text{A.6})$$

for $0 < k < N_1$. Again, one can argue that the ordering of the subchains is arbitrary and that for any result concerning the first subchain there is a similar result for the other subchains. Hence, the previous equation can be generalized so that the coefficients associated with the eigenvalues $\{(1 - \alpha_i)\lambda_{i,k}\}$ are

$$\beta'_{i,k} = \frac{\beta_{i,k}}{\alpha_1 \sum_j \frac{1}{\alpha_j}} \quad (\text{A.7})$$

for $0 < k < N_i, 1 \leq i \leq M$ as was claimed in equation (A.4).

Hence, each T_i contributes $N_i - 1$ eigenvalues to T . That means that

$\sum_{i=1}^M (N_i - 1) = N - M$ of the N eigenvalues of T have been accounted for. The remaining M eigenvalues of T can be found as follows. Define the stochastic matrix T_s to be

$$T_s = \begin{pmatrix} (1 - \alpha_1) & \alpha_1 & & & & \\ & (1 - \alpha_2) & \alpha_2 & & & \\ & & & \ddots & & \\ & & & & \ddots & \\ \alpha_M & & & & & (1 - \alpha_M) \end{pmatrix} \quad (\text{A.8})$$

and suppose T_i is diagonalizable with eigenvalues $\{\rho_k\}$ and associated right eigenvectors $\{\mathbf{v}_k\}$. Let $\mathbf{v}_k = (v_{k,1} \ v_{k,2} \ \dots \ v_{k,M})^T$. Consider the column vector whose first N_1 elements are $v_{k,1}$, next N_2 elements are $v_{k,2}$ and so on. This vector can be expressed in block vector notation as $\mathbf{v}'_k = (v_{k,1}\mathbf{1}_{N_1}^T \ v_{k,2}\mathbf{1}_{N_2}^T \ \dots \ v_{k,M}\mathbf{1}_{N_M}^T)^T$. Consider the product $T\mathbf{v}'_k$:

$$\begin{aligned} & \begin{pmatrix} (1 - \alpha_1)T_1 & \alpha_1\mathbf{1}_{N_1}s_2^T & & & & \\ & (1 - \alpha_2)T_2 & \alpha_2\mathbf{1}_{N_2}s_3^T & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ \alpha_M\mathbf{1}_{N_M}s_1^T & & & & & (1 - \alpha_M)T_M \end{pmatrix} \begin{pmatrix} v_{k,1}\mathbf{1}_{N_1} \\ v_{k,2}\mathbf{1}_{N_2} \\ \vdots \\ \vdots \\ v_{k,M}\mathbf{1}_{N_M} \end{pmatrix} \\ &= \begin{pmatrix} (1 - \alpha_1)v_{k,1}T_1\mathbf{1}_{N_1} + \alpha_1v_{k,2}\mathbf{1}_{N_1}s_2^T\mathbf{1}_{N_2} \\ (1 - \alpha_2)v_{k,2}T_2\mathbf{1}_{N_2} + \alpha_2v_{k,3}\mathbf{1}_{N_2}s_3^T\mathbf{1}_{N_3} \\ \vdots \\ \vdots \\ (1 - \alpha_M)v_{k,M}T_M\mathbf{1}_{N_M} + \alpha_Mv_{k,1}\mathbf{1}_{N_M}s_1^T\mathbf{1}_{N_1} \end{pmatrix}. \end{aligned}$$

But, $T_i\mathbf{1}_{N_i} = \mathbf{1}_{N_i}$ since T_i is stochastic. Also, $s_i^T\mathbf{1}_{N_i} = 1$. Using these simplifications,

$$T\mathbf{v}'_k = \begin{pmatrix} [(1 - \alpha_1)v_{k,1} + \alpha_1v_{k,2}]\mathbf{1}_{N_1}^T \\ [(1 - \alpha_2)v_{k,2} + \alpha_2v_{k,3}]\mathbf{1}_{N_2}^T \\ \vdots \\ \vdots \\ [(1 - \alpha_M)v_{k,M} + \alpha_Mv_{k,1}]\mathbf{1}_{N_M}^T \end{pmatrix}. \quad (\text{A.9})$$

