INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Some problems in stochastic partial realization

Liu, Duixian, Ph.D.
The Ohio State University, 1990

Copyright ©1990 by Liu, Duixian. All rights reserved.
Some Problems in Stochastic Partial Realization

A Dissertation
Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of the Ohio State University

by

Duixian Liu, M.S.E.E.

The Ohio State University
1990

Dissertation Committee:
Professor R. L. Moses
Professor A. Krishnamurthy
Professor M. C. Smith

Approved by:

Adviser
Department of Electrical Engineering
To My Parents
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and admiration to Professor R. L. Moses, my advisor, for his guidance and insight throughout the research, and for his suggestions and help beyond the academic field. I would also like to express my gratitude to Professor R. J. Garbacz, my M.S. advisor and mentor, for his guidance, suggestions and valuable help through my six tough years at the Ohio State University. Gratitude is also owed to Professor A. Krishnamurthy and Professor M. C. Smith for their review of the draft, to Professor R. J. Mayhan for arranging financial support, to Professor R. G. Kouyoumjian for teaching me the essence of electromagnetics, to Professor H. Hsu for his advice and help during these years.

The support I received for my M.S. program from the Chinese government is deeply appreciated.

Finally, I thank my father Huaiyi Liu, mother Yumin Liu, elder sister Caixia Liu, elder brother Shenlan Liu, younger sister Gaoping Liu, the youngest sister Jinping Liu, and my wife Shuang Huang, for their encouragement, support, love and patience during my six years absence while I completed my studies at the Ohio State University.
VITA

September 13, 1960 .................................Born—Xi’an, China

1982 ..........................................................B.S.E.E., Northwest Telecommunications
Engineering Institute, Xi’an, China

1986 ..........................................................M.S.E.E., The Ohio State University,
Columbus, Ohio

1986-1987 .................................................Graduate Research Associate
The Ohio State University
Columbus, Ohio

1988-1990 .................................................Graduate Teaching Associate
The Ohio State University
Columbus, Ohio

PUBLICATIONS

“Antenna synthesis and optimization using generalized characteristic
modes,” IEEE Transactions on Antennas and Propagation, vol. AP-
38, pp. 862-868, June 1990 (with R. J. Garbacz and D. M. Pozar).

“On determining the closest stable polynomial to an unstable one,” to
appear in IEEE Transactions on Acoustics, Speech, and Signal Process-
ing (with R. L. Moses).

“On nonnegative definiteness of estimated moving average autocovari-
ance sequences,” in Proceedings of the International Conference on
Acoustics, Speech, and Signal Processing, New York, April 11-14, 1988,
pp. 2340-2343. (with R. L. Moses)

“Generalized characteristic modes, generalized Inagaki modes and their
interrelationships,” in Proceedings of URSI, 1987 IEEE Antennas and
Propagation Society International Symposium and URSI Radio Science
FIELDS OF STUDY

Major Field: Electrical Engineering

Studies in Communications: Professor Randolph L. Moses
Studies in Electromagnetics: Professor Robert J. Garbacz
                      Professor Robert G. Kouyoumjian
Studies in Mathematics: Professor Joseph M. Rosenblatt
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGEMENTS</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>VITA</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. DETERMINING THE CLOSEST STABLE POLYNOMIAL TO AN UNSTABLE ONE</td>
<td>6</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>6</td>
</tr>
<tr>
<td>2.2 Problem Statement</td>
<td>9</td>
</tr>
<tr>
<td>2.3 Characterization of the Stability Set</td>
<td>9</td>
</tr>
<tr>
<td>2.4 Solution to the Stabilization Problem</td>
<td>11</td>
</tr>
<tr>
<td>2.5 Examples</td>
<td>16</td>
</tr>
<tr>
<td>2.6 Conclusions</td>
<td>28</td>
</tr>
<tr>
<td>III. OPTIMAL NON-NEGATIVE DEFINITE APPROXIMATIONS OF ESTIMATED MOVING AVERAGE COVARIANCE SEQUENCES</td>
<td>29</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>29</td>
</tr>
<tr>
<td>3.2 Problem Statement</td>
<td>32</td>
</tr>
<tr>
<td>3.3 Description of the Admissible Set $D$</td>
<td>34</td>
</tr>
<tr>
<td>3.4 Solution to the Minimization Problem</td>
<td>36</td>
</tr>
<tr>
<td>3.5 Further Characterization of Solutions</td>
<td>46</td>
</tr>
<tr>
<td>3.5.1 Variance Bounds</td>
<td>46</td>
</tr>
<tr>
<td>3.5.2 Characterization of Zero Frequencies</td>
<td>48</td>
</tr>
<tr>
<td>3.6 Examples</td>
<td>55</td>
</tr>
<tr>
<td>3.7 Conclusions</td>
<td>59</td>
</tr>
<tr>
<td>3.11 Introduction</td>
<td>62</td>
</tr>
<tr>
<td>4.2 Solutions to SPR</td>
<td>67</td>
</tr>
<tr>
<td>4.2.1 The Schur Parameters</td>
<td>67</td>
</tr>
<tr>
<td>4.2.2 Orthogonal Polynomial Descriptions</td>
<td>68</td>
</tr>
<tr>
<td>4.2.3 Remarks</td>
<td>70</td>
</tr>
<tr>
<td>4.3 Minimal Order Partial Realizations</td>
<td>71</td>
</tr>
<tr>
<td>4.4 Minimal Order Realization for Single Unknown Case</td>
<td>74</td>
</tr>
<tr>
<td>4.4.1 Stability Set</td>
<td>75</td>
</tr>
<tr>
<td>4.4.2 Non-Negative Definite Set</td>
<td>81</td>
</tr>
<tr>
<td>4.5 Some Further Results</td>
<td>85</td>
</tr>
<tr>
<td>4.5.1 General Case of Stability Set</td>
<td>85</td>
</tr>
<tr>
<td>4.5.2 General Case of Non-Negative Definite Set</td>
<td>90</td>
</tr>
<tr>
<td>4.6 Examples</td>
<td>95</td>
</tr>
</tbody>
</table>
4.7 Approximate Stochastic Partial Realizations ........................................... 100
4.8 Conclusions ............................................................................................. 109

V. SUMMARY .................................................................................................. 111

REFERENCES ............................................................................................... 114
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>TABLES</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Solutions to Example 2.2 Using Optimal Method.</td>
<td>18</td>
</tr>
<tr>
<td>2  Solutions to Example 2.2 Using Zero Moving Methods.</td>
<td>19</td>
</tr>
<tr>
<td>3  Solutions to Example 2.3 Using Optimal Method.</td>
<td>25</td>
</tr>
<tr>
<td>4  Values of $\delta_n$ for Various Orders $n$</td>
<td>53</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

FIGURES  PAGE

1  ARMA model. ................................................................. 2

2  Optimal solutions for $n = 2$. ........................................ 17

3  Zero distribution before stabilization of Example 2.2. ....... 19

4  Zero distribution after stabilization of Example 2.2, using optimal method. .................................................. 20

5  Zero distribution after stabilization of Example 2.2, using zero reflection method. .............................................. 22

6  Zero distribution after stabilization of Example 2.2, moving unstable zeros to the unit circle. ................................. 23

7  Zero distribution before stabilization of Example 2.3. ........ 26

8  Zero distribution after stabilization of Example 2.3, using the optimal method. ..................................................... 27

9  Non-Negative Definite Regions $N$ and $D$ for $n = 2$. ......... 37

10 The duality principle for the minimal distance from a point outside a convex set to the set. ................................. 41

11 Solutions on $N$ for $n = 2$. ................................................ 42

12 The variance bounds for the $n = 1$ case. ............................ 47
13 The plots of $g(\omega)$ and $f(\omega, \rho^*)$ for $n = 3$ with $s = [3.8, -6.16, 3.34, -1]^T$. 
   Note that $g(70.126^\circ) > 0$. ................................................................. 49
14 Geometric description for Conjecture C1. ............................................. 51
15 The plots of $g(\omega)$ and $f(\omega, \rho^*)$ for $n = 2$ and $s = [1, 1.5, -3]^T$. 
   Note that if $g(\omega) < 0$ on an interval, it is not necessarily true that 
   $f(\omega, \rho^*) = 0$ on that interval. .......................................................... 54
16 Geometric description for Conjecture C2. ............................................. 56
17 Example 3.1 Original, Optimal NND and Distance Measure functions. ................................................................. 57
18 Example 3.2 Original, Optimal NND and Distance Measure functions. ................................................................. 58
19 Example 3.3 Original, Optimal NND and Location functions. .......... 60
20 PR region for $m = 1$. ............................................................................. 66
21 Equation $3a_1 + 2a_2 = -1$ cuts the stability region. ......................... 96
22 The 3D plot of $f(\omega, c_4)$ for $C^{(3)} = \{5, 2, 3, 1\}$ ......................... 98
23 The locus of $[\rho_1/\rho_0, \rho_2/\rho_0]$ as $c_4$ changes from $-30$ to $10$ with $C^{(3)} = \{5, 2, 3, 1\}$. ................................................................. 99
24 The 3D plot of $f(\omega, c_4)$ for $C^{(3)} = \{5, 2, 1, 1\}$ ......................... 101
25 The locus of $[\rho_1/\rho_0, \rho_2/\rho_0]$ as $c_4$ changes from $-5$ to $5$ with $C^{(3)} = \{5, 2, 1, 1\}$. ................................................................. 102
26 Original, optimal, and suboptimal spectral density functions corre- 
   sponding to Example 4.10. ................................................................. 107
27 Original, optimal, and suboptimal spectral density functions using 
   $A_2(z) = z^2 - 1.029z + 0.95$ for the suboptimal method. .......... 108
CHAPTER I
INTRODUCTION

The modeling of stationary stochastic time series is a problem of great importance in digital signal processing and control system theory. Time series modeling is used in a wide variety of applications, including system identification [1,2,3,4], geophysics [5,6,7,8,9], radar and sonar processing [10,11,12,13], imaging [14], radio astronomy [15], biomedical signal processing [16], oceanography [17], ecological systems [18], speech processing [19,20,21], and economics [22], to name a few. The time series modeling problem has been considered using a variety of approaches, including input-output modeling [23,24,25], state space modeling [26,27], spectral estimation [24,25], and stochastic realization [27]-[42]. Many of these techniques are based on first estimating the covariance sequence of the time series data, then forming a time series (or spectral) model from the covariance sequence; the latter step is often referred to as stochastic (partial) realization.

In this dissertation we consider problems related to time series modeling using autoregressive moving average (ARMA) models. In order to define these problems, let us establish some basic definitions. We assume the stochastic time series \{x(k)\} to be modeled is the output of a linear system which is driven by a zero mean, unit variance white noise sequence \{w(k)\} as shown in Figure 1. The linear time invariant system \(H(z)\) is described by an ARMA model, so \(H(z) = \frac{B(z)}{A(z)}\) where

\[
A(z) = 1 + a_1z^{-1} + \cdots + a_mz^{-m}, \quad (1.1)
\]
The time series \( \{x(k)\} \) can be described by its corresponding autocovariance sequence \( \{c_k\}_{k=-\infty}^{\infty} \) where

\[
c_k = E\{x(n)x(n+k)\}. \tag{1.3}
\]

The power spectral density function \( P_x(\omega) \) corresponding to \( \{x(k)\} \) is given by

\[
P_x(\omega) = \sum_{k=-\infty}^{\infty} c_k e^{-jk\omega} = \left| \frac{B(e^{j\omega})}{A(e^{j\omega})} \right|^2. \tag{1.4}
\]

It can be seen from (1.4) that \( c_k \) is completely parameterized by the ARMA model parameters \( \{a_1, \ldots, a_m, b_0, \ldots, b_m\} \).

This dissertation is concerned with three problems related to stochastic partial realization. Briefly stated, stochastic partial realization is concerned with obtaining an ARMA model from the covariance \( c_k \) of \( \{x(k)\} \). Specifically, the Stochastic Partial Realization (SPR) problem is [27]-[42]

Given the partial covariance sequence \( C^{(n)} = \{c_0, c_1, \ldots, c_n\} \) which correspond to a wide sense stationary stochastic time series \( \{x(k)\} \), find
all possible ARMA time series models of \( \{x(k)\} \) which are consistent with \( C^{(n)} \).

Of particular interest is in finding ARMA models of low order, as these provide "simple" modeling representations of the time series [32].

A closely related problem is the Approximate Stochastic Partial Realization (ASPR) problem, in which one assumes that the given partial covariance sequence \( C^{(n)} \) is "noisy"; in this case one is interested in finding ARMA models of \( \{x(k)\} \) whose corresponding covariance sequence closely approximates the given sequence in some sense. In the ASPR problem, the given covariance sequence is often an estimate of the covariances obtained from a finite measurement set \( \{x(k)\}_{k=1}^{N} \). Thus, any time series modeling in which one first estimates covariances from data, then forms a model from these covariance estimates, can be considered as ASPR problem. As mentioned earlier, this is the approach taken by a wide variety of time series modeling methods [24,25], [43]-[47].

The SPR and ASPR problems are also closely related to spectral estimation problems. The connection arises from equation (1.4); if one obtains an ARMA model from the (A)SPR procedure, one has a spectral estimate via (1.4). Conversely, most AR, MA and ARMA spectral estimation methods involve first estimating covariances from a finite set of time series data \( \{x(k)\}_{k=1}^{N} \), then estimating the AR and MA parameters from these covariances [24,25], [43]-[47]. But this is just the ASPR problem.

There are two important restrictions which arise in modeling of stationary time series as in Figure 1. First, in order for \( \{x(k)\} \) to be stationary, the polynomial \( A(z) \) must be stable; that is, all its zeros must be inside the unit circle in the complex \( z \)-plane. Second, the power spectral density function \( P_x(\omega) \) is required
to be nonnegative and real; this is equivalent to requiring the infinite covariance sequence \( \{c_k\}_{k=-\infty}^{\infty} \) to be nonnegative definite [27]-[42]. If the partial sequence \( C^{(n)} \) is estimated, or if it is extended improperly, the NND requirement may be violated. These two issues are addressed in the dissertation.

This dissertation focuses on three problems associated with the (approximate) SPR problem. First, the problem \( A(z) \) must be stable. However, in many time series parameter or spectral estimation algorithms there is no guarantee that an estimate of \( A(z) \) is stable. Examples of AR estimates which do not guarantee stability including the covariance and prewindow based methods [24,25,3]. In fact, nearly all noniterative methods of ARMA modeling do not guarantee the stability of \( A(z) \) [24,25,48,49]. Second, \( P_x(\omega) \) must be nonnegative, but this non-negativity condition is not guaranteed in many algorithms [50]. This problem also occurs in spectral estimation algorithm in which the MA part of the spectrum is estimated using the estimated AR coefficients [45,54]. Third, we address the minimal order SPR problem [31,32,34,38]. Stochastic realizations of minimal order are of interest from the system-theoretic viewpoint as they provide the simplest explanation of the given covariance sequence [32]. These three problems are discussed in more detail below and in later chapters.

In Chapter II we consider the following stabilization problem: given a polynomial (corresponding to the AR part of the ARMA modeling) whose zeros do not all lie on or inside the unit circle, find the closest polynomial whose zeros are all on or inside the unit circle. The measure of closeness used is the weighted Euclidean distance in coefficient space. The algorithm can be extended to other measures of closeness as well. Because the direct minimization on the coefficient space is difficult, we approach the problem in Schur coefficient space. In this way, the stability
condition is easily guaranteed. We develop a very efficient algorithm for obtaining
the optimum solution.

In Chapter III we consider the following problem: Given an estimated covariance sequence which may not be Non-Negative Definite (NND), find the closest NND sequence to it. Here, closeness is measured by the weighted Euclidean distance of the covariances. We provide a solution to this problem by considering a set of linear minimization problems which are parameterized by the zero frequencies of the optimal solution. Some properties of the optimal NND solutions are established, and these properties are used to simplify the minimization procedure.

In Chapter IV we consider the minimal order stochastic partial realization problem. That is, given a sequence of real numbers $C^{(n)} = \{c_0, c_1, \ldots, c_n\}$, find all ARMA models of minimal order $m$ which are consistent with $C^{(n)}$. These ARMA models must satisfy both the stability and the nonnegativity properties discussed above. Without the minimal order constraint there are many solutions to the stochastic partial realization problem. However, if the minimal order is of interest, no general solution is known to the problem. We address this problem by considering the stability and the nonnegativity conditions separately. For the special case $n = 2m - 1$, this approach leads to a complete description of all $m$-th order realizations. For an arbitrary $n$, we develop necessary conditions for the existence of solutions.

Finally, in Chapter V we summarize the results we have obtained and discuss some open problems related to these results.
CHAPTER II

DETERMINING THE CLOSEST STABLE POLYNOMIAL TO AN UNSTABLE ONE

2.1 Introduction

In time series modeling and system identification problems, one often obtains an estimate of an autoregressive (denominator) polynomial. Depending on the particular estimator used, this polynomial may or may not be "stable"; that is, it may or may not have all its zeros inside the unit circle [25,3]. Examples of AR estimators which do not guarantee stability include the covariance and prewindow methods [25], and most singular value decomposition-based methods [48]. In addition, nearly all noniterative methods of ARMA modeling first estimate the AR coefficients by using some form of the extended Yule-Walker equations; these methods almost never guarantee that the estimated AR polynomial is stable [25,48,49]. Few system identification algorithms ensure stability of the estimate either [3,4].

Many applications require that the estimated denominator polynomial be stable. This is especially true in system identification applications, and in time series analysis applications which use the model as a synthesis filter (speech synthesis is one example [56]). Because most estimation algorithms do not guarantee stability, the following problem is of interest: given a polynomial whose zeros are not all inside the unit circle, find a "close" polynomial whose zeroes are all inside the unit circle. We call this the stabilization problem.

There are several ways to stabilize an unstable polynomial. One method is to
find the zeros of the unstable polynomial, and if any zero has magnitude greater than one, change it to have magnitude equal to (or slightly less than) one. In this case the stable polynomial is "close" to the original one in the sense of minimizing a distance measure based on the zero locations of the polynomials. In some applications this zero is reflected inside the unit circle by using the reciprocal of its magnitude; this choice has the property that $|B(z)| = |A(z)|$ on the unit circle (although the phases of these polynomials will differ). Another method based on the Schur parameters (or reflection coefficients) associated with a polynomial could be used: find the reflection coefficient sequence of the given polynomial (using the Levinson-Durbin recursions), and change any Schur parameter with magnitude greater than one to one which is (slightly less than) one in magnitude [57].

This chapter considers solutions to the stabilization problem that minimize the error between the polynomial coefficients of the original and stabilized polynomial. The reason for working in coefficient space is that most algorithms which estimate these polynomials directly estimate the coefficients of the polynomials (rather than the Schur parameters or the zeros corresponding to that polynomial). Since the polynomial coefficients are being estimated, it is natural to stabilize the polynomial by perturbing these estimated coefficients as little as possible.

More specifically, we use the weighted $\ell_2$ distance measure in coefficient space as the measure of closeness. The reason for this choice is that most of the polynomial coefficient estimation methods in time series analysis and system identification give coefficient estimates which are asymptotically Gaussian distributed as the number of data points used to estimate the coefficients becomes large [25,24]; this Gaussian distribution is obtained even when the data themselves are non-Gaussian. Since the polynomial coefficients are approximately Gaussian, the $\ell_2$ error is the
most natural distance metric to use in perturbing these coefficients. Moreover, asymptotic variance expressions for these coefficient estimates have been obtained for several algorithms [25,58]; in this case the inverse of the covariance matrix can be used as a weighting matrix in a weighted $\ell_2$ coefficient norm to form the distance measure. A stable polynomial whose (weighted) distance from the given polynomial is minimum has the interpretation of a minimum variance solution to the stabilization problem.

Although we minimize an error in coefficient space for the stabilization problem, we find that working in the Schur parameter space is easier because the stability condition is readily guaranteed. A related problem involving optimization of a covariance sequence was studied in [59], where the Schur parameter space was used to guarantee that a covariance sequence is nonnegative definite. We then employ the alternate minimization method [64] to derive a computationally efficient algorithm for solving the stabilization problem.

While the algorithm we present uses the $\ell_2$ norm, it readily generalizes to other $\ell_p$ norms as well; the only difference is that the error function is no longer quadratic in the parameters for $p \neq 2$, so the alternate minimization of the error function becomes more complex.

An outline of this chapter is as follows. In Section 2.2, we present a formal statement of the problem. In Section 2.3, some properties of the stability sets in both polynomial coefficients and Schur parameters are discussed. These properties form the foundations for our results. In Section 2.4, an efficient algorithm for solving this minimization problem is given. In Section 2.5, some examples are given to illustrate the algorithm.
2.2 Problem Statement

Assume we are given the real vector \( b = [b_1, \ldots, b_n]^T \), and that its associated polynomial

\[
B(z) = z^n + b_1 z^{n-1} + \cdots + b_{n-1} z + b_n \tag{2.1}
\]

has at least one zero \( z_0 \) satisfying \(|z_0| > 1\). We are interested in finding another vector \( a = [a_1, \ldots, a_n]^T \) which is close to \( b \), and such that its associated polynomial

\[
A(z) = z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n, \tag{2.2}
\]

has all its zeros on or inside the unit circle. The measure of error we use is the weighted Euclidean \( (\ell_2) \) distance

\[
J = \| a - b \|_W^2 \triangleq (a - b)^T W (a - b) \tag{2.3}
\]

for some given positive definite weighting matrix \( W \).

Consider the set of coefficients corresponding to stable polynomials:

\[
S_a = \{ a | A(z) = 0 \Rightarrow |z| \leq 1 \}. \tag{2.4}
\]

The stabilization problem can then be stated as follows:

**Problem SP:** Given a vector \( b \notin S_a \), find the vector \( a^\circ \in S_a \) such that \( J = (a^\circ - b)^T W (a^\circ - b) \) is minimized over all \( a \in S_a \).

2.3 Characterization of the Stability Set

In order to solve the above stabilization problem, it is useful to establish some basic properties of the stability set \( S_a \).
Let us first introduce the concepts of Schur parameters (also known as reflection coefficients in the signal processing literature). For any polynomial of degree $k$, we define [60]

$$\Phi_k^+(z) = z^k \Phi_k(z^{-1}).$$  
(2.5)

Then for an arbitrarily given real vector $[r_1, \ldots, r_n]$ (the vector of Schur parameters), we can determine a set of polynomials $\Phi_i(z), i = 0, 1, \ldots, n$ through the following recursive formulas:

$$\Phi_i(z) = z \Phi_{i-1}(z) + r_i \Phi_{i-1}^+(z), \quad \Phi_0(z) = 1$$  
(2.6)

The following two properties of $\Phi_n(z)$ are useful in the stability study of the discrete time system analysis [2]:

1) $\Phi_n(z)$ has all its zeros inside the unit circle $|z| = 1$ if and only if $|r_i| < 1$ for $i = 1, 2, \ldots, n$.

2) Assume $|r_i| < 1$ for $i = 1, 2, \ldots, (n - 1)$. Then $\Phi_n(z)$ has all its zeros on the unit circle $|z| = 1$ if and only if $|r_n| = 1$.

From these two properties, the following property readily follows:

3) If $|r_i| \leq 1$ for $i = 1, \ldots, n$, then $\Phi_n(z)$ has all its zeros on the unit disk $|z| \leq 1$.

With the help of property 3), we can count the number of zeros inside the unit circle and the number of zeros on the unit circle of $\Phi_n(z)$. Let us express $\Phi_k(z), k = 1, \ldots, n$ in terms of a coefficient vector $[a_{k1}, \ldots, a_{kk}]$ where

$$\Phi_k(z) = z^k + a_{k1}z^{k-1} + \cdots + a_{kk}.$$  
(2.7)
It is easy to show that the vector \([a_{k1}, \ldots, a_{kk}]\) and the vector \([r_1, \ldots, r_k]\) are related by

\[
\begin{cases}
  a_{ki} = a_{(k-1)i} + r_{k}a_{(k-1)(k-i)} & \text{for } i = 1, \ldots, (k - 1) \\
  a_{kk} = r_k
\end{cases}
\]

(2.8)

We know that the stability region of polynomial \(A(z) = \Phi_n(z)\) in terms of Schur parameters \([r_1, \ldots, r_n]\) is completely specified by the unit hypercube:

\[
S_r = \{|r_i| \leq 1, i = 1, \ldots, n\}.
\]

(2.9)

This set is a closed and bounded subset of \(R^n\). Since the \(a_i\) parameters are continuous functions of the Schur parameters, we have the following:

**Theorem 2.1:**

a) \(S_a\) is a bounded and closed subset of \(R^n\).

b) Let \(B_S\) denote the boundary of \(S_a\). Then if \(a \in B_S\), there is at least one zero \(z_0\) of (2.2) satisfying \(|z_0| = 1\).

It is obvious that the set \(S_r\) is convex, but the set \(S_a\) is not convex in general. As an example, \(a = [3, 3, 1]^T\) and \(b = [-3, 3, -1]^T\) are in \(S_a\), but \(\alpha a + (1 - \alpha)b = [0, 3, 0]^T\) with \(\alpha = 0.5\) is not in \(S_a\). Since \(S_a\) is not a convex set, there may not be a unique solution to the stabilization problem. However, since \(S_a\) is closed, any solution to the stabilization problem will lie on the boundary of \(S_a\).

### 2.4 Solution to the Stabilization Problem

From the previous discussion we know that the \(a_i\) coefficients can be expressed in terms of the Schur parameters \([r_1, \ldots, r_n]\) and the stability set \(S_a\) can be trans-
formed to the stability set $S_r$. We will find the optimal solution $\alpha^*$ by working on the set $S_r$.

For fixed $[r_1, \ldots, r_{k-1}, r_{k+1}, \ldots, r_n]$, each $a_i$ ($i = 1, \ldots, n$) is a linear function of $r_k$. Thus $J$ will be (at most) a quadratic function of $r_k$, and its minimum is easily found. Since we need only three points to specify a quadratic function, we calculate the following three points: $Q_{-1} = J$ at $r_k = -1$, $Q_0 = J$ at $r_k = 0$, and $Q_1 = J$ at $r_k = 1$. Then the quadratic function can be found as

$$J_k = J(r_k) = \alpha r_k^2 + \beta r_k + \gamma,$$  

where

$$\alpha = \frac{1}{2}[Q_{-1} + Q_1 - 2Q_0], \quad \beta = \frac{1}{2}[Q_1 - Q_{-1}], \quad \gamma = Q_0.$$

Note that $\alpha \geq 0$ and $\gamma \geq 0$. The minimum value of $J_k$ on $-1 \leq r_k \leq 1$ with respect to $r_k$ can be found very easily as

$$\begin{cases} 
\gamma - 0.25\beta^2/\alpha & \text{if } -1 \leq -0.5\beta/\alpha \leq 1, \\
\gamma + \beta & \text{if } \alpha = 0 \text{ and } \beta \leq 0, \\
\gamma - \beta & \text{if } \alpha = 0 \text{ and } \beta \geq 0,
\gamma & \text{if } \alpha = 0 \text{ and } \beta = 0
\end{cases}$$

Since $J_k = J(r_k)$ is nonnegative for any $r_k$, thus if $\alpha = 0$, we also have $\beta = 0$.

To use this quadratic function property, the alternative minimization procedure is used (see [64] for a description of the alternative maximization procedure). The algorithm for finding the optimal Schur parameters $r^*$ is as follows:

1) Initialize $r^0 \in S_r$ and accuracy $\epsilon$. Set $i = 1$.

2) Calculate $r_k^i$ by using equations (2.10) and (2.12) successively for $k = 1, \ldots, n.$
3) If \( \max_{1 \leq k \leq n} |r_{k-1}^i - r_k^i| \leq \epsilon \), then go to 4); otherwise increase \( i \) and go to 2).

4) Calculate \( a^0 \) from \( r^i \) by equation (2.8).

Note that the above algorithm is based on a (weighted) \( \ell_2 \) error norm, which results in a quadratic error function in equation (2.10). If an \( \ell_p \) norm is used for \( 1 \leq p \leq \infty \), the above algorithm can still be used; the only difference is that \( J_k \) is no longer quadratic in \( r_k \), so its minimum on \(-1 \leq r_k \leq 1\) must be found by some other means. For \( 1 < p < \infty \), \( J_k \) is a continuously differentiable function of \( r_k \), and can be minimized using standard techniques. For \( p = 1 \) or \( p = \infty \), \( J_k \) is not continuously differentiable, and more care must be taken to find its minimum.

In the case \( \alpha = 0 \) and \( \beta = 0 \), \( J_k \) is independent of \( r_k \). Theoretically, we can choose any number on \([-1, 1]\) for \( r_k \). However, we must be careful about the selection of \( r_k \) values, otherwise the algorithm can get "stuck". As an example, for \( n = 2 \) we know

\[
a_1 = r_1(1 + r_2), \quad a_2 = r_2.
\]

Assume \( b = [2, -2]^T \) and that the initial value is \( r^0 = [0, -1]^T \). Then if we do the minimization procedure with respect to \( r_1 \) with \( r_2 = -1 \) fixed, we have \( \alpha = 0 \) and \( \beta = 0 \). If we select \( r_1 = 0 \), then the minimization with respect to \( r_2 \) gives \( r_2 = -1 \). Thus, the \( r^1 = [0, -1]^T \) and the algorithm fails to converge to the true solution of \( r^\infty = [1, 0.5]^T \). However, if we select any value for \( r_1 \) on \([-1, 1]\) except 0, we can continue the minimization procedure. In fact if we chose any number \( r_1 \in [-1, 1] \) and any number \( r_2 \in (-1, 1] \) as the initial values, the algorithm converges to the true minimum. In general, if the algorithm "sticks" at an iteration, we can perturb the coefficients slightly and proceed. In our algorithm we choose \( r_k^i = r_k^{i-1} \) for the case \( \alpha = 0 \) and \( \beta = 0 \).
Since \( S_a \) is not convex, there is the possibility of convergence to a local minimum. One way to overcome this situation is to try several initial values. Because we can use the quadratic function property, we have a very efficient minimization algorithm for each initial condition. The choice of a “good” initial guess can often eliminate problems of convergence to local minima. In most applications, the given polynomial is an estimate of a stable polynomial, and is therefore expected to be not too far away from a stable polynomial. In this case a good initial guess can be found using the following procedure [57].

- From \( B(z) \) compute the Schur parameters \( r^b = [r_1^b, \ldots, r_n^b]^T \).
- For each \( i \), if \( |r_i^b| > 1 \) replace it by \( \text{sign}(r_i^b) \cdot (1 - \alpha) \) for some small positive \( \alpha \) (we use \( \alpha = 0.001 \)). This modified Schur sequence is used as the initial sequence \( r^0 \).

Note that \( r^0 \in S_r \). Also, \( r^0 \) is close to \( r^b \), and their corresponding polynomials are close to each other [57].

Another initial guess can be found using the following procedure.

- Find all the zeros \( \{z_i\} \) of \( B(z) = 0 \).
- Form another sequence \( \{z_i'\} \), where \( z_i' = z_i \) if \( |z_i| < 1 \) and \( z_i' = (1 - \alpha)z_i/|z_i| \), for some small positive \( \alpha \), if \( |z_i| \geq 1 \).
- Use \( \{z_i'\} \) as the zeros to form a polynomial \( B'(z) \).
- From \( B'(z) \) compute the Schur parameters, and use them as the initial \( r^0 \).

This initial guess requires more computation, as the zeros of \( B(z) \) must be found. However, we have found that this method gives a better initial guess than does the previous method.
Note that in the case where $A(z)$ is required to be strictly stable, equation (2.12) is easily modified as

$$
\begin{align*}
\gamma - 0.25\beta^2/\alpha & \quad \text{if } -(1 - \epsilon_b) \leq -0.5\beta/\alpha \leq 1 - \epsilon_b, \\
\gamma + \beta(1 - \epsilon_b) & \quad \text{if } \alpha = 0 \text{ and } \beta < 0, \\
\gamma - \beta(1 - \epsilon_b) & \quad \text{if } \alpha = 0 \text{ and } \beta > 0, \\
\gamma & \quad \text{if } \alpha = 0 \text{ and } \beta = 0.
\end{align*}
$$

(2.13)

where $\epsilon_b$ is a small positive number. Since the mapping from $[a_1, \ldots, a_n]$ to $[r_1, \ldots, r_n]$ is bijective inside the stability regions, the quadratic function $J_k$ can not be degenerated. Thus the $\alpha = 0$ case can not occur.

We close this section by extending the minimization to the case where the leading polynomial coefficient is not unity. If the unstable polynomial is given by

$$B(z) = b_0z^n + b_1z^{n-1} + \cdots + b_n,$$

(2.14)

and the corresponding stable polynomial is

$$C(z) = c_0z^n + c_1z^{n-1} + \cdots + c_n.$$  

(2.15)

We can write $C(z)$ as

$$C(z) = c_0[z^n + a_1z^{n-1} + \cdots + a_n] = c_0A(z)$$

(2.16)

with $a_i = c_i/c_0$, $i = 0, 1, \ldots, n$. Thus we can determine $a_i$ in terms of Schur parameter vector $r$, and $c_0$ can be determined by minimizing

$$J_k = (c - b)^TW(c - b) = (a^TWa)c_0^2 - 2(a^TWb)c_0 + b^TWb,$$

(2.17)

where $a = [1, a_1, \ldots, a_n]^T$, $b = [b_0, b_1, \ldots, b_n]^T$, and $c = [c_0, c_1, \ldots, c_n]^T$.

Taking the derivative with respect to $c_0$ to equation (2.17) and setting it to zero, we have
\[(a^T Wa)c_0 - a^T Wb = 0, \quad (2.18)\]

which gives
\[c_0 = \frac{a^T Wb}{a^T Wa}. \quad (2.19)\]

The corresponding \(J_k\) function is given by
\[J_k = b^T Wb - f(r_k), \quad (2.20)\]

where
\[f(r_k) = \frac{(a^T Wb)^2}{a^T Wa} = \frac{d_0 r_k^2 + d_1 r_k + d_2}{e_0 r_k^2 + e_1 r_k + e_2}. \quad (2.21)\]

It is obvious that \(d_0 \geq 0, e_0 \geq 0, d_2 \geq 0, e_2 \geq 0, \) and
\[d_0 r_k^2 + d_1 r_k + d_2 \geq 0, \quad e_0 r_k^2 + e_1 r_k + e_2 > 0. \quad (2.22)\]

Minimizing \(J_k\) is equivalent to maximizing \(f(r_k)\) with respect to \(r_k\) over \([-1, 1]\).

Except when \(f(r_k)\) is a constant, we can prove that \(f(r_k)\) has a unique maximum on \([-1, 1]\), and this maximum can be found analytically in a similar manner as before. In this way, we can still make use of the alternating projection algorithm in the more general case.

### 2.5 Examples

Below we present some examples which illustrate the theory discussed above. To study the computation time, we consider searching the minimum in all directions once (\(J_k, k = 1, \ldots, n\)) as one step. In all cases the accuracy is \(\epsilon = 10^{-5}\).

**Example 2.1:** \(n = 2\) and \(W = I\).
All of the examples for $n = 2$ are quite simple. Moreover, the minimum point to the stabilization problem can be found immediately by inspection of Figure 2. For the all $n = 2$ cases we obtain the optimal solution within two iterations of the algorithm.

**Example 2.2:**

Consider the stable polynomial from [66]

$$A(z) = z^4 - 2.7607z^3 + 3.8106z^2 - 2.6535z + 0.9238,$$
Table 1: Solutions to Example 2.2 Using Optimal Method.

<table>
<thead>
<tr>
<th></th>
<th>Before Stabilization</th>
<th>After Stabilization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True Value</td>
<td>Mean</td>
</tr>
<tr>
<td>$a_1$</td>
<td>-2.7607</td>
<td>-2.7621</td>
</tr>
<tr>
<td>$a_2$</td>
<td>3.8106</td>
<td>3.8086</td>
</tr>
<tr>
<td>$a_3$</td>
<td>-2.6535</td>
<td>-2.6555</td>
</tr>
<tr>
<td>$a_4$</td>
<td>0.9238</td>
<td>0.9237</td>
</tr>
<tr>
<td>Avg Distance</td>
<td>$3.9556 \times 10^{-4}$</td>
<td>$2.7901 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

so $a = [-2.7607, 3.8106, -2.6535, 0.9238]^T$.

We add Gaussian noise $N(0, 0.0001)$ to the vector $a$. If the perturbed coefficients give an unstable polynomial, we stabilize it by using the algorithm discussed; if the perturbed coefficients give a stable polynomial, we will not do anything. The weighting matrix $W$ is chosen as the identity matrix, and 50 Monte Carlo simulations are performed.

Table 1 shows the mean and standard deviation of the polynomial coefficients before and after stabilization. In this case, 30 of the 50 polynomials required stabilization. It can be seen that the standard deviation is lower for the stabilized polynomial coefficients. Table 1 also shows the average distance square ($J$) between the true polynomial coefficients and the perturbed polynomial coefficients (before and after stabilization, respectively). For the stabilized case, a smaller distance measure is observed. Figures 3 and 4 show the zero distribution of 50 simulations before and after stabilization, respectively. In this example, the minimization procedure converged to the optimal solutions for 26 out of 30 stabilizations using the modified Schur parameters as the initial conditions. For the case of finding the
Table 2: Solutions to Example 2.2 Using Zero Moving Methods.

<table>
<thead>
<tr>
<th>$a$</th>
<th>True Value</th>
<th>Zero Reflecting</th>
<th>Zero to Unit Circle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Stdev</td>
<td>Mean</td>
</tr>
<tr>
<td>$a_1$</td>
<td>-2.7606</td>
<td>-2.7167</td>
<td>0.0730</td>
</tr>
<tr>
<td>$a_2$</td>
<td>3.8106</td>
<td>3.6874</td>
<td>0.1905</td>
</tr>
<tr>
<td>$a_3$</td>
<td>-2.6535</td>
<td>-2.5348</td>
<td>0.1876</td>
</tr>
<tr>
<td>$a_4$</td>
<td>0.9238</td>
<td>0.8689</td>
<td>0.0833</td>
</tr>
<tr>
<td>Avg Distance</td>
<td>$8.3744 \times 10^{-2}$</td>
<td>$2.2723 \times 10^{-2}$</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: Zero distribution before stabilization of Example 2.2.
Figure 4: Zero distribution after stabilization of Example 2.2, using optimal method.
To compare this stabilization method with others, stabilized the polynomial using two other methods. In the first method, we computed the zeros of the noisy polynomial, and any zeros outside the unit circle were reflected inside; that is, the magnitude of the zero was inverted, and the angle of the zero was left unchanged. We note that this method is the asymptotic result of the Planar Least Squares Inverse (PLSI) techniques applied twice to the polynomial [67, pp. 173–174], [68, pp. 234–236]. In the PLSI technique, a Levinson recursion is used to find a stable inverse polynomial; applying this method twice gives a stable approximant to the original unstable polynomial. The PLSI stabilization method is computationally efficient because we need only compute two Levinson recursions. As the order of the first inverse polynomial approaches infinity, the PLSI method approaches the zero reflecting method discussed above.

The second method of stabilization is similar to the first method except that the magnitude of the zero is set to one instead of being inverted. This method provides the minimum distance stable polynomial, where distance is measured in "zero space".

The results of the above two stabilization methods are shown in Table 2. It can be seen that both of these stabilization methods give much higher distances to the true polynomials than the optimal method does; in this case, the squared error $J$ is about two orders of magnitude higher. Figures 5 and 6 show the zero plots for the two other stabilization methods. Comparing these with the zero plots for the original and optimally stabilized polynomials in Figures 3 and 4, we see similar results, with the optimal method giving slightly tighter zero clusters than
Figure 5: Zero distribution after stabilization of Example 2.2, using zero reflection method.

the other two methods.