Next, use the fact that $T_s\mathbf{v}_k = \rho_k\mathbf{v}_k$ and expand:

$$T_s \mathbf{v}_k = \begin{pmatrix} (1 - \alpha_1)v_{k,1} + \alpha_1 v_{k,2} \\ (1 - \alpha_2)v_{k,2} + \alpha_2 v_{k,3} \\ \vdots \\ \vdots \\ (1 - \alpha_M)v_{k,M} + \alpha_M v_{k,1} \end{pmatrix} = \begin{pmatrix} \rho_k v_{k,1} \\ \rho_k v_{k,2} \\ \vdots \\ \vdots \\ \rho_k v_{k,M} \end{pmatrix} = \rho_k \mathbf{v}_k.$$

Hence, $(1 - \alpha_1)v_{k,1} + \alpha_1 v_{k,2} = \rho_k v_{k,1}$, $(1 - \alpha_2)v_{k,2} + \alpha_2 v_{k,3} = \rho_k v_{k,2}$ and so on. Making these substitutions in equation (A.7) yields

$$T \mathbf{v}'_k = \begin{pmatrix} \rho_k v_{k,1} \mathbf{1}_{N_1} \\ \rho_k v_{k,2} \mathbf{1}_{N_2} \\ \vdots \\ \vdots \\ \rho_k v_{k,M} \mathbf{1}_{N_M} \end{pmatrix} = \rho_k \mathbf{v}'_k$$

and proves that $\{\rho_k\}$ are eigenvalues of T . A similar result relates left eigenvectors of T_s to left eigenvectors of T : if $\mathbf{u}_k = (u_{k,1} \ u_{k,2} \ \cdots \ u_{k,M})$ is the left eigenvector of T_s , associated with the eigenvalue ρ_k , then $\mathbf{u}'_k = (u_{k,1} \mathbf{1}_{N_1}^T \ u_{k,2} \mathbf{1}_{N_2}^T \ \cdots \ u_{k,M} \mathbf{1}_{N_M}^T)^T$ is the left eigenvector of T associated with the eigenvalue ρ_k .

Next, consider the spectral coefficients for these eigenvalues. As before, split the expression, which is given by $\beta_k = \mathbf{a}^T S \mathbf{v}'_k (\mathbf{u}'_k)^T \mathbf{a}$, into two parts. First consider $\mathbf{a}^T S \mathbf{v}'_k$:

$$\mathbf{a}^T S \mathbf{v}'_k = \frac{\mathbf{a}^T}{\sum_j \frac{1}{\alpha_j}} \begin{pmatrix} \frac{1}{\alpha_1} S_1 & & & & \\ & \frac{1}{\alpha_2} S_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \frac{1}{\alpha_M} S_M \end{pmatrix} \begin{pmatrix} v_{k,1} \mathbf{1}_{N_1} \\ v_{k,2} \mathbf{1}_{N_2} \\ \vdots \\ \vdots \\ v_{k,M} \mathbf{1}_{N_M} \end{pmatrix}$$

$$\begin{aligned}
&= \frac{1}{\sum_j \frac{1}{\alpha_j}} (\mathbf{a}_1^T \quad \mathbf{a}_2^T \quad \cdots \quad \mathbf{a}_M^T) \begin{pmatrix} \frac{v_{k,1}}{\alpha_1} \mathbf{S}_1 \mathbf{1}_{N_1} \\ \frac{v_{k,2}}{\alpha_2} \mathbf{S}_2 \mathbf{1}_{N_2} \\ \cdot \\ \cdot \\ \cdot \\ \frac{v_{k,M}}{\alpha_M} \mathbf{S}_M \mathbf{1}_{N_M} \end{pmatrix} \\
&= \frac{1}{\sum_j \frac{1}{\alpha_j}} \left(\sum_i \frac{v_{k,i}}{\alpha_i} \mathbf{a}_i^T \mathbf{S}_i \mathbf{1}_{N_i} \right).
\end{aligned}$$