Example 2.3:

This example considers stabilization of an estimated polynomial $B(z)$ obtained by AR modeling of a stochastic time series. The data is generated by an AR(4) model with

$$A(z) = z^4 + 0.1z^3 + 1.66z^2 + 0.093z + 0.8649.$$ 

This example was taken from [66]. Thus, the data $x(n)$ is generated by the recursion
Figure 6: Zero distribution after stabilization of Example 2.2, moving unstable zeros to the unit circle.
\[ x(k) = -\sum_{i=1}^{4} a_i x(k - i) + u(k), \]  

(2.23)

where \( \{u(k)\} \) is a \( N(0, 0.5) \) white noise sequence. We generate 50 sets of data points \( \{x(k)\}_{k=0}^{19} \) and from each set we obtain an estimate of the AR parameter vector \( a \) using the covariance method (see [25]). If the estimated polynomial is not stable, we stabilize it using the minimization algorithm.

It is known that such an AR parameter vector estimate is asymptotically a Gaussian distributed random vector. The covariance matrix of the estimate can be found in, e.g. [24, p.212]. For this vector, the natural distance metric is the weighted \( \ell_2 \) norm with the weighting matrix \( W \) is chosen to be the inverse of the asymptotic covariance matrix of AR parameter vector; in this way, the stabilization procedure corresponds to a minimum variance update. Therefore, we have used the inverse of the covariance matrix as the weighting matrix in this example.

Table 3 shows the means and standard deviations of the estimated AR coefficients before and after stabilization (20 of the 50 polynomials were unstable). Table 3 also shows the average distance square \( (J) \) from the true polynomial coefficients to the realized polynomial coefficients before and after stabilization, respectively. For the stabilized case, a smaller distance measure is also observed. The distance measure is only slightly smaller because the variance of the estimates is the dominant factor in this distance.

Figures 7 and 8 shows the zero distribution of 50 simulations before and after stabilization, respectively. While the stabilized zeros are all within the unit circle, there was not much movement needed to change the unstable polynomials to stable ones. In this example, the minimization procedure did not converge to the optimal solution only for one of the 20 stabilizations using the modified Schur parameters as
Table 3: Solutions to Example 2.3 Using Optimal Method.

<table>
<thead>
<tr>
<th>a</th>
<th>True Value</th>
<th>Mean</th>
<th>Stddev</th>
<th>Mean</th>
<th>Stddev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>0.1000</td>
<td>0.0457</td>
<td>0.1896</td>
<td>0.0425</td>
<td>0.1943</td>
</tr>
<tr>
<td>$a_2$</td>
<td>1.6600</td>
<td>1.5601</td>
<td>0.2105</td>
<td>1.5305</td>
<td>0.2161</td>
</tr>
<tr>
<td>$a_3$</td>
<td>0.0930</td>
<td>0.0446</td>
<td>0.2076</td>
<td>0.0425</td>
<td>0.1957</td>
</tr>
<tr>
<td>$a_4$</td>
<td>0.8469</td>
<td>0.7599</td>
<td>0.2434</td>
<td>0.7327</td>
<td>0.2369</td>
</tr>
<tr>
<td>Avg Distance</td>
<td></td>
<td>0.1826</td>
<td></td>
<td>0.1789</td>
<td></td>
</tr>
</tbody>
</table>

the initial conditions. It converged to the optimal solutions for all 20 stabilizations for the initial conditions from the zeros of $B(z)$.

The spectral peaks corresponding to the stabilized polynomials will be very sharp because the stabilized AR estimates have poles on the unit circle. However, if it is known \textit{a priori} that the spectral peaks must have a certain minimum bandwidth, then the stabilization procedure can be modified to give poles which lie in the disk $\{|z| < 1 - \varepsilon\}$ for some appropriate value of $\varepsilon$. Such a modification will be dependent on the particular application of the modeling procedure.
Figure 7: Zero distribution before stabilization of Example 2.3.
Figure 8: Zero distribution after stabilization of Example 2.3, using the optimal method.
2.6 Conclusions

We have considered the problem of finding the closest stable polynomial to a given unstable one. The measure of error between these two polynomials is the weighted Euclidean distance in coefficient space. This problem has no closed form solution in general. We developed an efficient minimization procedure using the alternating projection approach in Schur parameter space. In each step in the iteration we minimize a scalar quadratic function, which is very efficient to implement. As a result, each iteration of the algorithm requires only $4.5n^2 + 1.5n + 2$ multiplications and one division. For the case $W = I$, identity matrix, there are only $1.5n^2 + 1.5n + 2$ multiplications and one division per iteration. Simulation examples illustrate the effectiveness of the algorithm for both a polynomial stabilization application and an Autoregressive (AR) modeling example.

As a final note, we used an $\ell_2$ distance as a measure of closeness in polynomial coefficient space. However, other distance measures could easily be employed in this procedure. If other distance measures are used, the alternating projection approach can still be used, with only a small change in the error minimization procedure for the error function $J(r_k)$. 
CHAPTER III

OPTIMAL NON-NEGATIVE DEFINITE APPROXIMATIONS OF ESTIMATED MOVING AVERAGE COVARIANCE SEQUENCES

3.1 Introduction

There are many problems in which one is interested in obtaining a parametric model of the spectrum of a time series. The autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA) models are widely used in many engineering problems. In obtaining MA and ARMA spectral estimates, a problem which often arises is that of ensuring that the resulting spectral estimate is non-negative and real on the unit circle \([50]\). For example, a commonly used method of MA spectral estimation is to estimate the first \(n + 1\) autocovariances \(\{\gamma_k\}_0^n\) of a time series from some measurements of that time series. Depending on the estimator used for \(\{\gamma_k\}\), the spectral estimate may not be non-negative and real. A similar problem occurs in ARMA spectral estimation algorithms in which the AR parameters are estimated in a first step, and the MA part of the spectrum is estimated using the AR coefficient estimates \([51,52,53,54,44]\).

It is well-known that a necessary and sufficient condition for the MA spectrum to be nonnegative real is that the MA covariance sequence \(\{\gamma_k\}_{k=\infty}^0\) be non-negative definite (NND). Here, the sequence is nonzero only for \(|k| \leq n\) where \(n\) is the order of the MA process. If the MA covariance sequence estimate is not NND, there are various ways in which one can alter the estimate to make it NND. A common procedure entails multiplying the estimated autocovariances by some
window sequence (such as the Bartlett window or an exponential window) [50,54]. For some estimates, the window can be chosen in such a way as to guarantee NND estimates; however, such a window imposes a bias on the resulting estimate [50]. A second approach is to use a data adaptive window, in which a parameter in the window is chosen to ensure NND estimates, with a minimum of bias for that particular window. For example, an exponential window \( w_k = \alpha^{|k|} \) can be used, where \( \alpha \) is chosen by a one-dimensional search procedure to be as small as possible so that the sequence \( \{w_k \gamma_k\} \) is NND. This variable window method imposes less bias than the fixed window method, but it requires iteration to find \( \alpha \).

In this chapter we consider an optimum approach to obtaining a NND covariance sequence. Given an estimated covariance sequence \( \{\gamma_k\}_{k=0}^n \) of an MA time series, we wish to find the closest NND sequence to that estimate, where closeness is measured in terms of a \( \ell_2 \) error norm in coefficient space. This is a nonlinear minimization problem. We discuss necessary and sufficient conditions for the solution to this problem. We then derive an algorithm for finding the global minimum.

This problem is closely related to the approximate stochastic realization problem as considered in [44,47]. In [44,47], the approach taken is to parameterize the covariance sequence in terms of the parameters of an ARMA model which admits this sequence. Then the ARMA model which yields the closest covariance sequence to a given one is found by minimizing an error functional; this involves a nonlinear minimization procedures on the coefficients of the ARMA model.

This work is also related to recent work by Steinhardt and others [38,39,41,42]. In this work the authors have characterized the set of all partial covariance sequences \( \{r_0, \ldots, r_n\} \) which admit an MA or ARMA model of given order. They also develop the expanding hull algorithm to find the smallest order model which
can admit a given partial sequence \( \{r_0, \ldots, r_n\} \) which is NND. The problem considered here is somewhat different: given a sequence \( \{\gamma_0, \ldots, \gamma_n\} \) which is not NND, we wish to find the closest NND sequence \( \{r_0, \ldots, r_n, 0, 0, \ldots\} \) (corresponding to an MA(\( n \)) model).

In this work, the approach taken is to consider the minimization problem in the space of covariance parameters. One reason to consider the problem in covariance space is to gain insight on the geometry of the problem. Specifically, if a covariance sequence is not NND, then its Fourier Transform (power spectral density function) is negative for some frequencies. The closest NND approximant to this sequence will have a Fourier Transform which touches zero at some frequencies. We derive a minimization procedure which is based on locating these zero frequencies. These frequencies correspond to tangent hyperplanes in the set of solutions, and the minimization problem can be seen as a standard orthogonal projection problem onto these hyperplanes. The nonlinearity of this problem arises because these frequencies are not known \textit{a priori}, and must be found using iterative minimization techniques. On the other hand, the number of zero frequencies is small; it is at most \( \lfloor \frac{3}{2} \rfloor + 1 \) where \( n \) is the MA order. Thus, iterative minimization is carried out in a space of lower dimension than if the solution is found in the space of MA parameters as in [47,44].

An outline of this chapter is as follows. In Section 3.2 we present a formal statement of the problem. In Section 3.3 some properties of the non-negativity region are described. In Section 3.4 we derive a solution to the minimization problem which uses the Lagrange method in terms of known zero-spectrum frequencies. In Section 3.5 we further characterize the optimal solution. Section 3.6 presents examples of the algorithm, and Section 3.7 concludes the chapter.
3.2 Problem Statement

Let \( \{\gamma_k\}_{k=0}^n \) denote a sequence of real numbers. This sequence can be thought of as estimates of the first \( n + 1 \) covariances of an MA(\( n \)) process. Consider the function

\[
S_\gamma(z) = \sum_{k=-n}^{n} \gamma_k |z|^{-k}.
\] (3.1)

In order to ensure that \( S_\gamma(z) \) is a valid spectral density function, we must have \( S_\gamma(e^{i\omega}) \geq 0 \), or equivalently, that

\[
g(\omega) = s_0 + s_1 \cos \omega + \cdots + s_n \cos n\omega \geq 0 \quad \forall \omega \in [0, \pi],
\] (3.2)

where \( s_0 = \gamma_0 \), and \( s_k = 2\gamma_k \) for \( k = 1, \cdots, n \). Nearly all covariance estimators guarantee that \( \gamma_0 > 0 \), but often do not guarantee that equation (3.2) is satisfied; thus, we will assume \( s_0 > 0 \) in the following discussion.

Assume condition (3.2) is not satisfied; that is \( \{\gamma_k\} \) is not a NND sequence. In this case, we are interested in finding a covariance sequence which is NND and which is close to the given sequence. To this end, let \( \rho = [\rho_0, \rho_1, \ldots, \rho_n]^T \) and define

\[
f(\omega, \rho) = \rho_0 + \rho_1 \cos \omega + \cdots + \rho_n \cos n\omega.
\] (3.3)

Define the non-negative definite set \( D \) by

\[
D = \{\rho | f(\omega, \rho) \geq 0 \text{ for } \omega \in [0, \pi]\}. \] (3.4)

Then the problem of finding the closest NND sequence can be stated as follows:

**Problem P:** Given a vector \( s = [s_0, s_1, \ldots, s_n]^T \notin D \), find the vector \( \rho^{opt} \in D \) such that \( Q = \|\rho^{opt} - s\|_2^2 \leq \|\rho - s\|_2^2 \) for all \( \rho \in D \), where \( \| \cdot \| \) is the \( \ell_2 \) (Euclidean) vector norm.
Note that even though $s$ and $\rho$ have the same dimension in the above problem statement, one can find the NND solution for a different order than that of $s$. If the closest NND sequence of order $l$ is desired for $l < n$, $s$ is replaced by $[s_0, \ldots, s_l]^T$. If the desired order $l$ is greater than $n$, $s$ is replaced by the $l + 1$-vector $[s_0, \ldots, s_n, 0, \ldots, 0]^T$.

Once $\rho^{opt}$ is found, an MA($n$) filter which realizes $\rho^{opt}$ is obtained by performing a spectral factorization on the function

$$S_{\rho^{opt}}(z) = \sum_{k=-n}^{n} r_k |z|^{-k}, \quad (3.5)$$

where $r_0 = \rho_0^{opt}$ and $r_k = \rho_k^{opt}/2$ for $1 \leq k \leq n$.

There are several ways to approach the solution to this minimization problem. One method involves parameterizing the autocovariance sequence by an underlying model (such as an MA, AR, or ARMA model), and finding the coefficients of this model (which in turn yields the NND covariance sequence) [44,47]. This procedure entails a nonlinear minimization, as the covariances are nonlinear functions of the ARMA parameters. The functional dependence on the AR parameters is highly nonlinear, and convergence to local minima is a problem [44,47]. For the special case of an MA model, $Q$ is quartic in the MA parameters, so we have a quartic minimization problem [44]. In either case, the minimization problem is of dimension equal to the number of ARMA (MA) parameters.

An alternative method is to consider the minimization problem directly in the space of autocovariance coefficients. One reason for studying the problem in this space is to gain insight into its geometry. It turns out that the geometrical properties relate closely with properties of the corresponding spectral density functions associated with the covariances. In fact, we show that if the original (estimated)
set of covariances is not in $D$, then the solution $\rho^{opt}$ has power spectral density which is zero for at least one frequency. We then use this property to characterize possible zero frequencies of $S_{\rho^{opt}}(e^{j\omega})$, and derive a solution to Problem $P$ which is based on these zero frequencies. The resulting minimization algorithm has the property that the minimization is carried out in a space of smaller dimension than if ARMA coefficients are used, and this dimension is related to the way in which the original $\gamma_k$ sequence deviates from NND. The dimension of the minimization is equal to the number of distinct zero frequencies of the optimal solution. For $\gamma_k$ sequences which are "close" to NND, the nonlinear minimization problem has low dimension; the worst case dimension of the minimization is $[\frac{n}{2}] + 1$.

3.3 Description of the Admissible Set $D$

The minimization problem is nontrivial because the set $D$ is a complicated function of the $\rho$ vector. We first establish some properties of $D$.

**Theorem 3.1:**

a) $D \subset R^{n+1}$ is a closed convex cone with vertex at the origin.

b) Let $\partial D$ denote the boundary of $D$. If $\rho \in \partial D$, there is at least one frequency $\omega_0 \in [0, \pi]$ such that $f(\omega_0, \rho) = 0$.

c) If $\rho^* \in \partial D$ and $f(\omega_0, \rho^*) = 0$, then the hyperplane

$$H_{\omega_0} = \{\rho | f(\omega_0, \rho) = 0\}$$

(3.6)

is tangent to $D$.

d) There is a unique solution $\rho^{opt}$ to the minimization problem $P$. 
Define the half spaces $H^+_\omega$

$$H^+_\omega = \{ \rho \in \mathbb{R}^{n+1} | f(\omega, \rho) \geq 0 \} = \{ \rho \in \mathbb{R}^{n+1} | (b(\omega), \rho) \leq 0 \}, \quad (3.7)$$

where

$$b(\omega) = [-1, \cos \omega, \cos 2\omega, \ldots, \cos n\omega]^T \quad (3.8)$$

is the normal vector for $H_\omega$. Then $D$ is the intersection of these half spaces:

$$D = \bigcap_{\omega \in [0, \pi]} H^+_\omega. \quad (3.9)$$

Proof:

a) It is readily verified that if $\rho_1, \rho_2 \in D$, $\alpha \rho_1 + (1 - \alpha) \rho_2 \in D$ for any $\alpha \in [0, 1]$, so $D$ is convex. It is also clear from equation (3.3) that if $\rho \in D$, then $\alpha \rho \in D$ for all $\alpha \geq 0$. Thus $D$ is a convex cone with vertex at the origin.

To show $D$ is closed, consider any $\rho^0 \in D^c$, where $D^c$ is the complement of $D$. Then $f(\omega, \rho^0) < 0$ for some $\omega \in [0, \pi]$. Since $f(\omega, \rho)$ is a continuous function of $\rho$ for any fixed $\omega$, there exists a neighborhood of $\rho^0$ such that $f(\rho, \omega) < 0$ for any $\rho$ in that neighborhood. This implies that $D^c$ is open, and thus $D$ is closed.

b) This follows readily from the definition of $f(\omega, \rho)$.

c) Note that $\rho^* \in H_{\omega_0}$ so $H_{\omega_0}$ intersects $D$. Moreover, $H_{\omega_0}$ cannot contain an interior point of $D$, because every interior point $\rho$ of $D$ has the property that $f(\omega, \rho) > 0$ for all $\omega \in [0, \pi]$. Thus, $H_\omega$ is tangent to $D$.

d) This follows immediately from the fact that $D$ is a closed and convex cone, and $Q$ is a distance function [55, Theorem 1 on page 69].
e) Equations (3.7) and (3.9) follow immediately from the definition of $D$.

We remark that a set which is related to $D$ can be found by considering the slice of $D$ obtained by setting $\rho_0$ to a constant. For any $\rho \in D$ with $\rho_0 > 0$, we can write

$$f(\omega, \rho) = \rho_0 [1 + \eta_1 \cos \omega + \cdots + \eta_n \cos n\omega], \quad \text{(3.10)}$$

where $\eta_i = \rho_i / \rho_0$ for $i = 1, \ldots, n$. We can then define a set $N \subset \mathbb{R}^n$ by

$$N = \{ \eta \mid 1 + \eta_1 \cos \omega + \cdots + \eta_n \cos n\omega \geq 0 \text{ for } \omega \in [0, \pi]\} \quad \text{(3.11)}$$

with $\eta = [\eta_1, \ldots, \eta_n]^T$. The set $N$ is the $\rho_0 = 1$ slice of $D$. This set has been studied in [38]. In particular, it is shown in [38] that $N$ is a compact, convex subset of $\mathbb{R}^n$. One can define the minimization problem in terms on $N$ instead of $D$; but this gives a different solution which has higher error. However, it is convenient to use $N$ instead of $D$ to visualize the geometry of the minimization problem. This is especially true when $n = 3$, because $N$ is a subset of $\mathbb{R}^2$ for this case.

Figure 9 shows the regions $N \subset \mathbb{R}^2$ and $D \subset \mathbb{R}^3$ for $n = 2$. Also shown are three hyperplanes (lines) on $N$ corresponding to three frequency values for which $f(\omega, \rho) = 0$. These three lines correspond to tangent planes of $D$ in $\mathbb{R}^3$, each defined by $\rho_0 + \rho_1 \cos \omega + \rho_2 \cos 2\omega = 0$.

3.4 Solution to the Minimization Problem

One way to solve the minimization problem is to define a grid of points $\omega_1, \ldots, \omega_k$ in the interval $[0, \pi]$ and to find the solution to

$$\rho^* = \arg \min_{\rho \in H_k^+} \|s - \rho\| \quad \text{for } \rho \in H_k^+ = \bigcap_{i=1}^k H_{\omega_i}^+.$$
Figure 9: Non-Negative Definite Regions $N$ and $D$ for $n = 2$. 
In this case, $\rho^*$ is the closest point to $s$ which lies on the supporting polygonal cone which contains $D$. As the number of grid points increases, $\rho^*$ approaches the optimal solution $\rho^{opt}$. The minimization problem (3.12) is a quadratic minimization problem with $k$ linear inequality constraints, and can be solved using standard techniques (see [61,62,63]).

The approach we take is based on this idea, but incorporates some structure of the problem to simplify the minimization. Because $\rho \in \mathbb{R}^{n+1}$, $f(\omega, \rho) = 0$ for at most $\lfloor \frac{n}{2} \rfloor + 1$ distinct frequencies in $[0, \pi]$. As a result, at most $\lfloor \frac{n}{2} \rfloor + 1$ of the linear inequality constraints in (3.12) are active. The approach we take makes use of this fact by using $k$ constraints in (3.12) for $k = 1, \ldots, \lfloor \frac{n}{2} \rfloor + 1$, and by forcing all constraints to be active. By varying the frequencies corresponding to these constraints, we span over all possible points on the boundary of $D$, and thus span over all possible solutions to the minimization problem.

The minimization problem in equation (3.12) can be stated as a constrained minimization problem by defining

$$\omega = [\omega_1, \ldots, \omega_k]$$
$$Q_k(\omega) = \|\rho - s\|^2 - 2 \sum_{j=1}^{k} A_j f(\omega_j, \rho)$$

where each $A_j$ is a Lagrange multiplier. Let $Q_k^*(\omega)$ denote the minimum of $Q_k(\omega)$ with respect to $\rho$ for a given frequency vector $\omega$, and let $\rho^*$ denote the minimum point. Then minimization of $Q_k(\omega)$ gives the point $\rho^*$ which is closest to $s$ under the constraint that $\rho^*$ lies on the hyperplane $H_k = H_{\omega_1} \cap \cdots \cap H_{\omega_k}$, and $Q_k^*(\omega) = \|\rho^* - s\|^2$.

The solution $\rho^*$ to the above problem is the orthogonal projection of the point $s$ onto the hyperplane $H_k$, and can be found by solving a set of linear equations as
The solution to equation (3.15) is given by:

\[
\rho^* = s - C(C^T C)^{-1} g
\]

\[
A^* = -(C^T C)^{-1} g
\]

\[
Q_k^*(\omega) = g^T (C^T C)^{-1} g
\]

where

\[
g = C^T s = [g(\omega_1), \ldots, g(\omega_k)]^T.
\]

Because each \( H_{\omega_i} \) is tangent to \( D \), any solution \( \rho^* \) lies either outside \( D \) or on the boundary \( \partial D \). If the frequency vector \( \omega \) is chosen appropriately, \( \rho^* \) lies in \( \partial D \) and coincides with the optimal solution \( \rho^{opt} \). Below we develop conditions for which \( \rho^* = \rho^{opt} \). From these conditions, we develop an algorithm for finding \( \rho^{opt} \) based on solving the projection problem in equation (3.14). To this end, the following theorems are of interest.

**Theorem 3.2:** Let \( \rho^* \) and \( A^* \) be the solutions to (3.18) and (3.19), and let \( f(\omega, \rho^*) = 0 \) for \( k \) distinct frequencies \( \omega_1, \ldots, \omega_k \). Then \( A_j^* \geq 0 \) for \( j = 1, \ldots, k \).
Remark: Theorem 3.2 states that $s - \rho^*$ is in the convex cone spanned by the normals of the hyperplanes $H_{\omega_i}$ for $i = 1, \ldots, k$.

Proof: This is a standard result from optimization theory (see, e.g. [55, pp. 213–270]); the conditions $A_j^* \geq 0$ for $j = 1, \ldots, k$ are the Kuhn-Tucker necessary conditions.

Theorem 3.3: Let $\rho^{opt}$ be the solution to the minimization problem $P$, and let $\{\omega_1^0, \ldots, \omega_k^0\}$ be the set of $k$ distinct frequencies for which $f(\omega, \rho^{opt}) = 0$. Then the functional $Q_k^*(\omega)$ in equation (3.20) has a local maximum at the point $\omega = (\omega_1^0, \ldots, \omega_k^0)$.

Proof: According to the duality principle [55], the minimum distance between a convex set and a point outside this set is equal to the maximum of the distances from the point to supporting hyperplanes separating the point and the convex set. Figure 10 shows this principle. In addition, the point $\rho^*$ (and consequently, the distance $Q_k^*(\omega)$ between $s$ and $\rho^*$) is completely parameterized by the zero frequencies of $f(\omega, \rho^*)$. Thus, varying $\rho$ to minimize $||\rho - s||^2$ is equivalent to varying the zero frequencies $\omega_1, \ldots, \omega_k$ to find local maxima of $Q_k^*(\omega)$.

We illustrate the above theorems by means of a simple example on the set $N$ as shown in Figure 11. Consider first the point $a$ as the given covariance vector $s$. For any frequency $\omega \in [0, \pi]$, the point $\rho^*(\omega)$ is the orthogonal projection of $a$ onto the hyperplane $H_{\omega}$, which in this case is a line. The hyperplanes $H_0$ and $H_\pi$ are shown, and for $\omega \in (0, \pi)$ the hyperplane $H_{\omega}$ is the tangent line to $D$, which rotates clockwise from $H_0$ around to $H_\pi$ as $\omega$ increases. The value of $Q^*(\omega)$ is the squared distance between $a$ and the hyperplane $H_{\omega}$. For the point $a$, this distance achieves a local maximum at $\omega = \pi/4$ and at $\omega = 3\pi/4$; however, $A^*(\pi/4) > 0$ and
Figure 10: The duality principle for the minimal distance from a point outside a convex set to the set.
Figure 11: Solutions on $N$ for $n = 2$.

$A^*(3\pi/4) < 0$ because $a$ and $D$ are on opposite sides of the hyperplane $H_{\pi/4}$ but on the same side of $H_{3\pi/4}$. The projection $\rho^*(\pi/4)$ is the point $b$, and this is also $\rho^{opt}$ for this problem.

Next, consider the point $c$. For this point $Q^*(\omega)$ has a local maxima at $\omega = 0$, $\pi/2$, and $\pi$, and the corresponding $\rho^*$ points are marked $d$, $e$, and $f$, respectively. For points $d$ and $f$, the corresponding $A^*(\omega) \geq 0$, but they are outside $D$; the point $e$ is on the boundary $\partial D$, but the corresponding $A^*(\omega) < 0$, so none of these points are admissible solutions. Thus, we must consider the two constraint problem. The
only possible two-constraint set for this order is \( \omega = 0, \omega = \pi \), and \( H_0 \cap H_\pi \) gives the point \( g \), which is the optimum solution \( \rho^{opt} \) for this case.

The above theorems and examples motivate a method for finding the optimum solution to the minimization problem \( P \) via projections onto the tangent subspaces \( H_k \), as given by the following theorem.

**Theorem 3.4:** Let \( \omega^* = [\omega_1, \ldots, \omega_k] \) be a set of \( k \) distinct frequencies, each in the interval \([0, \pi]\). Assume

1) \( Q_k(\omega) \) achieves a local maximum at \( \omega^* \)

2) \( A_j^* \geq 0 \) for \( j = 1, \ldots, k \)

3) \( \rho^* \) corresponding to \( \omega^* \) is an element of \( D \).

Then \( \rho^* = \rho^{opt} \).

**Proof:** We know that

\[
\|s - \rho^*\| \leq \|s - \rho\| \quad \forall \rho \in H_k^+,
\]

with equality if and only if \( \rho = \rho^* \). We also know that \( D \subset H_k^+ \). These two statements imply

\[
\|s - \rho^*\| \leq \|s - \rho\| \quad \forall \rho \in D.
\]

But the above statement is the definition of \( \rho^{opt} \), and since there is a unique solution to the minimization problem, \( \rho^* = \rho^{opt} \).

**Description of the Algorithm**

The above theorems provide a means for finding \( \rho^{opt} \). Start with one frequency constraint \((k = 1)\) in equation (3.14). Find local maxima of \( Q^*(\omega_1) \) in equation
(3.20) and corresponding $\rho^*$ and $A^*$ in equations (3.18) and (3.19) (note that this involves a nonlinear maximization in one variable $\omega_1 \in [0, \pi]$). For each local maximum, check conditions 2) and 3) of Theorem 3.4. If both are satisfied, $\rho^* = \rho^{opt}$. If not, increment $k$ and continue.

Note that if $n$ is even, $f(\omega, \rho) = 0$ for at most $\lfloor \frac{n}{2} \rfloor$ distinct frequencies in $(0, \pi)$. There can be up to $\lfloor \frac{n}{2} \rfloor + 1$ distinct zeros in $[0, \pi]$, but only if $f(\omega, \rho) = 0$ at both $\omega = 0$ and $\omega = \pi$. Thus, the maximum value of $k$ is $\lfloor \frac{n}{2} \rfloor + 1$, and the dimension of the nonlinear maximization problem is at most $\lfloor \frac{n}{2} \rfloor$. Similarly, if $n$ is odd there are at most $\lfloor \frac{n}{2} \rfloor$ zeros of $f(\omega, \rho)$ in $(0, \pi)$, and $k$ is at most $\lfloor \frac{n}{2} \rfloor + 1$. In either case, then, $k \leq \lfloor \frac{n}{2} \rfloor + 1$. Also, because of the symmetry of $Q_k^*(\omega)$ with respect to interchanging of two frequencies, it suffices to find maxima of $Q_k^*(\omega)$ on the set

$$I_k = \{\omega \in [0, \pi] \big| \omega_1 \in [0, \pi], \omega_2 \in [0, \omega_1), \ldots, \omega_k \in [0, \omega_{k-1})\}. \quad (3.24)$$

The set $I_k$ can be further restricted as described later in Theorem 3.6. Finally, we mention that in our implementation of the above procedure we used the alternative maximization method as described in [64] to perform the nonlinear maximization step, although other methods could be used.

One step in the procedure requires testing if $\rho^* \in D$. This test can be implemented using a Schur-Cohn algorithm as we now describe. From $\rho^*$ form the $2n$ degree polynomial $z^n S_{\rho^*}(z)$ as defined in equation (3.5). This polynomial has zeros at $e^{\pm j \omega_i}$ for $i = 1, 2, \ldots, k$, so we divide this polynomial by

$$C(z) = \prod_{i=1}^{k} \left( z^2 - 2 \cos \omega_i z + 1 \right) \quad (3.25)$$

to form the remainder polynomial $R(z)$. Now, $\rho^* \in D$ if $R(z)$ has no zeros of odd multiplicity on the unit circle. (In fact, $R(z)$ will have no zeros on the unit
circle except in the rare case that the optimum solution is found using \( k \) frequency constraints, when in fact there are more than \( k \) zero frequencies; this is a degenerate case, as it occurs for points \( s \) on a set of measure zero.) The needed zero test for \( R(z) \) can be implemented using a Schur-Cohn test; see \([57,65]\) for details.

We summarize the above discussion by a concise statement of the algorithm.

1. Set \( k = 1 \).

2. Form \( Q^*(\omega) \) in equation (3.20). Find local maxima of \( Q^*(\omega) \) on the region \( I_k \) defined in (3.24) (using, for example, the alternative maximization method described in \([64]\)). For each local maximum:

   (a) Find \( A^* \) using equation (3.19) and check condition 2) of Theorem 3.4.

   (b) Find \( \rho^* \) using equation (3.18) and check if \( \rho^* \in D \) using the Schur-Cohn procedure described above.

3. If an admissible solution is found in the previous step, it is the optimal solution \( \rho^{opt} \). If not, increment \( k \) and go to Step 2.

One important feature of the above algorithm is that the nonlinear maximizations are carried out in low dimensional space. This is in contrast to methods which parameterize the covariances in terms of MA or ARMA parameters, where the dimension of the minimization problem is fixed at \( n \). The procedure presented above starts with \( Q^*(\omega_1) \); that is, we maximize a function over a single dimensional variable \( \omega_1 \), which itself lies in a compact region \( \omega_1 \in [0, \pi] \). If an admissible solution is found using \( k \) constraints, this solution is the optimum one (by Theorem 3.4) and we need not search for a solution using a larger number of constraints. For many problems, admissible solutions are found for small numbers of constraints.
k, and thus the nonlinear maximization is a low dimensional problem. For an MA order of n, maximization of \( Q_k^*(\omega) \) is a maximization over at most \( \lfloor \frac{n}{2} \rfloor + 1 \) variables, compared to an \( n + 1 \) dimensional minimization if MA parameters are used as in [44].

3.5 Further Characterization of Solutions

In this section we consider some geometrical properties of the solution to the minimization problem \( P \). These properties provide some insight into the spectral properties of the optimum solution with respect to \( s \), or equivalently, to \( g(\omega) \). These properties lead to additional necessary conditions on the solution to the minimization problem, and these conditions can be used to reduce the regions over which one needs to find local maxima of \( Q^*(\omega) \).

3.5.1 Variance Bounds

An immediate result of Theorem 3.2 is given below:

**Corollary 3.1:** Assume \( s \notin D \) and \( p^{opt} \) is the optimal solution to the minimization problem \( P \). Then \( p_0^{opt} > s_0 \).

**Proof:** From equations (3.18) and (3.19) we have \( p_0^* - s_0 = \sum_{i=1}^{k} A_i^* \). Since \( A_i^* \geq 0 \) for \( i = 1, \ldots, k \), and not all \( A_i^* = 0 \), the result follows.

The above corollary shows that the estimated variance of the MA(\( n \)) process is always increased to arrive at the closest NND covariance sequence. An upper bound on this variance can also be obtained. If \( M = - \min_{0 \leq \omega \leq \pi} g(\omega) \), then \( p = [s_0 + M, s_1, \ldots, s_n] \) is an admissible solution with cost error \( Q = M^2 \); thus we have

\[
s_0 < p_0^{opt} \leq s_0 + M. \tag{3.26}
\]
Figure 12: The variance bounds for the $n = 1$ case.
This case is shown on Figure 12.

3.5.2 Characterization of Zero Frequencies

It is of interest to obtain properties of the solution to the minimization problem in terms of the original vector $s$ or, equivalently, in terms of $g(\omega)$. In particular, the solution to the minimization problem can be simplified if we find restrictions on the regions of frequencies for which $f(\omega, \rho^{opt}) = 0$. To this end, we consider the following two conjectures:

Conjecture C1: If $f(\omega_0, \rho^{opt}) = 0$, then $g(\omega_0) \leq 0$.

Conjecture C2: If $g(\omega) < 0$ for $\omega \in (a, b)$, and if $g(a) = g(b) = 0$, then $f(\omega, \rho^{opt}) = 0$ for at least one $\omega \in [a, b]$.

Simply stated, these conjectures say that the spectral density function corresponding to the optimum solution to the minimization problem is zero at $\omega_0$ if and only if $g(\omega)$ is negative there. Conjectures C1 and C2 seem reasonable from an approximation point of view. In fact, as the MA order $n \to \infty$, the solution to the minimization problem approaches

$$f_\infty(\omega, \rho^{opt}) = \max\{g(\omega), 0\} \text{ for } 0 \leq \omega \leq \pi \quad (3.27)$$

and in this case, both conjectures are satisfied. It turns out that for finite MA order $n$, neither conjecture is true in general. We discuss each conjecture below.

Conjecture C1

In general, it is not true that if $f(\omega, \rho^{opt}) = 0$ at some frequency $\omega_0$, then $g(\omega_0) < 0$. A counterexample for $n = 3$ is given by

$$s = [3.8, -6.16, 3.34, -1.]^T.$$
Figure 13: The plots of $g(\omega)$ and $f(\omega, \rho^{opt})$ for $n = 3$ with $s = [3.8, -6.16, 3.34, -1]^T$. Note that $g(70.126^\circ) > 0$.

In this case, optimal NND solution is found to be

$$\rho^{opt} = [3.80538, -6.15485, 3.34475, -0.99528]^T.$$ 

It is readily verified that $f(\omega, \rho^{opt}) = 0$ for $\omega = 70.126^\circ$, but $g(70.126^\circ) > 0$. This case is plotted on Figure 13.

The reason that Conjecture C1 does not hold in general results from acute angles on the boundary of $D$. A simplified sketch of the problem is shown in Figure 14. In this figure, $H_{\omega_1}$ is the hyperplane corresponding to frequency $\omega_1$;
that is, \( H_{\omega_1} \) is the set of all points \( \rho \) for which \( f(\omega_1, \rho) = 0 \). \( H_{\omega_2} \) is similarly defined. It can be seen that \( \rho^{\text{opt}} \) is the closest point in \( D \) to \( s \), and that \( f(\omega, \rho^{\text{opt}}) = 0 \) at frequencies \( \omega_1 \) and \( \omega_2 \). Also, \( g(\omega_1) < 0 \) and \( g(\omega_2) > 0 \), which follows from the fact that \( s \) is on the same side of \( H_{\omega_2} \) as \( D \), but on the opposite side of \( H_{\omega_1} \) from \( D \).

If the angle between \( H_{\omega_1} \) and \( H_{\omega_2} \) were not acute, then this situation could not arise. The angle is not acute if and only if \( \langle b(\omega_1), b(\omega_2) \rangle > 0 \) where \( b(\omega) \) is the normal to the hyperplane \( H_\omega \).

With the above condition in mind, we can state more precise versions of conjecture C1 in the following theorem and corollary:

**Theorem 3.5:** Let \( \rho^{\text{opt}} \) be the solution to the minimization problem \( P \), and assume that \( f(\omega, \rho^{\text{opt}}) = 0 \) for \( k \) distinct frequencies \( \omega_1, \ldots, \omega_k \). If \( \langle b(\omega_i), b(\omega_j) \rangle \geq 0 \) for \( j = 1, \ldots, k \), then \( g(\omega_i) \leq 0 \).

**Proof:** Taking the partial derivative of equation (3.14) with respect to \( \rho_i \) for \( i = 0, \ldots, n \), we obtain

\[
A_i^* b(\omega_1) + \ldots + A_k^* b(\omega_k) = A_i^* b(\omega_i),
\]

(3.28)

Since \( \langle b(\omega_i), \rho^* \rangle = 0 \) and \( \langle s, b(\omega_j) \rangle = -g(\omega_j) \), we have

\[
g(\omega_i) = -\sum_{j=1}^{K} A_j^* \langle b(\omega_i), b(\omega_j) \rangle.
\]

(3.29)

Since \( A_j^* \geq 0 \) and \( \langle b(\omega_i), b(\omega_j) \rangle \geq 0 \) for \( j = 1, \ldots, k \), the result follows from equation (3.29).

An immediate result of Theorem 3.5 is:

**Corollary 3.2:** Under the assumptions of Theorem 3.2, there is at least one zero frequency \( \omega_i \) such that \( g(\omega_i) < 0 \). Furthermore, if \( k = 1 \), then \( g(\omega_1) < 0 \).
Figure 14: Geometric description for Conjecture C1.
Proof: Equation (3.28) implies
\[ \langle s, s - \rho^* \rangle = \sum_{i=1}^{k} A_i^*(s, b(\omega_i)). \] (3.30)

Using the formulas \( Q^*(\omega) = \langle s - \rho^*, s - \rho^* \rangle = \langle s, s - \rho^* \rangle \) and \( g(\omega_i) = -\langle s, b(\omega_i) \rangle \),
we have
\[ Q^*(\omega) = \sum_{i=1}^{k} A_i^*[-g(\omega_i)]. \] (3.31)

Since \( Q^*(\omega) > 0 \) and \( A_i^* \geq 0 \), we have the result.

Theorem 3.5 provides an equation which restricts the frequency intervals over which \( Q^*(\omega) \) must be maximized. However, these restrictions are difficult to use in practice, except in simple cases. From Corollary 3.2, we can restrict \( \omega_1 \) to be in a region for which \( g(\omega) \leq 0 \), and this restriction is simple to implement. We have found that the overhead in computing the general frequency restrictions provided in Theorem 3.5 is too high to be of practical use in the minimization algorithm.