But, $\mathbf{S}_i \mathbf{1}_{N_i} = \mathbf{s}_i$ which yields

$$\beta_k = \frac{1}{\sum_j \frac{1}{\alpha_j}} \left(\sum_i \frac{v_{k,i}}{\alpha_i} \mathbf{a}_i^T \mathbf{s}_i \right).$$

Finally, $\mathbf{a}_i^T \mathbf{s}_i = \mu_i = 0$ by a previous assumption. The consequence of this is that $\beta_k = 0$ for all of the eigenvalues $\{\rho_k\}$. This is fortunate since these eigenvalues are much more difficult to work with than the $\{(1 - \alpha_i) \lambda_{i,k}\}$ eigenvalues.

Now suppose that T_s is not diagonal. As an example, suppose that T_s is a 3x3 matrix with Jordan form

$$J = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \rho & 1 \\ 0 & 0 & \rho \end{pmatrix}$$

That is, T_s has two eigenvalues: 1, which has unit multiplicity; and ρ , which has algebraic multiplicity 2, but geometric multiplicity 1. Further, suppose the modal matrix of T_s is given in block vector notation as

$$V = (\mathbf{v}_0 \quad \mathbf{v}_1 \quad \mathbf{v}_2)$$

so that $TV = VJ$. Let us write out this equation in terms of the component vectors of V and the eigenvalues:

$$T(v_0 \ v_1 \ v_2) = (v_0 \ v_1 \ v_2) \begin{pmatrix} 1 & 0 & 0 \\ 0 & \rho & 1 \\ 0 & 0 & \rho \end{pmatrix}$$

Therefore,

$$T_s v_0 = v_0,$$

$$T_s v_1 = \rho v_1,$$

$$\text{and } T_s v_2 = v_1 + \rho v_2$$

Now let $v_k = (v_{k,1} \ v_{k,2} \ \cdots \ v_{k,M})^T$, and let v'_k be the vector

$v'_k = (v_{k,1} \mathbf{1}_{N_1}^T \ v_{k,2} \mathbf{1}_{N_2}^T \ \cdots \ v_{k,M} \mathbf{1}_{N_M}^T)^T$. Then, using the procedure as in the diagonalizable case given above, it is possible to show that

$$T v'_0 = v'_0,$$

$$T v'_1 = \rho v'_1,$$

$$\text{and } T v'_2 = v'_1 + \rho v'_2.$$

Hence, 1 and ρ are eigenvalues of T . However, in this case, ρ has algebraic multiplicity 2, but geometric multiplicity 1.

Now, this is precisely the situation discussed in Chapter II in the section concerning the power spectra of MCRP with nondiagonalable transition probability matrices. Hence, from Equation (II.35), it is clear that the autocorrelation of this MCRP will have a ρ^n term and its spectral coefficient β_1 will be

$$\beta_1 = a^T S v'_1 (u'_1)^T a + a^T S v'_2 (u'_2)^T a$$

where \mathbf{u}'_1 and \mathbf{u}'_2 are the corresponding left eigenvectors of T associated with ρ . Additionally, the autocorrelation of this MCRP will have a $n\rho^{n-1}$ term and its spectral coefficient β_2 will be

$$\beta_2 = \mathbf{a}^T \mathbf{S}(\mathbf{v}'_1(\mathbf{u}'_2)^T) \mathbf{a}$$

Formally, the autocorrelation also includes $\beta_0(1)^n$, but it was shown in Chapter II that β_0 is always equal to zero for a process like this one with zero mean.