A less restrictive theorem characterizing the zero frequency locations is given below:

**Theorem 3.6:** Let \( \rho^{opt} \) be the solution to the minimization problem \( P \), and assume that \( f(\omega, \rho^{opt}) = 0 \) for \( k \) distinct frequencies \( \omega_1, \ldots, \omega_k \). Define
\[ \delta_n = \min_{0 \leq \omega_1, \omega_2 \leq \pi} \left( \frac{b(\omega_1)}{\|b(\omega_1)\|}, \frac{b(\omega_2)}{\|b(\omega_2)\|} \right). \] (3.32)

Note that \( \delta_1 = 0 \), \( \delta_n < 0 \) for \( n > 1 \) and depends only on the model order \( n \). Then
\[ g(\omega_i) \leq -M \delta_n \|b(\omega_i)\|. \] (3.33)

Proof: Equation (3.32) states that the cosine of the angle between the \( b(\omega_i) \) vectors is no less than \( \delta_n \). From Theorem 3.2 we know that the vector \( s - \rho^{opt} \) is in the convex cone spanned by the \( b(\omega_i) \) vectors. These two statements imply
Table 4: Values of $\delta_n$ for Various Orders $n$

<table>
<thead>
<tr>
<th>$n$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-\delta_n$</td>
<td>0</td>
<td>0.07</td>
<td>0.14</td>
<td>0.18</td>
<td>0.22</td>
<td>0.24</td>
<td>0.26</td>
<td>0.28</td>
<td>0.29</td>
<td>0.30</td>
</tr>
<tr>
<td>$n$</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>$-\delta_n$</td>
<td>0.31</td>
<td>0.32</td>
<td>0.22</td>
<td>0.33</td>
<td>0.24</td>
<td>0.23</td>
<td>0.24</td>
<td>0.25</td>
<td>0.24</td>
<td>0.25</td>
</tr>
</tbody>
</table>

\[
\left\langle \frac{s - \rho^\text{opt}}{\|s - \rho^\text{opt}\|}, \frac{b(\omega_i)}{\|b(\omega_i)\|} \right\rangle \geq \delta_n, \quad (3.34)
\]
equivalently,

\[-g(\omega_i) + f(\omega_i, \rho^\text{opt}) \geq \delta_n \|s - \rho^\text{opt}\| \cdot \|b(\omega_i)\| \quad (3.35)\]
or

\[g(\omega_i) \leq -\delta_n \|s - \rho^\text{opt}\| \cdot \|b(\omega_i)\| \leq -\delta_n M \|b(\omega_i)\|. \quad (3.36)\]

Thus we have proved the theorem.

Note that $1 \leq \|b(\omega_i)\| \leq \sqrt{n + 1}$. Note also that if $s$ is close to $D$, $M$ is small.

Thus for a good estimator $s$, the zero frequencies are mainly in the intervals where $g(\omega) \leq 0$. Table 4 gives the $\delta_n$ values for $n = 1, \ldots, 20$

**Conjecture C2**

Conjecture C2 does not hold in general, as the following example shows. For $n = 2$ and $s = [1, 1.5, -3]^T$, it is readily verified that $g(\omega) < 0$ for $\omega \in [0, 180^\circ)$ and $\omega \in (134.5^\circ, 180^\circ]$ (see Figure 15). The optimal solution is found to be $\rho^\text{opt} = [13/6, 1/3, -11/6]^T$; $f(\omega, \rho^\text{opt}) = 0$ has a zero at $\omega_1 = 180^\circ$. The loss function $Q^* = 4.0833$ and $A^* = 7/6$. The solution obtained by enforcing two constraints
Figure 15: The plots of \( g(\omega) \) and \( f(\omega, \rho^{opt}) \) for \( n = 2 \) and \( s = [1, 1.5, -3]^T \). Note that if \( g(\omega) < 0 \) on an interval, it is not necessarily true that \( f(\omega, \rho^{opt}) = 0 \) on that interval.

at \( \omega_1 = 0 \) and \( \omega_2 = 180^\circ \) gives \( \rho = [2, 0, -2]^T \) with \( Q^* = 4.25, A_1^* = -0.25, \) and \( A_2^* = 1.25 \), so this solution is not optimal.

Figure 16 shows the geometric situation which allows Conjecture C2 to be violated. Here \( g(\omega_1) < 0 \) and \( g(\omega_2) < 0 \) because \( s \) is on the opposite side of the hyperplanes \( H_{\omega_1} \) and \( H_{\omega_2} \) from \( D \). However, the projection onto \( D \) is such that \( \rho^{opt} \) is on the other side of \( H_{\omega_1} \), so \( f(\omega_1, \rho^{opt}) > 0 \). We note that by considering this problem in \( N \), the optimum solution would have zeros at \( \omega = 0 \) and \( \omega = \pi \) (as
this problem corresponds to point $g$ in Figure 11. Thus, the optimal solutions using $D$ or $N$ as the constraint set give results with different geometrical properties.

3.6 Examples

Below we consider three examples which illustrate the algorithm for finding the solution to the minimization problem $P$.

**Example 3.1:** $n = 2$, $s = [1, 2, 3]^T$

For $n = 2$, $f(\omega, \rho^\text{opt})$ has at most two distinct zero frequencies. We first consider $k = 1$ constraint. A plot of $Q_1^*(\omega)$ is shown in Figure 17. It can be seen that this function has a local maximum at $\omega_1 = 102.85^\circ$. From equations (3.18)-(3.20) we find that

$$\rho^* = [2.154, 1.743, 1.960]^T, \quad A^* = 1.154, \quad Q_1^*(\omega_1) = 2.479.$$  

Since $A_1^* \geq 0$ and $f(\omega, \rho^*) \geq 0$, it follows from Theorem 3.4 that $\rho^* = \rho^\text{opt}$. The functions $g(\omega)$ and $f(\omega, \rho^\text{opt})$ are shown in Figure 17.

Note that $g(\omega)$ has only one negative interval, given by $\omega \in (64.26^\circ, 140.14^\circ)$, and it follows from Corollary 3.2 that the zero frequency corresponding to the optimal solution must lie in this region.

**Example 3.2:** $n = 3$, $s = [1, -2, 0, 0]^T$

In Example 3.2, we have appended two zeros to $s$ to find the closest 3rd order NND sequence to a given first order sequence which is not NND. For $n = 3$, $f(\omega, \rho^\text{opt})$ has at most two distinct zero frequencies.

From $s$ we find $g(\omega)$ as shown in Figure 18. We first consider $k = 1$ constraint, and $Q_1^*(\omega)$ is also shown in Figure 18. Note that $g(\omega)$ is negative for $\omega \in [0, 60^\circ)$,
Figure 16: Geometric description for Conjecture C2.
Figure 17: Example 3.1 Original, Optimal NND and Distance Measure functions.
so by Corollary 3.2, \( f(\omega, \rho^{opt}) = 0 \) somewhere in this interval. Maximization of \( Q_1^*(\omega) \) over the interval \([0, 60^\circ]\) gives \( \omega_1 = 23.47^\circ \), with corresponding \[
\rho^* = [1.345, -1.684, 0.235, 0.116]^T, \quad A^* = 0.345, \quad Q_1^*(\omega_1) = 0.288.
\]

This solution is admissible and in \( D \), so it is the optimal solution \( \rho^{opt} \). \( f(\omega, \rho^{opt}) \) is also shown in Figure 18.

**Example 3.3**: \( n = 4, \ s = [1, 2, 3, 4, 5]^T \)

This case is shown in Figure 19. For \( n = 4, \ f(\omega, \rho^{opt}) \) has at most three distinct zero frequencies. According to Theorem 3.6, the possible zero frequencies should
be in the intervals where \( g(\omega) \leq -M\delta_4\|b(\omega)\| \) with \( M = 6.603 \) and \( \delta_4 = -0.18 \). The location function \( L(\omega) = g(\omega) - [-M\delta_4\|b(\omega)\|] \) is also plotted in Figure 19. Thus, the zero frequencies should be in the intervals where the function \( L(\omega) \) is non-positive. In this case there are only two non-positive intervals. Since these two intervals do not contain \( \omega = 0 \) or \( 180^\circ \), \( f(\omega, \rho^{opt}) \) has at most two distinct zero frequencies.

Using \( k = 1 \) constraint gives no admissible solutions at maxima of \( Q^*_1(\omega) \). With \( k = 2 \) constraints (placing the two possible zero frequencies in the two non-positive intervals, respectively), a local maximum of \( Q^*_2(\omega) \) is found at \( (\omega_1, \omega_2) = (55.6^\circ, 138.3^\circ) \), with corresponding \( A^* = [2.120, 0.436]^T \), \( Q^*_2(\omega) = 33.75 \), and

\[
\rho^* = [3.556, 2.872, 2.283, 2.187, 3.011]^T.
\]

Again according to Theorem 3.4 this is the optimal solution.

### 3.7 Conclusions

We have considered the problem of finding the closest non-negative definite MA covariance sequence to a given estimate which may not be non-negative definite. We developed an algorithm which is based on a set of constrained minimization problems, each parameterized by the zero frequencies of the spectral density function corresponding to the optimal solution. The algorithm entails first solving a simple minimization problem with linear constraints whose closed-form solution is given by a projection onto a subspace. These solutions lie either outside the set of NND sequences, or on its boundary; if the solution lies on the boundary, it is the optimal solution.
Figure 19: Example 3.3 Original, Optimal NND and Location functions.
One property of this algorithm is that we consider the problem directly in the space of covariance sequence elements. As a result, the nonlinear maximization step is performed on sets of low dimension (up to $\left\lfloor \frac{n}{2} \right\rfloor$, where $n$ is the MA order). In addition, by considering the minimization problem in this space, we were able to characterize some of the geometrical properties of the optimal solution in terms of the locations of its zero frequencies.
CHAPTER IV
MINIMAL ORDER STOCHASTIC PARTIAL REALIZATION PROBLEM

4.1 Introduction

A fundamental problem in engineering is that of obtaining a parametric model of a time series from a finite set of measurements. For stochastic systems, one of the most widely used model is the so-called Autoregressive Moving Average (ARMA) model of order \( m \), in which time series \( \{ x(k) \} \) is described as the output of a finite order filter with system function

\[
H_m(z) = \frac{B_m(z)}{A_m(z)} = \frac{b_0 + b_1z^{-1} + \cdots + b_mz^{-m}}{1 + a_1z^{-1} + \cdots + a_mz^{-m}}
\] (4.1)

which is driven by zero mean, unit variance white noise (see Figure 1 in Chapter I).

Assume that we are given a finite set \( C^{(n)} = \{ c_0, c_1, \cdots, c_n \} \) of noise-free covariances which correspond to the output of the system. Based on this partial information, what can we say about the underlying system? Can we ascertain the possible minimum order of the system, or give bounds on the power spectrum of \( \{ x(k) \} \)? Such information would be available from the solution to the so-called Stochastic Partial Realization (SPR) problem, which can be stated as follows.

Given a sequence of real numbers \( C^{(n)} = \{ c_0, c_1, \cdots, c_n \} \), find an extension \( c_{n+1}, c_{n+2}, \cdots \) of \( C^{(n)} \) and a corresponding function \( v(z) \) such that;
Here the condition 1) guarantees that \( v(z) \) is consistent with the given sequence \( C(n) \), condition 2) guarantees that the realization is stable, and condition 3) guarantees that the realization is nonnegative definite (NND). If a function \( v(z) \) satisfy conditions 2) and 3), we also say that \( v(z) \) is positive real (PR).

This problem is usually called Carathéodory Extension Problem (CEP) [83]. It was proved [84] that a solution to the CEP exists if and only if the corresponding Toeplitz matrix of \( C(n) \) is nonnegative, that is,

\[
T_n = \begin{bmatrix}
c_0 & c_1 & \cdots & c_n \\
c_1 & c_0 & \cdots & c_{n-1} \\
\vdots & \vdots & \ddots & \vdots \\
c_n & c_{n-1} & \cdots & c_0
\end{bmatrix} \geq 0. \tag{4.5}
\]

Perhaps, the most natural and useful class of solutions of the CEP for \( C(n) \) is the class of rational functions. Under the condition \( T_n > 0 \), there exist infinitely many solutions in this class. In the present study, we are interested in finite order extensions (strictly speaking for \( T_n \)). For this case, we also require that \( v(z) \) satisfy

\[
4) \quad v(z) + v(z^{-1}) = \frac{B_m(z)B_m(z^{-1})}{A_m(z)A_m(z^{-1})} \tag{4.6}
\]

The stochastic partial realization problem is closely related to the problem of spectral estimation. If \( \{x(k)\} \) is a weakly stationary stochastic process with covariance sequence \( \{c_k\} \), then its power spectral density is given by
\[ P_x(e^{j\omega}) = v(e^{j\omega}) + v(e^{-j\omega}) = H_m(e^{j\omega})H_m(e^{-j\omega}) = |H_m(e^{j\omega})|^2. \] (4.7)

An important problem in spectral estimation is that of characterizing the set of all power spectral density function \( P_x(e^{j\omega}) \) of order \( m \) which are consistent with \( C^{(n)} \). This problem is equivalent to the SPR problem. Thus, advances in the SPR knowledge base have direct applications in spectrum estimation.

In general, we have the following theorem [85]:

**Theorem 4.1:** The power series \( v(z) = \frac{1}{2}c_0 + \sum_{k=1}^{\infty} c_k z^{-k} \) converges outside the unit circle \(|z| = 1\) and has there a positive real part if and only if the Toeplitz matrix \( T_n \) (for \( n = 0, 1, \cdots \)) are positive; The function \( v(z) \) is identically zero if and only if all determinants \( D_n \) of \( T_n \) vanish; If \( D_n > 0 \) for \( n \leq m - 1 \) and \( D_n = 0 \) for \( n \geq m \), then \( v(z) \) has the form

\[ v(z) = \sum_{k=1}^{m} \rho_k \frac{z^{m+k}}{z^{-k}} \quad m \geq 1 \] (4.8)

where \( \rho_k > 0 \), \(|\varepsilon_k| = 1\) and \( \varepsilon_k \neq \varepsilon_l \) if \( k \neq l \).

The last part occurs if and only if \( \{x(k)\} \) with covariances \( c_i = E\{x(k)x(k+i)\} \) can be realized by superposition of sinusoidal signals with frequencies determined from the poles of \( v(z) \).

Here, we use an example to show the last part of Theorem 4.1.

**Example 4.1:**

Given a sequence 5, 2, 5, 2, 5, 2, \cdots, it is easy to see that

\[ D_0 = 5, \quad D_1 = 21, \quad D_n = 0 \quad \text{for} \ n \geq 2 \]

So \( m = 2 \) and the corresponding function \( v(z) \) in CEP is

\[ v(z) = \frac{2.5z^2 + 2z + 2.5}{z^2 - 1} = 1.75 \frac{z + 1}{z - 1} + 0.75 \frac{z - 1}{z + 1}. \]
It is easy to show that

\[ x(k) = \sqrt{3} \cos(k\pi + \phi_1) + \sqrt{7} \cos(2\pi k + \phi_2), \]

where \( \phi_1 \) and \( \phi_2 \) are independent random variables distributed uniformly on \([0, 2\pi]\), respectively, and

\[ c_k = 1.5 \cos(k\pi) + 3.5 \cos(2\pi k). \]

Before going further, we use the following example to show the difficulties of the stochastic partial realizations.

**Example 4.2:**

Assume \( c_0 = 1 \), \( c_1 \), and \( c_2 \) are two arbitrary covariances. Find the conditions on \( c_1 \) and \( c_2 \) so we can realize \( c_0, c_1, \) and \( c_2 \) with rational function \( v(z) \) of order 1.

Here we have

\[ v(z) = \frac{d_0 z + d_1}{z + a_1} = \frac{1}{2} + c_1 z^{-1} + c_2 z^{-2} + \cdots. \]  \hspace{1cm} (4.9)

This gives

\[ d_0 = \frac{1}{2}, \quad d_1 = c_1 + \frac{1}{2} a_1, \quad c_2 + c_1 a_1 = 0. \]  \hspace{1cm} (4.10)

\( a_1 \) is found as \( a_1 = -c_2/c_1 \). It is obvious that the function \( v(z) \) is stable if and only if \( |c_2| < |c_1| \).

From \( v(e^{j\omega}) + v(e^{-j\omega}) \geq 0 \), we have

\[ |c_1 - \frac{c_2}{c_1}| \leq 0.5 - c_2 + \frac{c_2^2}{2 c_1^2}. \]  \hspace{1cm} (4.11)

The region for \( c_1 \) and \( c_2 \) in which we can realize \( c_0, c_1, c_2 \) with a positive real \( v(z) \) of order 1 is shown in Figure 20.
Figure 20: PR region for \( m = 1 \).
From this example, it is obvious that the region in which $C^{(2)}$ corresponding a first order (minimal order) realization is not only very limited but also very complicated.

4.2 Solutions to SPR

4.2.1 The Schur Parameters

A convenient description of autocovariance sequences makes use of the so-called Schur parameters. These parameters are defined as follows. Let us consider a positive sequence $C^{(n)}$. The positivity of $C^{(n)}$ implies that the corresponding Toeplitz matrix is positive. We define a sequence of real numbers $\{r_i\}_{i=1}^n$ by

$$r_i = \frac{(-1)^{i-1}}{D_{i-1}},$$

where $D_{i-1}$ is determinant of $T_{i-1}$. The numbers $r_i$ are called the Schur parameters associated with $C^{(n)}$. The importance of the Schur parameters rests on the following algebraic identity [27,29]

$$D_n = D_{n-1} \cdot \sigma_n^2$$

with

$$\sigma_n^2 = \prod_{k=1}^n (1 - r_k^2).$$

From this identity it follows that the sequence $C^{(n)}$ is positive only if $|r_k| < 1$ for each $k > 0$. Conversely, a sequence $\{r_k\}_{k=1}^n$ of real numbers satisfying $|r_k| < 1$
1 for all k, along with a positive real number c_0, determines a unique positive sequence C^{(n)} through equation (4.12). Thus, for any n > 0, there is a one-to-one correspondence between the positive sequence C^{(n)} and the set \{c_0, r_1, \cdots, r_n\} where |r_k| < 1 for all 1 \leq k \leq n. Similar relationships exist between nonnegative autocovariance sequences and Schur parameters which are less than or equal to one in magnitude.

The application of the above result to the SPR problem is immediate. Given any partial covariance sequence C^{(n)}, with corresponding Schur parameters \{r_i\}_{i=1}^{n}, the set of all positive definite extensions to C^{(n)} is given by the set of all extensions to the Schur parameter sequence \{r_k\}_{k=n+1}^{\infty} such that |r_k| < 1. However, this characterization provides little information on the order of these realizations. It is not clear how to choose these Schur parameters so that the resulting v(z) function gives a realization of order m.

4.2.2 Orthogonal Polynomial Descriptions

For any polynomial of degree k, we define [60]

\[ P_k^*(z) = z^k P_k(z^{-1}). \] \hspace{1cm} (4.15)

The sequence of Schur parameters determines two sets of polynomials \(\Psi_i(z), i = 0, 1, \cdots, n\) and \(\Phi_i(z), i = 0, 1, \cdots, n\) through the following recursive formulas:

\[ \Phi_i(z) = z\Phi_{i-1}(z) - r_i\Phi_{i-1}^*(z), \quad \Phi_0(z) = 1 \] \hspace{1cm} (4.16)

\[ \Psi_i(z) = z\Psi_{i-1}(z) + r_i\Psi_{i-1}^*(z), \quad \Psi_0(z) = 1 \] \hspace{1cm} (4.17)

The polynomials \(\Phi_i(z)\) and \(\Psi_i(z)\) are called orthogonal polynomials on the unit circle of the first and second kind, respectively.

\(\Phi_i(z)\) and \(\Psi_i(z)\) are related by
\[ \Phi_i(z)\Psi_i^*(z) + \Phi_i^*(z)\Psi_i(z) = 2\sigma_i^2z^i \]  (4.18)

or

\[ \Phi_i(z)\Psi_i(z^{-1}) + \Phi_i(z^{-1})\Psi_i = 2\sigma_i^2. \]  (4.19)

It is obvious that \( \Phi_n(z) \) is stable if and only if \( \Psi_n(z) \) is stable.

These orthogonal polynomials have been used to obtain several important characterizations of finite order stochastic realizations [27,29,30,33]. Some of the more pertinent results can be summarized by the following theorems which describe the set of all functions in CEP for a given sequence \( C(n) \).

**Theorem 4.2:** Under the condition \( T_n > 0 \), a function \( u(z) \) is a solution to CEP for \( C(n) \), if and only if, it can be represented as [33]

\[
u(z) = \frac{c_0 z\Psi_n(z) - u(z)\Psi_n^*(z)}{2 z\Phi_n(z) + u(z)\Phi_n^*(z)},
\]  (4.20)

where \( u(z) \) is an arbitrary function analytic on \(|z| > 1\) satisfying \(|u(e^{i\omega})| < 1\) for any \( \omega \).

This theorem gives a complete parameterization of all stochastic realizations which are consistent with \( C(n) \). If \( u(z) \) is a rational function, the resulting realization is rational. For the case \( u(z) = 0 \), equation (4.20) becomes

\[
u(z) = \frac{c_0 \Psi_n(z)}{2 \Phi_n(z)},
\]  (4.21)

which is the Maximum Entropy (ME) realization of \( C(n) \).

**Example 4.3:**

Given \( c_0 = 5, c_1 = 3, c_2 = 2 \), we have \( r_1 = \frac{3}{5}, r_2 = \frac{1}{16} \) and

\[ \Phi_2(z) = z^2 - \frac{9}{16}z - \frac{1}{16} \]
\[ \Psi_2(z) = z^2 + \frac{51}{80}z + \frac{1}{16} \]

For the maximum entropy solution, we have

\[
v(z) = \frac{c_0 \Psi_2(z)}{2 \Phi_2(z)} = 2.5 \frac{z^2 + \frac{51}{80}z + \frac{1}{16}}{z^2 - \frac{9}{16}z - \frac{1}{16}} = 2.5 + 3z^{-1} + 2z^{-2} + \cdots.
\]

The ME realization is equivalent to the covariance extension obtained by setting the Schur parameters \( r_k = 0 \) for \( k > n \). It is not clear, however, how to choose other \( u(z) \) functions in order to obtain realizations of a particular order \( m \), especially for \( m < n \).

Another result which parameterizes realizations of order \( n \) (based on equation (4.20)) is given in the following theorem (obtained independently by Georgiou [29] and Kimura [28]).

**Theorem 4.3:** A rational function \( v(z) \) is a solution to the CEP for \( C^{(n)} \) of order \( n \), if and only if \( v(z) \) is stable, nonnegative, and can be represented as

\[
v(z) = \frac{c_0 \Psi_n(z) + \alpha_1 \Psi_{n-1}(z) + \cdots + \alpha_n \Psi_0(z)}{2 \Phi_n(z) + \alpha_1 \Phi_{n-1}(z) + \cdots + \alpha_n \Phi_0(z)},
\]

where \( \alpha_i, i = 1, \cdots, n, \) are some real numbers.

This theorem parameterizes all realizations of order \( n \) in terms of \( n \) \( \alpha_i \) coefficients. These \( \alpha_i \) coefficients must be chosen so that \( v(z) \) is both stable and nonnegative; however, general conditions on the coefficients \( \alpha_i \) to meet these requirements are not known.

**4.2.3 Remarks**

The above theory provides a means of constructing general ARMA models with order at least as large as the number of given covariance samples. In addition,
this theory provides a complete description of all admissible covariance extensions. However, these descriptions give little information on how to find stable, nonnegative definite descriptions. Moreover, neither description provides a direct means of finding or characterizing realizations of order less than \( n \). From an engineering standpoint, it is natural and often necessary to consider realizations of low order [31,32]. For the partial covariance sequence \( C^{(n)} \), it seems natural to use only realizations of order up to \( n \) (In fact, depending on \( C^{(n)} \) there might not exist any realization of lower order [40], whereas, there exists always one of order equal to \( n \)-the \( AR(n) \) model that corresponds to the ME extension). So the question has been raised regarding realizations of minimal order [32,34]. However, the existence of extensions to \( C^{(n)} \) of minimal order and a characterization of all such extensions, if they exist, remain unsolved problems.

### 4.3 Minimal Order Partial Realizations

Suppose we are given a covariance sequence \( C^{(n)} \). The question is how to find all realizations of minimum order.

Assume the given autocovariances \( C^{(n)} \) correspond to an \( m^{th} \) order ARMA model. Then \( v(z) \) is of the form

\[
v(z) = \frac{D_m(z)}{A_m(z)} = \frac{d_0 z^m + d_1 z^{m-1} + \cdots + d_m}{z^m + a_1 z^{m-1} + \cdots + a_m}
\]

\[
= \frac{1}{2} c_0 + c_1 z^{-1} + \cdots + c_n z^{-n} + \sum_{j=n+1}^{\infty} c_j z^{-j},
\]

where \( c_j \) are the extensions of \( C^{(n)} \) and where (c.f. equation (4.6)).

\[
B_m(z)B_m(z^{-1}) = A_m(z)D_m(z^{-1}) + A_m(z^{-1})D_m(z).
\]

From equation (4.24), we have the following equations:
\begin{align}
\begin{bmatrix}
c_m & c_{m-1} & \cdots & c_1 \\
c_{m+1} & c_m & \cdots & c_2 \\
\vdots & \vdots & \ddots & \vdots \\
c_{n-1} & c_{n-2} & \cdots & c_{n-m}
\end{bmatrix}
\begin{bmatrix}
a_1 \\
a_2 \\
\vdots \\
a_m
\end{bmatrix}
= -
\begin{bmatrix}
c_{m+1} \\
c_{m+2} \\
\vdots \\
c_n
\end{bmatrix},
\end{align}
(4.26)

\begin{align}
\begin{bmatrix}
d_0 \\
d_1 \\
\vdots \\
d_m
\end{bmatrix}
= 
\begin{bmatrix}
\frac{1}{2}c_0 & 0 & \cdots & 0 \\
c_1 & \frac{1}{2}c_0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
c_m & c_{m-1} & \cdots & \frac{1}{2}c_0
\end{bmatrix}
\begin{bmatrix}
a_1 \\
a_2 \\
\vdots \\
a_m
\end{bmatrix},
\end{align}
(4.27)

\begin{align}
c_k = -c_{k-1}a_1 - \cdots - c_{k-m}a_m, \quad k > n.
\end{align}
(4.28)

It is obvious that \( c_j \) for \( j > n \) are just the linear combinations of \( c_1, \ldots, c_n \) through equation (4.28). \( d_0, d_1, \ldots, d_m \) can be determined as far as \( a_1, \ldots, a_m \) are determined.

Denote by \( a \) and \( d \) the coefficient vectors in equations (4.26) and (4.27). For a given order \( m \leq \left\lfloor \frac{n}{2} \right\rfloor \), one can compute the unique vectors \( a \) and \( d \) by solving equations (4.26) and (4.27). Note that equation (4.26) may not have a solution, in which case no realization of that order exists. If equation (4.26) does have a solution, then two additional constraints must be satisfied in order for \( a \) to correspond to an \( m \)th order realization:

1. \( a \in S_a \), where:

\begin{align}
S_a = \{ a : A_m(z) = 0 \Rightarrow |z| < 1 \}
\end{align}
(4.29)

2. \( a \in N_a \), where:

\begin{align}
N_a = \{ a : A_m(z)D_m(z^{-1}) + A_m(z^{-1})D_m(z) \geq 0, \ |z| = 1 \}
\end{align}
(4.30)
It is clear from the above discussion that given $C^{(n)}$, one can recursively test for realizations of order 1, 2, \cdots, \lfloor \frac{m}{2} \rfloor$. For each order $m \leq \lfloor \frac{m}{2} \rfloor$, one solves for $a$ and $d$ using equations (4.26) and (4.27) (if a solution exists), and tests the above two conditions. If the conditions are all satisfied, the unique realization of order $m$ has been found.

For orders greater than $\lfloor \frac{m}{2} \rfloor$, equations (4.26) and (4.27) still must be satisfied. However, the covariances $c_{n+1}, \cdots, c_{2m}$ in these equations are unknown. The remainder of this chapter considers the questions of whether realizations exist for these orders, and if so, how they may be characterized.

If $n = 2m$, then (4.26) is a set of $m$ equations, and can be used to solve for $a_1, \cdots, a_m$ if the Hankel matrix

$$H_m = \begin{bmatrix}
c_1 & c_2 & \cdots & c_m \\
c_2 & c_3 & \cdots & c_{m+1} \\
\vdots & \vdots & \cdots & \vdots \\
c_m & c_{m+1} & \cdots & c_{2m-1}
\end{bmatrix}$$

(4.31)

is nonsingular. Note that $H_m$ is just a column permutation of the matrix in (4.26).

Thus, from [31] we can have the following theorem

**Theorem 4.4:** If $H_m$ is nonsingular and $v(z)$ obtained from equations (4.26) and (4.27) is positive real, then $v(z)$ is a minimal order realization of $C^{(2m)}$ and is unique.

Here we use an example to make the procedure clear.

**Example 4.4:**

Given $c_0 = 5, c_1 = 2, c_2 = 3, c_3 = 1, c_4 = 1$, and $m = 2$, equations (4.26) and (4.27) in this case become
Finally

\[ v(z) = \frac{17.5z^2 + 11.5z + 14}{7z^2 - z - 2}. \]

It is easy to show \( v(z) \) is positive real.

If \( C(n) \) is given with \( m < n < 2m \), we need \( 2m - n \) more equations in (4.26) to determine \( a_1, \ldots, a_m \). At this moment no theory is available as how to extend \( C(n) \) to \( C(2m) \) so that the corresponding \( v(z) \) is PR.

4.4 Minimal Order Realization for Single Unknown Case

In this section we consider the following problem:

Given \( C(2m-1) \), determine whether or not realizations of order \( m \) exist, and if so, find the set of all stable \( m^{th} \) order realizations and find the set of all NND \( m^{th} \) order realizations. The intersection of these two sets gives all stable and NND \( m^{th} \) order realizations. Such realizations, if they exist, have the minimal order \( m \).
This problem has a single unknown coefficient $c_{2m}$; thus, this problem is equivalent to finding the set of all $c_{2m}$ coefficients such that $C^{(2m)}$ corresponds to a stable stochastic realization of order $m$, and to finding the set of all $c_{2m}$ coefficients such that $C^{(2m)}$ corresponds to a NND stochastic realization of order $m$. Throughout this discussion, we assume that $C^{(2m-1)}$ does not admit a realization of order less that $m$. The intersection of these sets gives all stable and NND realizations with the minimal order $m$.

4.4.1 Stability Set

Solving equation (4.26) in terms of $c_{2m}$, we have

$$a_i = \alpha_i + \beta_i c_{2m}, \quad i = 1, 2, \ldots, m,$$

where

$$
\begin{bmatrix}
\alpha_1 \\
\alpha_2 \\
\vdots \\
\alpha_m
\end{bmatrix} =
- \begin{bmatrix}
v_{11} & v_{12} & \cdots & v_{1(m-1)} \\
v_{21} & v_{22} & \cdots & v_{2(m-1)} \\
\vdots & \vdots & \ddots & \vdots \\
v_{m1} & v_{m2} & \cdots & v_{m(m-1)}
\end{bmatrix}
\begin{bmatrix}
c_{m+1} \\
c_{m+2} \\
\vdots \\
c_{2m-1}
\end{bmatrix},
$$

(4.33)

$$
\begin{bmatrix}
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_m
\end{bmatrix} =
- \begin{bmatrix}
v_{1m} \\
v_{2m} \\
\vdots \\
v_{mm}
\end{bmatrix},
$$

(4.34)

and $v_{ij}$ is $ij$-th element of matrix $V$ given by
\[
V = \begin{bmatrix}
  c_m & c_{m-1} & \cdots & c_1 \\
  c_{m+1} & c_m & \cdots & c_2 \\
  \vdots & \vdots & \ddots & \vdots \\
  c_{2m-1} & c_{2m-2} & \cdots & c_m
\end{bmatrix}^{-1}
\]

(4.35)

Working on the closed compact subset \( S_a \) directly turns out to be very difficult, so we will transform the polynomial coefficients in discrete-time case to the continuous-time case by the bilinear transformation

\[ z = \frac{1 + s}{1 - s}. \]  

(4.36)

This gives

\[ A_m(s) = (1 - s)^m A_m\left(\frac{1 + s}{1 - s}\right) = P_m(s) + c_{2m} Q_m(s) \]  

(4.37)

with

\[ P_m(s) = p_0 s^m + p_1 s^{m-1} + \cdots + p_m, \]  

(4.38)

\[ Q_m(s) = q_0 s^m + q_1 s^{m-1} + \cdots + q_m, \]  

(4.39)

\[ p_k = \sum_{i=0}^{m} \alpha_i b_{mk}^i, \]  

(4.40)

\[ q_k = \sum_{i=0}^{m} \beta_i b_{mk}^i, \]  

(4.41)

where \( b_{mk}^i = \sum_{j=0}^{k} (-1)^j C_j^i C_{m-i}^{k-j}, \alpha_0 = 1, \beta_0 = 0, \) and \( k = 0, 1, \ldots, m. \)

In general, \( P_m(s) \) and \( Q_m(s) \) are polynomials of order \( m. \)

Since the bilinear transformation is bijective, the stability intervals for \( c_{2m} \) in \( A_m(z) \) are the same as the stability intervals for \( c_{2m} \) in \( A_m(s) \). Also since the
stable polynomial coefficient sets in the discrete-time and continuous-time are not convex in general, there may be more than one stability interval for $c_{2m}$.

To find the stability intervals for $c_{2m}$ in $A_m(s)$, the following generalized version of Theorem 3.2 in [94] is used.

The Hurwitz matrix of an $m$-th order polynomial

$$A(s) = a_0 s^m + a_1 s^{m-1} + \cdots + a_m$$

has dimension $m \times m$ and is denoted by

$$H(A) \triangleq \begin{bmatrix}
    a_1 & a_3 & a_5 & \cdots & 0 \\
    a_0 & a_2 & a_4 & \cdots & 0 \\
    0 & a_1 & a_3 & \cdots & 0 \\
    0 & a_0 & a_2 & \cdots & 0 \\
    \vdots & \vdots & \vdots & \cdots & \vdots \\
    \vdots & \vdots & \vdots & \cdots & a_m
\end{bmatrix}.$$  \hfill (4.43)

We have the following theorem:

**Theorem 4.5:** Given an $m_1$-th order polynomial

$$P(s) = p_0 s^{m_1} + p_1 s^{m_1-1} + \cdots + p_{m_1}$$

and an $m_2$-th order polynomial

$$Q(s) = q_0 s^{m_2} + q_1 s^{m_2-1} + \cdots + q_{m_2}.$$ \hfill (4.45)

Then, the $n$-th ($n = \max\{m_1, m_2\}$) order polynomial

$$A_r(s) = P(s) + rQ(s)$$ \hfill (4.46)
has negative real part roots if and only if the set \( S_r = \bigcup_{i=1}^{k} (r_{\min}^i, r_{\max}^i) \) is not empty and \( r \in S_r \), where \( k \) is the number of stability intervals, each \( r_{\min}^i \) and \( r_{\max}^i \) is in general from \(-\lambda, the real eigenvalues of the equation\)

\[
H(P)x = \lambda H(Q)x
\]

or a solution of \( p_0 + rq_0 = 0 \).

Here \( H(P) \) and \( H(Q) \) are computed by treating \( P(s) \) and \( Q(s) \) as \( n \)-th order polynomials. It is possible that \( r_{\min}^1 = -\infty \) or \( r_{\max}^k = \infty \).

**Proof:** Since the coefficients of \( A_r(s) \) are linear functions of \( r \), thus when \( r \) varies on \([-\infty, \infty]\), the coefficients vary along a hyperline in \( \mathbb{R}^n \). The stability intervals can be found as intersections of the hyperline and the stability region of an \( n \)-th order polynomial in its coefficient space. The problem here is how to find two end values for each interval and the number of intervals \( k \). We have two cases here. Assume equation (4.47) has \( L \) real eigenvalues \( \lambda_i, i = 1, \ldots, L \), and set \( r_i = -\lambda_i, i = 1, \ldots, L \).

**Case 1:** \( n = m_1 \geq m_2 \) and \( P(s) \) is a strictly stable polynomial.

Let us define

\[
\begin{align*}
r_{\min}^* &= \begin{cases} 
-p_0/q_0 & \text{if } m_1 = m_2 \text{ and } p_0q_0 > 0, \\
-\infty & \text{otherwise,}
\end{cases} \\
r_{\max}^* &= \begin{cases} 
-p_0/q_0 & \text{if } m_1 = m_2 \text{ and } p_0q_0 < 0, \\
+\infty & \text{otherwise,}
\end{cases} \\
r_{\min}^- &= \max\{r_i, \text{ for } r_i < 0\}, \\
r_{\max}^+ &= \min\{r_i, \text{ for } r_i > 0\},
\end{align*}
\]

(4.48) (4.49) (4.50) (4.51)
\[ r_{\text{min}} = \max\{r_{\text{min}}^*, r_{\text{min}}^-, r_{\text{min}}\}, \quad (4.52) \]
\[ r_{\text{max}} = \min\{r_{\text{max}}^*, r_{\text{max}}^+, r_{\text{max}}\}. \quad (4.53) \]

Then from \([94]\), \(A_r(s)\) is a strictly stable polynomial if \(r \in (r_{\text{min}}, r_{\text{max}})\). Here \(r_{\text{min}} \geq r_{\text{min}}^*\) and \(r_{\text{max}} \leq r_{\text{max}}^*\) in order to guarantee that every \(A_r(s)\) for \(r \in (r_{\text{min}}, r_{\text{max}})\) is an \(n\)-th order polynomial.

Case 2: \(n = m_1 \geq m_2\) and \(P(s)\) is not a strictly stable polynomial, or \(n = m_2 > m_1\).

In this case we assume that \(A'(s)\) is strictly stable for some \(r_0\). We can define
\[ A'(s) = P(s) + r_0 Q(s), \quad (4.54) \]
and study \(A_r(s) = A'(s) + r^l Q(s)\). The method in Case 1 can be applied here now. \(A_r(s)\) is a strictly stable polynomial if \(r \in (r_0 + r_{\text{min}}^*, r_0 + r_{\text{max}}^*)\). However, it is easy to show that each \(-(r_0 + r')\) is also an eigenvalue of equation (4.47) (since we have \(H(P)x = -(r_0 + r')H(Q)x\) or a solution of \(p_0 + rq_0 = 0\). Thus we have proved the theorem.

In this generalized case, the polynomial \(P(s)\) is not required to be a strictly stable one and has an order which may be less than the order of \(Q(s)\).

Note that this generalized theorem only gives the source of the end points. We need the following procedure to find \(S_r\) from this source.

1) Form matrices \(H(P)\) and \(H(Q)\), and solve equation (4.47) to get real eigenvalues (in decreasing order) \(\lambda_1, \ldots, \lambda_L\).

2) For the case \(m_1 = m_2\), obtain \(r^*\) as \(r^* = -p_0/q_0\).
3) Let \( r_i, i = 1, \ldots, L + 1 \) be \( r^* \) and \( \{-\lambda_j\}_1^L \) in increasing order. Note if \( m1 \neq m2 \), then \( i = 1, \ldots, L \).

4) Set \( r_0 = -\infty \) and \( r_{L+2} = \infty \).

5) Form intervals \( I_i = (r_{i-1}, r_i), i = 1, \ldots, L + 2 \).

6) Decide the stable intervals. For \( i = 1, \ldots, L + 2 \), if \( A_{0.5}(r_{i-1}+r_i)(s) \) is stable, then \( (r_{i-1}, r_i) \) is a stable interval, otherwise it is not.

For our particular case, we have:

**Theorem 4.6:** Given an \( m \)-th order polynomial \( P_m(s) \) and an \( m \)-th order polynomial \( Q_m(s) \). Then the \( m \)-th order polynomial

\[
A_m(s) = P_m(s) + c_{2m} Q_m(s)
\]

(4.55)

has only negative real part roots if and only if \( S_c = \bigcup_{i=1}^k (c_{\min i}, c_{\max i}) \) is not empty and \( c_{2m} \in S_c \).

The procedures to find \( S_c \) is summarized as following:

- Use equation (4.26) to calculate \( \alpha_i, \beta_i \), for \( i = 1, \ldots, m \). Set \( \alpha_0 = 1 \) and \( \beta_0 = 0 \).