To show that β_1 and β_2 are zero, consider $\mathbf{a}^T \mathbf{S} \mathbf{v}'_k(\mathbf{u}'_j)^T \mathbf{a}$ for $1 \leq k, j \leq 2$. First consider $\mathbf{a}^T \mathbf{S} \mathbf{v}'_k$:

$$\begin{aligned} \mathbf{a}^T \mathbf{S} \mathbf{v}'_k &= \frac{\mathbf{a}^T}{\sum_j \frac{1}{\alpha_j}} \begin{pmatrix} \frac{1}{\alpha_1} \mathbf{S}_1 & & & & & \\ & \frac{1}{\alpha_2} \mathbf{S}_2 & & & & \\ & & \ddots & & & \\ & & & \ddots & & \\ & & & & \frac{1}{\alpha_M} \mathbf{S}_M & \\ & & & & & \end{pmatrix} \begin{pmatrix} v_{k,1} \mathbf{1}_{N_1} \\ v_{k,2} \mathbf{1}_{N_2} \\ \vdots \\ \vdots \\ \vdots \\ v_{k,M} \mathbf{1}_{N_M} \end{pmatrix} \\ &= \frac{1}{\sum_j \frac{1}{\alpha_j}} (\mathbf{a}_1^T \quad \mathbf{a}_2^T \quad \cdots \quad \mathbf{a}_M^T) \begin{pmatrix} \frac{v_{k,1}}{\alpha_1} \mathbf{S}_1 \mathbf{1}_{N_1} \\ \frac{v_{k,2}}{\alpha_2} \mathbf{S}_2 \mathbf{1}_{N_2} \\ \vdots \\ \vdots \\ \frac{v_{k,M}}{\alpha_M} \mathbf{S}_M \mathbf{1}_{N_M} \end{pmatrix} \\ &= \frac{1}{\sum_j \frac{1}{\alpha_j}} \left(\sum_i \frac{v_{k,i}}{\alpha_i} \mathbf{a}_i^T \mathbf{S}_i \mathbf{1}_{N_i} \right). \end{aligned}$$

Noting that $S_i \mathbf{1}_N = s_i$,

$$\begin{aligned} \mathbf{a}^T S \mathbf{v}'_k &= \frac{1}{\sum_j \frac{1}{\alpha_j}} \left(\sum_i \frac{v_{k,i}}{\alpha_i} \mathbf{a}_i^T s_i \right) \\ &= 0 \end{aligned}$$

since $\mathbf{a}_i^T s_i = \mu_i = 0$ by a previous assumption. The consequence of this is that $\beta_1 = 0$ and $\beta_2 = 0$. Hence, even in this case when T_s is not diagonalizable, its eigenvalues do not influence the autocorrelation or power spectrum of the MCRP.

To summarize, each T_i matrix contributes $N_i - 1$ eigenvalues to T , and each of these eigenvalues can have nonzero spectral coefficients. The remaining M eigenvalues of T are found from the T_s matrix defined in (A.8). However, these eigenvalues do not influence the autocorrelation and power spectrum of the ring-structured MCRP.

2

VITA

William Scott King

Candidate for the Degree of

Doctor of Philosophy

Thesis: THE SPECTRAL PROPERTIES OF MARKOV CHAIN RANDOM PROCESSES WITH APPLICATION TO SPEECH MODELING AND SPREAD SPECTRUM COMMUNICATIONS

Major Field: Electrical Engineering

Biographical:

Personal Data: Born in Stillwater, Oklahoma, November 6, 1963, the son of William and June King.

Education: Graduated from Stillwater High School, Stillwater, Oklahoma, in May 1982; received Bachelor of Science degree in Electrical Engineering from Oklahoma State University in May, 1987; attended University of Texas, Austin, the fall of 1987 and spring of 1988; received Master of Science degree from Oklahoma State University in May 1989; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in December, 1991.

Professional Experience: Co-op Engineer at General Dynamics, Fort Worth Division, the summer of 1985, spring of 1986, and fall of 1986. Research and Teaching Assistant, School of Electrical and Computer Engineering, Oklahoma State University, August, 1988, to December, 1991.