- Use equations (4.40) and (4.41) to calculate \( p_i, q_i \), for \( i = 0, 1, \ldots, m \).

- Form the stability intervals for \( c_{2m} \) using the procedure above.

**Theorem 4.7:** Given \( C^{(2m-1)} \), we can realize it with a stable system of order \( m \) if and only if \( S_c \) is not empty.
4.4.2 Non-Negative Definite Set

The set of \( c_{2m} \) coefficients which yield nonnegative definite (NND) solutions must be found. Note that a necessary condition for a NND solution is that \( |T_{2m}| \geq 0 \); this determinant is quadratic in \( c_{2m} \), and an interval for \( c_{2m} \) which yield NND realizations can readily be computed. However, this is not a sufficient condition because \( c_{2m} \) uniquely specifies all remaining covariances; the extended sequence must then satisfy \( |T_n| > 0 \) for all \( n \geq 2m \), not just for \( n = 2m \).

Necessary and sufficient conditions for NND realizations can be found by considering

\[
f(\omega, c_{2m}) = A_m(e^{j\omega})D_m(e^{-j\omega}) + A_m(e^{-j\omega})D_m(e^{j\omega})
\]  

(4.56)

as a function of \( c_{2m} \). From equation (4.26), it can be seen that \( a \) vector is a linear function of \( c_{2m} \). Moreover, from equation (4.27), \( d \) is also a linear function of \( c_{2m} \). Thus, \( f(\omega, c_{2m}) \) is quadratic in \( c_{2m} \). This function has the form

\[
f(\omega, c_{2m}) = f_0(\omega) + f_1(\omega)c_{2m} + f_2(\omega)c_{2m}^2,
\]

(4.57)

where (cf. (4.26) and (4.27))

\[
f_0(\omega) = \alpha^T W \alpha,
\]

(4.58)

\[
f_1(\omega) = 2\alpha^T W \beta,
\]

(4.59)

\[
f_2(\omega) = \beta^T W \beta.
\]

(4.60)
with

\[ \alpha = [1, \alpha_1, \ldots, \alpha_m]^T, \]  
\[ (4.61) \]

\[ \beta = [0, \beta_1, \ldots, \beta_m]^T, \]  
\[ (4.62) \]

\[ Z = [e^{j\omega_1}, \ldots, e^{j\omega}, 1]^T, \]  
\[ (4.63) \]

\[ S = \begin{bmatrix} \frac{1}{2}c_0 & 0 & \cdots & 0 \\ c_1 & \frac{1}{2}c_0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ c_m & c_{m-1} & \cdots & \frac{1}{2}c_0 \end{bmatrix}, \]  
\[ (4.64) \]

\[ W = ZZ^*S + S^TZZ^*. \]  
\[ (4.65) \]

where "t" and "*" denote transpose and complex conjugate transpose, respectively.

We will find a set \( N_c \) such that \( f(\omega, c_{2m}) \geq 0 \) for \( \omega \in [0, \pi] \) and for any \( c_{2m} \in N_c \).

At this point, it requires only to find \( c_{2m} \) such that \( f(\omega, c_{2m}) \geq 0 \) for all \( \omega \in [0, \pi] \).

One method is to test equation (4.57) for various \( \omega \), specifically, given an \( \omega \in [0, \pi] \), from equation (4.57) we determine intervals for \( c_{2m} \) in which \( f(\omega, c_{2m}) \geq 0 \). Finally, the set \( N_c \) is the intersection of all intervals found above changing \( \omega \) from 0 to \( \pi \).

This method is conceptually easy, but it is hard to implement since the set \( N_c \) contains many intervals, for example, for the case \( m = 2 \), \( N_c \) contains as many as three intervals.
The purpose of the following discussions is to find frequencies from which we can obtain the boundary points of \( N_c \).

We can rewrite equation (4.57) as

\[
f(\omega, c_{2m}) = f_2(\omega)[c_{2m} + \frac{f_1(\omega)}{2f_2(\omega)}]^2 + T(\omega), \quad \text{for } f_2(\omega) \neq 0 \tag{4.66}
\]

with

\[
T(\omega) = f_0(\omega) - \frac{f_1^2(\omega)}{4f_2(\omega)}. \tag{4.67}
\]

Two conclusions follow from equation (4.66). First, if there is an interval \([\omega_1, \omega_2] \subset [0, \pi]\) such that for any \( \omega \in [\omega_1, \omega_2] \) \( T(\omega) < 0 \) and \( f_2(\omega) < 0 \), then no \( c_{2m} \) exists such that \( f(\omega, c_{2m}) \geq 0 \) for all \( \omega \in [0, \pi] \), and therefore, no stochastic realization of order \( m \) exists for \( C^{(2m-1)} \). Second, if \( f_2(\omega) > 0 \) and \( T(\omega) > 0 \) for any \( \omega \in [\omega_1, \omega_2] \subset [0, \pi] \), \( f(\omega, c_{2m}) > 0 \) for any \( c_{2m} \) on the interval \([\omega_1, \omega_2]\). This implies that the frequencies which give the boundary points of \( N_c \) do not belong to \([\omega_1, \omega_2]\). From the above discussions the following theorem is useful.

**Theorem 4.8:** The frequencies from which we can obtain the end points of \( N_c \) are from one of the following two cases:

1. an interval \([\omega_{+1}, \omega_{+2}] \subset [0, \pi] \), in which \( f_2(\omega) > 0 \) and \( T(\omega) < 0 \).
2. an interval \([\omega_{-1}, \omega_{-2}] \subset [0, \pi] \), in which \( f_2(\omega) < 0 \) and \( T(\omega) > 0 \).

Furthermore, the boundary points of \( N_c \) can be found from the following four kinds of points:

\[
c_{+1} = \min\left\{ -\frac{f_1(\omega) - \sqrt{f_1^2(\omega) - 4f_0(\omega)f_2(\omega)}}{2f_2(\omega)}, \quad \omega \in [\omega_{+1}, \omega_{+2}] \right\}, \tag{4.68}
\]
Proof: The first part is obvious from the properties of a quadratic function. For any \( \omega \in [\omega_+1, \omega_+2] \), \( f(\omega, c_{2m}) > 0 \) if \( c_{2m} \to \infty \) or \( c_{2m} \to -\infty \) since \( f_2(\omega) > 0 \), \( T(\omega) \) and \( f_1(\omega)/[2f_2(\omega)] \) are finite. Furthermore, for any fixed \( \omega \in [\omega_+1, \omega_+2] \), \( f(\omega, c_{2m}) \) is a quadratic function of \( c_{2m} \), we can find \( c_{2m1}(\omega) \) and \( c_{2m2}(\omega) \) such that \( f(\omega, c_{2m}) \) are nonnegative for \( c_{2m} \in (-\infty, c_{2m1}(\omega)] \) or \( c_{2m} \in [c_{2m2}(\omega), \infty) \) with

\[
c_{2m1}(\omega) = \frac{-f_1(\omega) - \sqrt{f_1^2(\omega) - 4f_0(\omega)f_2(\omega)}}{2f_2(\omega)},
\]

\[
c_{2m2}(\omega) = \frac{-f_1(\omega) + \sqrt{f_1^2(\omega) - 4f_0(\omega)f_2(\omega)}}{2f_2(\omega)}.
\]

Thus

\[
c_{+1} = \min\{c_{2m1}(\omega) \mid \omega \in [\omega_+1, \omega_+2]\}, \quad (4.74)
\]

\[
c_{+2} = \max\{c_{2m2}(\omega) \mid \omega \in [\omega_+1, \omega_+2]\}. \quad (4.75)
\]

In the same way, we can prove \( c_{-1} \) and \( c_{-2} \) as the end points from the case \( f_2(\omega) < 0 \) and \( T(\omega) > 0 \) for \( \omega \in [\omega_-1, \omega_-2] \).

The procedure to find \( N_c \) is summarized below:
- Divide \([0, \pi]\) into four kinds of subintervals according to \(f_2(\omega) > 0\) and \(T(\omega) > 0\), \(f_2(\omega) < 0\) and \(T(\omega) < 0\), \(f_2(\omega) < 0\) and \(T(\omega) > 0\), and \(f_2(\omega) > 0\) and \(T(\omega) < 0\).

- If there is an interval \([\omega_1, \omega_2] \subseteq [0, \pi]\) in which \(f_2(\omega) < 0\) and \(T(\omega) < 0\), \(N_c\) is empty. NND realization does not exist for the order \(m\).

- Calculate \(c_{-1}\) and \(c_{-2}\) for each interval \([\omega_{-1}, \omega_{-2}]\) in which \(f_2(\omega) < 0\) and \(T(\omega) > 0\). \(f(\omega, c_{2m}) \geq 0\) if \(\omega \in [\omega_{-1}, \omega_{-2}]\) and \(c_{2m} \in [c_{-1}, c_{-2}]\). \(N_c\) is empty if \(c_{-1} > c_{-2}\).

- Calculate \(c_{+1}\) and \(c_{+2}\) for \([\omega_{+1}, \omega_{+2}]\) in which \(f_2(\omega) > 0\) and \(T(\omega) < 0\). \(f(\omega, c_{2m}) \geq 0\) if \(\omega \in [\omega_{+1}, \omega_{+2}]\) and \(c_{2m} \in (-\infty, c_{+1}] \cup [c_{+2}, \infty)\).

- \(N_c\) is the intersections of the intervals found above.

Finally, we have

**Theorem 4.9:** Given \(C^{(2m-1)}\), we can realize it with a NND system of order \(m\) if and only if \(N_c\) is nonempty.

We summarize the results of this section in the following theorem:

**Theorem 4.10:** Given \(C^{(2m-1)}\), the set of all \(c_{2m}\) such that \(C^{(2m)}\) yields an \(m\)th order realization is given by \(c_{2m} \in N_c \cap S_c\).

### 4.5 Some Further Results

#### 4.5.1 General Case of Stability Set

The method in previous section can be readily extended to stability regions for realizations of higher order. The resulting algorithm provides a simple
means for determining whether or not a stable solution of order \( m = \lceil \frac{n}{2} \rceil + k \) exists for any \( k \in \{1, \ldots, \lfloor \frac{n+1}{2} \rfloor \} \).

The algorithm proceeds as follows. Assume we are given \( C^{(n)} \), and wish to find a realization of order \( m \). The first \( n - m \) rows of equation (4.26) are constraint equations. Each can be thought of as a hyperplane in the space \( R^m \) of \( a \) vectors. In order for a stable realization to exist, each hyperplane must have nonempty intersection in the compact set \( S_a \) of stable solutions. Consider, for example, the hyperplane defined by the first row of equation (4.26), with \( c_{m+1} \) as a variable. As \( c_{m+1} \) varies, the hyperplane sweeps through the stability set \( S_a \). In fact, there is a compact subset \( S_{m+1} \) of \( c_{m+1} \) for which a stable solution is possible. It is straightforward to find the set \( S_{m+1} \). If the given \( c_{m+1} \) coefficient does not lie in this set, no stable realization for this order is possible. This argument can be repeated for \( c_{m+2}, \ldots, c_n \). If all of these autocovariances fall within their corresponding stability interval, then a stable realization of order \( m \) can be found.

The above procedure can be continued for \( c_{n+1}, \ldots, c_{2m} \) to determine necessary and sufficient conditions on these coefficients for stable realizations. The set of all realizations of order \( m \) is given by the set of all coefficients \( c_{n+1}, \ldots, c_{2m} \) which satisfy their corresponding stability interval constraint. However, the stability set constraint on, for example, \( c_{n+2} \) is a function of \( c_{n+1} \) (c.f. equation (4.26)), which itself must lie in a stability set. Thus, one obtains a family of stability sets \( c_{n+2} \) parameterized on \( c_{n+1} \). This recursive dependence quickly becomes complicated. Thus, the above algorithm does not yet appear to provide a "clean" characterization of all solutions of a
given order. On the other hand, the present method is useful for constructing solutions of minimal order.

These stability results can be summarized in the following theorem.

**Theorem 4.11:** Given $C^{(n)}$, a stochastic realization of order $m$ which satisfies the stability constraint can be found if and only if:

$$c_i \in S_i, \quad i = m + 1, m + 2, \ldots, n$$

where each set $S_i$ is determined from $c_1, \ldots, c_{i-1}$ as discussed above. Moreover, any stable realization of order $m$ satisfies:

$$c_i \in S_i, \quad i = n + 1, n + 2, \ldots, 2m$$

where each set $S_i$ is determined from $c_1, \ldots, c_{i-1}$.

The single unknown case is just a special case of this result with $n = 2m - 1$. Note that although it is straightforward to find the set $S_{n+1}$, it is very tedious to do so. The following method can be used to find a set $\hat{S}_{n+1} = (s_{\min}, s_{\max})$ such that if $c_{n+1} \notin \hat{S}_{n+1}$, then there is no stable realization existing for $C^n$.

To find $\hat{S}_{n+1}$, the following theorem [95] can be used.

**Theorem 4.12:** The convex hull of $S_a$ is a polyhedron whose vertices correspond to all polynomials $A_m(z)$ with zeros in the set $\{-1, 1\}$.

This theorem states that we get $m + 1$ vectors $a$ representing $m + 1$ linearly independent points in $R^m$. These $m + 1$ points define uniquely a geometric $m$-simplex with the $m + 1$ points as vertices. Each point is the coefficients of the polynomial.
\[ A_m(z) = (z + 1)^i(z - 1)^{m-i}, \quad i = 0, 1, \ldots, m. \] (4.76)

The convex hull has \( m + 1 \) faces, each face is defined by

\[ x_{i1}a_1 + x_{i2}a_2 + \cdots + x_{im}a_m = 1, \] (4.77)

where \( i \) denotes the \( i \)th face. To find coefficients \( \{x_{i1}, x_{i2}, \ldots, x_{im}\} \) we only need to force equation (4.77) to pass through \( m \) vertices out of the \( m + 1 \) vertices defined in Theorem 4.12.

The convex hull can be expressed as

\[ H = \{a : x_{i1}a_1 + \cdots + x_{im}a_m \leq 1, i = 1, \ldots, m + 1\}. \] (4.78)

Solving equation (4.26) in terms of \( a_{n-m+1}, \ldots, a_m \), we have

\[ [a_1, \ldots, a_{n-m}]^T = e + F[a_{n-m+1}, \ldots, a_m], \] (4.79)

where

\[
\begin{bmatrix}
    c_m & \cdots & c_{2m-n-1} \\
    \vdots & \ddots & \vdots \\
    c_{n-1} & \cdots & c_m
\end{bmatrix}
{-1}
\begin{bmatrix}
    c_{m+1} \\
    \vdots \\
    c_n
\end{bmatrix},
\]

(4.80)

\[
F = -
\begin{bmatrix}
    c_m & \cdots & c_{2m-n-1} \\
    \vdots & \ddots & \vdots \\
    c_{n-1} & \cdots & c_m
\end{bmatrix}
{-1}
\begin{bmatrix}
    c_{2m-n-2} & \cdots & c_1 \\
    \vdots & \cdots & \vdots \\
    c_{m+1} & \cdots & c_{n-m}
\end{bmatrix}, \] (4.81)

\( c_{n+1} \) can be expressed as

\[ c_{n+1} = -[c_n a_1 + \cdots + c_{n-m+1} a_m] = \gamma + [a_{n-m+1}, \ldots, a_m] \beta, \] (4.82)
where

\[ \gamma = [c_n, \ldots, c_{m+1}]e, \] (4.83)

\[ \beta' = [c_n, \ldots, c_{m+1}]F + [c_{m+2}, \ldots, c_{n-m+1}]. \] (4.84)

Each hyperplane defined by equation (4.77) can now be modified as

\[ x_i(n-m+1)a_{n-m+1} + \cdots + x_imam = s_i, \] (4.85)

where

\[ s_i = 1 - [x_{i1}, \ldots, x_{i(n-m)}] [e + F[a_1, \ldots, a_{n-m}]^T]. \] (4.86)

Finally, \( s_{\min} \) and \( s_{\max} \) can be found as

\[ s_{\min} = \gamma + \min\{[a_{n-m+1}, \ldots, a_m] \beta\} \] (4.87)

subject to

\[ x_i(n-m+1)a_{n-m+1} + \cdots + x_imam \leq s_i, \quad i = 1, \ldots, m+1, \] (4.88)

and

\[ s_{\max} = \gamma + \max\{[a_{n-m+1}, \ldots, a_m] \beta\} \] (4.89)

subject to (4.88). \( s_{\min} \) and \( s_{\max} \) can be obtained using standard linear programming.

If \( s_{\min} \) or \( s_{\max} \) does not exists, then there is no stable realization of order \( m \).

Example 4.5:

Let \( c_1 = 5, c_2 = 4, c_3 = 3, c_4 = 2, c_5 = -1. \) For \( m = 3 \), the convex hull of the stability region is defined by
\[ -a_1 - a_2 - a_3 \leq 1 \]
\[ -\frac{1}{3}a_1 + \frac{1}{3}a_2 + a_3 \leq 1 \]
\[ a_1 - a_2 + a_3 \leq 1 \]
\[ \frac{1}{3}a_1 + \frac{1}{3}a_2 - a_3 \leq 1 \]

From equation (4.26), we have \( a_1 = -10 + a_3, \ a_2 = 7 - 2a_3, \) and
\[ c_6 = -(c_5a_1 + c_4a_2 + c_3a_3) = -24 + 2a_3. \]

Inserting expressions for \( a_1 \) and \( a_2 \) into (4.90) yields
\[ 3 \leq 1, \ \frac{17}{3} \leq 1, \ -17 + 6a_3 \leq 1, \ -1 - \frac{4}{3}a_3 \leq 1. \]

It is obvious that all points \([a_1, a_2, a_3]\) are in the outside of convex hull \( H \). Thus, there does not exist any \( c_6 \) to have a stable realization.

From the above discussion we obtain the following necessary condition for the existence of a stochastic realization of order \( m \).

**Theorem 4.13:** Given \( C^{(n)} \), for a stable realization of order \( m \) to exist it is necessary that \( \hat{S}_{n+1} \) is not empty.

### 4.5.2 General Case of Non-Negative Definite Set

The procedure in the single unknown case can be generalized to higher order cases. In the general case, the function \( f(\omega) \) depends on more than one autocovariance, and the dependence is no longer quadratic. On the other hand, \( f(\omega) \) is a bilinear function of \( a \). Thus, the problem of finding the NND region from \( f(\omega) \) is conceptually similar to the single unknown case.

Using equations (4.26) and (4.27), we have
\[ f(\omega) = A(e^{-j\omega})D(e^{j\omega}) + D(e^{-j\omega})A(e^{j\omega}) = Z^*EZ, \quad (4.91) \]

where \( Z = [e^{jm\omega}, \ldots, e^{j\omega}, 1]^T \) and

\[
E = \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} \begin{bmatrix} 1 \\ a_1 & \cdots & a_m \end{bmatrix}S + S \begin{bmatrix} 1 \\ a_1 & \cdots & a_m \end{bmatrix}. \quad (4.92)
\]

\( f(\omega) \) can also be expressed as

\[
f(\omega) = \rho_0 + 2\rho_1 \cos \omega + \cdots + 2\rho_m \cos m\omega, \quad (4.93)
\]

where \( \rho_j = \sum_{i=0}^{m-j} e_{i(i+j)} \) for \( j = 0, 1, \ldots, m \), and \( e_{ij} \) is the \( ij \)-th element of \( E \). \( a_1, \ldots, a_{n-m} \) can also be expressed in terms of \( a_{n-m+1}, \ldots, a_m \) (see (4.79)).

It can be shown that \( \rho_0 \) is always positive. For \( \rho_0, \ldots, \rho_m \) to be a covariance sequence of a MA model, it is necessary that \( |\rho_i| \leq \rho_0 \) for \( i = 1, \ldots, m \). Thus if there is a \( \rho_k \) such that \( |\rho_k| > \rho_0 \) for all values of \( a_{n-m+1}, \ldots, a_m \), there is no NND realization of order \( m \) existing.

For convenience, let us define

\[
\eta_i = \frac{2\rho_i}{\rho_0}, \quad i = 1, \ldots, m. \quad (4.94)
\]

Then

\[
f(\omega) = \rho_0 \hat{f}(\omega), \quad (4.95)
\]

where
\[ \hat{f}(\omega) = 1 + \eta_1 \cos \omega + \cdots + \eta_m \cos m\omega. \] (4.96)

Note that \( f(\omega) \) is NND if and only if \( \hat{f}(\omega) \) is NND. By introducing \( \hat{f}(\omega) \), we can find a tight convex polyhedron \( H_\eta \) of the NND set \( N_\eta \) of \( \hat{f}(\omega) \) in terms of \( \eta = [\eta_1, \ldots, \eta_m]^T \) defined by

\[ N_\eta = \{ \eta : \hat{f}(\omega) \geq 0 \text{ for } \omega \in [0, \pi] \}. \] (4.97)

As explained in Chapter III, \( N_\eta \) is a compact convex set.

If we rewrite \( \hat{f}(\omega) \) as

\[ \hat{f}(\omega) = \text{Re}\{ \sum_{i=0}^{m} \eta_i z^{-i} \} = \text{Re}\{ z^{-m} \sum_{i=0}^{m} \eta_i z^{m-i} \}, \] (4.98)

where \( z = e^{j\omega} \) and \( \eta_0 = 1 \), then we have the following \( m \)-th order polynomial

\[ A_m(z) = z^m + \eta_1 z^{m-1} + \cdots + \eta_m. \] (4.99)

This polynomial is stable if \( \eta \in S_a \). For any \( \eta \in S_a \), \( A_m(z) \neq 0 \) for any \( z = e^{j\omega} \). This implies that \( |z^{-m} \sum_{i=0}^{m} \eta_i z^{m-i}| \neq 0 \) for any \( z = e^{j\omega} \). Thus for any \( \eta \) in the interior of \( N_\eta \), it also belongs to \( S_a \). This implies \( N_\eta \in S_a \). On the other hand, \( S_a \subset H \), therefore \( N_\eta \subset H \).

It is well-known that \( |\eta_m| \leq 1 \) if \( \hat{f}(\omega) \geq 0 \). This can be seen here since \( |\eta_m| \leq 1 \) if \( A_m(z) \) is stable.

Note that the convex hull \( H \) defined previously in (4.78) is much tighter than the one defined by \( |\rho_i/\rho_0| \leq 1 \) for the \( m = 2 \) case.

We now have the convex polyhedron \( H_\eta \) as

\[ H_\eta = H \cap \{|\eta_i| \leq 2, i = 1, \ldots, m\}. \] (4.100)
From the above discussion we obtain the following necessary condition for the existence of a stochastic realization of order $m$

**Theorem 4.14:** If $\eta$ obtained from $a_{n-m+1}, \ldots, a_m$ does not belong to $H_{\eta}$, then NND realizations of order $m$ do not exist.

**Example 4.6:**

Given $\eta_1 = 1.8, \eta_2 = 1.8, \eta_3 = 0.1$. Here each $|\eta_i| < 2$. Substitute $\eta_i$ into the convex hull $H$ in (4.90) we have

$$-3.7 \leq 1, \quad 0.1 \leq 1, \quad 0.1 \leq 1, \quad 1.1 \leq 1.$$  

Because of the fourth inequality, these $\eta_i$'s are not in $H_{\eta}$. Therefore, the $\eta_i$'s are not a NND sequence.

We can also draw some sufficient conditions from $\hat{f}(\omega)$. Since $\hat{f}(\omega)$ is NND for $\eta \in \Theta$, and the set $N_{\eta}$ is a compact convex set, the set

$$H_N = \{ \eta : \alpha_1 \eta_1 + \cdots + \alpha_m \eta_m \leq 1, \quad i = 1, \ldots, 2^m \}$$  

(4.102)

with $\alpha_i = \pm 1$ is a compact convex subset of $N_{\eta}$. Since the boundary of $H_N$ is specified by equation

$$\alpha_1 \eta_1 + \cdots + \alpha_m \eta_m = 1$$  

(4.103)

we can check easily whether those $\rho_0, \rho_1, \ldots, \rho_m$ (or $\eta_1, \ldots, \eta_m$) obtained from $a_{n-m+1}, \ldots, a_m$ form a valid covariance sequence of a MA model.
Another convenient sufficient condition can be obtained from the set $H_N$. Since the nonnegative region of $f(\omega)$ in terms of $\rho = [\rho_0, \rho_1, \ldots, \rho_m]^T$ is a closed convex cone, the angle between the axis ($\rho_0$ axis) and a ray on the surface of the cone can be found as

$$\beta = \cos^{-1} \frac{\rho_0}{\sqrt{\rho_0^2 + \cdots + \rho_m^2}}. \quad (4.104)$$

We can determine the minimum value of $\beta$ for those values of $\rho$ corresponding to the boundary of the set $H_N$. Without loss of generality, we assume $\rho_0 = 1$, thus $\rho_i = \eta_i$ for $i = 1, \ldots, m$. Now we can write $\beta$ as

$$\beta = \cos^{-1} \frac{1}{\sqrt{1 + \eta_1^2 + \cdots + \eta_m^2}}. \quad (4.105)$$

We can find the minimal value of $\beta$ under the constraint (4.103). Equivalently we need to find the minimal value of $\sum_{i=1}^m \eta_i^2$ under the constraint (4.103). Let us define

$$Q = \sum_{i=1}^m \eta_i^2 + 2\lambda(1 - \sum_{i=1}^m \alpha_{ik}\eta_i), \quad (4.106)$$

where $\lambda$ is the Lagrange multiplier. With the condition $\alpha_{ik} = \pm 1$, it is easy to show that $Q$ reaches its minimal value at $\eta_i = \frac{\alpha_{ik}}{m}$. Therefore, at $\eta_i = \frac{\alpha_{ik}}{m}$ the corresponding $\beta$ can be easily found as $\beta_{\text{min}} = \cos^{-1}\sqrt{\frac{m}{m+1}}$.

Note that $\beta$ is also a function of $a_{n-m+1}, \ldots, a_m$. Thus, if the $\beta$ value is less than or equal to $\beta_{\text{min}}$ for some $a_{n-m+1}, \ldots, a_m$ values, then there exist NND solutions. In practice, minimizing the angle $\beta$ increases the possibility of the existence of NND realizations.
4.6 Examples

In this section we present three examples to show the theories developed in this chapter. Here we concentrate mainly on the second order \((m = 2)\) models, so we can see geometrically what is going on. The first two examples are simple for the stability study; in fact, we can find the stability intervals for \(c_4\) by simple computations. However, the nonnegativity intervals for \(c_4\) are not simple at all.

Example 4.7:

Given \(C^{(3)} = \{5, 2, 3, 1\}\), for \(m = 2\), find the stability intervals and the nonnegativity intervals for \(C^4\), respectively.

We only have one equation from equation (4.26)

\[-c_3 = c_2a_1 + c_1a_2 \quad \text{or} \quad 3a_1 + 2a_2 = -1\]

for obtaining the coefficients of \(A_2(z)\). However, the stability region for this case is very simple. We have plotted this equation on the stability region (see Figure 21). This equation cuts the boundary of the stability region at \([a_1, a_2] = [-1, 1]\) and \([0.2, -0.8]\). The stability interval for \(c_4\) can be found from these two points using (see equation (4.26))

\[c_4 = -[c_3a_1 + c_2a_2] = -[a_1 + 3a_2]\]

as \((-2, 2.2)\). Thus, any \(c_4 \in (-2, 2.2)\) gives a stable realization of order 2.

Now let us find the nonnegativity intervals for \(c_4\). From equation (4.57) we have
Figure 21: Equation $3a_1 + 2a_2 = -1$ cuts the stability region.
\[ f(\omega, c_4) = f_0(\omega) + f_1(\omega)c_4 + f_2(\omega)c_4^2, \]

with

\[ f_0(\omega) = 4.9184 - 2.1633 \cos \omega + 5.7143 \cos 2\omega, \]

\[ f_1(\omega) = -2.3673 + 4.1224 \cos \omega - 3.1429 \cos 2\omega, \]

\[ f_2(\omega) = 0.8367 - 0.8980 \cos \omega. \]

Figure 22 shows the function \( f(\omega, c_4) \) for \( \omega \in [0, \pi] \) and \( c_4 \in [-5, 10] \). For any fixed \( \omega \) we can see that \( f(\omega, c_4) \) is a quadratic function of \( c_4 \). Figure 23 shows the locus of \([\rho_1/\rho_0, \rho_2/\rho_0]\) of the MA spectrum function \( \rho_0 + \rho_1 \cos \omega + \rho_2 \cos 2\omega \) as \( c_4 \) changes from \(-30\) to \( 10 \) on the NND region. As we can see that there are six \( c_4 \) values as end points. Also we can see that as \( c_4 \rightarrow \pm \infty \), \([\rho_1/\rho_0, \rho_2/\rho_0]\) falls outside the NND region. The NND intervals for \( c_4 \) are thus found as

\[ [-27.667, -2.878], [0.614, 2.2], [3.352, 5]. \]

The the PR region for \( c_4 \) as the intersection of the stability interval \((-2, 2.2)\) and the NND intervals is \((0.614, 2.2)\).

**Example 4.8:**

Given \( C^{(3)} = \{5, 2, 1, 1\} \), for \( m = 2 \), find the stability intervals and the nonnegativity intervals for \( c_4 \), respectively.

Using the method of the previous example, the stability interval for \( c_4 \) is found to be \((1/3, 1)\).
Figure 22: The 3D plot of $f(\omega, c_4)$ of $C^{(3)} = \{5, 2, 3, 1\}$ for $\omega \in [0, \pi]$ and $c_4 \in [-5, 10]$. 
Figure 23: The locus of $[\rho_1/\rho_0, \rho_2/\rho_0]$ as $c_4$ changes from $-30$ to $10$ with $C^{(3)} = \{5, 2, 3, 1\}$. 
Figure 24 shows the function $f(\omega, c_4)$ for $\omega \in [0, \pi]$ and $c_4 \in [-5, 10]$. For any fixed $\omega$ we can see that $f(\omega, c_4)$ is a quadratic function of $c_4$. Figure 25 shows the locus of $[\rho_1/\rho_0, \rho_2/\rho_0]$ as $c_4$ changes from $-5$ to $5$. As we can see that there are four $c_4$ values as end points. In this case, we can see that as $c_4 \Rightarrow \pm \infty$, $[\rho_1/\rho_0, \rho_2/\rho_0]$ falls inside the NND region. The NND intervals for $c_4$ can be found as

$$[-\infty, 1/3], \ [0.429, 1], \ [1.154, \infty].$$

Finally, the the PR region for $c_4$ as the intersection of the stability interval $(1/3, 1)$ and the NND intervals is $(0.429, 1)$.

**Example 4.9:**

Given $C^{(5)} = \{10, 2, 3, 4, 1, 5\}$, for $m = 3$, find the stability intervals for $c_6$.

Solving the eigenvalue equation for this case we have the following three possible end points for the stability intervals:

$$-12.6307, \ 0.7653, \ 1.75.$$

$A(z)$ will be degenerated for $c_6 = -0.2083$.

After checking the stability condition, we only have one stability interval available for $c_6$ as $(-0.2083, 0.7653)$.

**4.7 Approximate Stochastic Partial Realizations**

As discussed in the previous sections, given an arbitrary covariance sequence $C^{(n)}$ with $m < n < 2m$, it is possible that we can not realize $C^{(n)}$ with an ARMA model of order $m < n$. Because of this, Makhoul and Steinhardt [40]
Figure 24: The 3D plot of $f(\omega, c_4)$ of $C^{(3)} = \{5,2,1,1\}$ for $\omega \in [0, \pi]$ and $c_4 \in [-5,10]$. 
Figure 25: The locus of \([\rho_1/\rho_0, \rho_2/\rho_0]\) as \(c_4\) changes from \(-5\) to \(5\) with \(C^{(3)} = \{5, 2, 1, 1\}\).
concluded that pole-zero modeling based on minimizing some error criterion might be preferable to exact constraint matching [44,47]. In practice, we may only have the estimates of a covariance sequence. Depending on problems and algorithms used, these estimates may not satisfy those properties (like non-negativity) of the covariance sequence. This makes approximate stochastic partial realization (ASPR) very attractive in practical applications.

There are many algorithms for obtaining a stochastic realization from an approximate covariance sequence [44]-[47]. An iterative algorithm based on minimizing the following error criterion can be found in [47],

\[ F_Q = (R - c)^T Q (R - c) \]  

(4.111)

where \( R = [R_0, R_1, \ldots, R_n]^T \) is the given set of (estimated or noisy) covariances, \( c = [c_0, c_1, \ldots, c_n]^T \) corresponds to a covariance sequence obtained from an ARMA model of order \( m \), and \( Q \) is a nonnegative definite weighting matrix. In this method, no structure (such as positive definiteness) is assumed on the given sequence \( C^{(n)} \). Even though this method gives an optimal solution to the ASPR problem, it is difficult to use because of convergence to local optimal solutions and tremendous computation efforts.

In many cases, the estimated covariance sequence is a good approximation to the underlying process even if it does not satisfy the stability or the nonnegativity or even the both conditions. For these cases, the optimization methods developed in Chapter II and Chapter III may be applied. The methods here only give suboptimal results in the sense of (4.111) but they are easy to implement. Thus, the methods in Chapter II and III provide a computationally simple alternative to ASPR modeling.
If an estimate of a covariance sequence is given, and if one attempts to obtain a stochastic realization of order $m$, then one of the following three cases may occur:

1. The stability condition is not satisfied, that is, $A_m(z)$ has at least one zero outside the unit circle.

2. The nonnegativity condition is not satisfied, that is,

$$A_m(z)D_m(z^{-1}) + A_m(z^{-1})D_m(z) \geq 0, |z| = 1$$

(4.112)

is not satisfied.

3. Both conditions above are not satisfied.

To obtain a valid stochastic realization, we apply the following procedure. For the first case, only the stability procedure is applied to $A_m(z)$ while keeping $A_m(z)D_m(z^{-1}) + A_m(z^{-1})D_m(z)$ unchanged. For the second case, only the nonnegativity procedure is applied to $A_m(z)D_m(z^{-1}) + A_m(z^{-1})D_m(z)$ while keeping $A_m(z)$ unchanged. For the third case we will first apply the stability procedure to $A_m(z)$, then apply the nonnegativity procedure to $A_m(z)D_m(z^{-1}) + A_m(z^{-1})D_m(z)$ using either original or stabilized $A_m(z)$ polynomial. A combination of the optimal and the suboptimal methods could be used to reduce the computation time when using the optimal method alone. This gives another suboptimal method.

In the following example, the results obtained from these suboptimal methods are compared to the optimal method developed in [47].

**Example 4.10:**

In this example, we use the estimated covariance $C^{(4)} = [5, 2, 3, 1, -2.1]$. 
It is easy to show $C^{(4)}$ is not positive definite. If we realize this sequence with a second order rational function we have

$$v(z) = \frac{D_2(z)}{A_2(z)} = \frac{2.5z^2 - 0.571z + 3.55}{z^2 - 1.029z + 1.043},$$

(4.113)

and with $z = e^{j\omega}$,

$$v(z) + v(z^{-1}) = \frac{13.580 - 14.780 \cos \omega + 12.314 \cos 2\omega}{A_2(z)A_2(z^{-1})}.$$  

(4.114)

It is obvious that $A_2(z)$ is not stable and the corresponding MA spectrum is not non-negative.

For the following optimization procedures, we restrict the Schur parameters in the range $-0.95 \leq r \leq 0.95$ for stability checks. This forces the stabilized polynomial to have zeros strictly inside the unit circle. We choose $Q$ as the identity matrix.

If we use the method developed in [47], we obtain an ARMA(2, 2) model as

$$H_2(z) = \frac{B_2(z)}{A_2(z)} = \frac{1.394z^2 - 1.960z + 1.394}{z^2 - 1.603z + 0.934},$$

(4.115)

and the corresponding first five covariances as

$$C^{(4)} = [5.165, 2.212, 1.608, 0.605, 0.452].$$

If we use the suboptimal method we proposed, we obtain

$$v(z) = \frac{D_2(z)}{A_2(z)} = \frac{11.200z^2 - 0.320z - 4.702}{z^2 - 1.029z + 0.95},$$

(4.116)

$$v(z) + v(z^{-1}) = \frac{14.123 - 14.613 \cos \omega + 11.875 \cos 2\omega}{A_2(z)A_2(z^{-1})},$$

(4.117)
and the corresponding first five covariances as

\[ C^{(4)} = [22.4, 11.2, -3.822, -14.572, -11.358]. \quad (4.118) \]

Figure 26 shows the spectral density functions for the original case (not NND), the suboptimal case corresponding to \( C^{(4)} \) in (4.118), and the optimal case corresponding to the method in [47]. For the suboptimal method, the nonnegativity step uses the original \( A_2(z) \) polynomial. We can see that the method we proposed here gives a large \( L_2 \) distance measure from the original "spectrum" \( S_g(\omega) \). However, from Figure 26 we see that the solution from this method basically retains the original spectrum function's characteristics, such as pole and zero locations.

The large error from the suboptimal method in Figure 26 is primarily caused by the denominator polynomial.

Figure 27 shows the results of the suboptimal method (a combination of the optimal and the suboptimal methods) with \( A_2(z) \) fixed at \( z^2 - 1.029z + 0.95 \) (which is the optimal denominator polynomial, see (4.116)). Thus the optimization method [47] is applied only to \( B_2(z) \). This gives a suboptimal solution since \( A_2(z) \) is fixed. However, the optimization procedure is performed on a smaller dimension. The other two curves in Figure 27 are the same as in Figure 26. In this case, the suboptimal procedure gives an ARMA model

\[ H_2(z) = \frac{B_2(z)}{A_2(z)} = \frac{2.097z^2 - 1.973z + 2.097}{z^2 - 1.029z + 0.95}, \quad (4.119) \]

and the corresponding first five covariances as
Figure 26: Original, optimal, and suboptimal spectral density functions corresponding to Example 4.10.
Figure 27: Original, optimal, and suboptimal spectral density functions using $A_2(z) = z^2 - 1.029z + 0.95$ for the suboptimal method.
From above discussions, the optimal method gives a smaller distance measure in the covariance space while the suboptimal method gives smaller distance measure in the sense of retaining zero-pole locations of the original spectrum. It is thus not clear which of the two methods is better from an approximate point of view.

4.8 Conclusions

We have considered the problem of obtaining minimal order stochastic partial realizations; We consider the problem differently from other approaches: we considered the stability and the nonnegativity conditions of the PR function \( v(z) \) separately, so we can handle the minimal order requirement easily.

For the special case \( n = 2m - 1 \), we first found the ranges of \( c_{2m} \) in which \( v(z) \) is a stable function. It turns out that the boundary values of \( c_{2m} \) can be obtained from the eigenvalues of some Hurwitz matrices. Then we found the ranges of \( c_{2m} \) for which \( v(z) + v(z^{-1}) \) is nonnegative on the unit circle. We need to perform some nonlinear minimization procedures to find those boundary values. If either one of these two kinds of ranges is empty, no realization of minimal order \( m \) exists. For a minimal order \( m \) realization, we need the intersection of these two kinds of ranges to be nonempty.

For an arbitrary data length \( n \) \((m < n < 2m - 1)\), we can only give some necessary conditions for the stability and the nonnegativity requirements. These necessary conditions are easy to use in practice. It is interesting that the coefficients of the MA part spectrum have to be inside the corresponding sta-
bility set. Future work should concentrate on finding necessary and sufficient conditions.
CHAPTER V

SUMMARY

This dissertation has considered three problems which arise in the modeling of stochastic time series: the stability problem, the nonnegativity problem, and the minimal order stochastic partial realization problem.

First, we considered the problem of optimally stabilizing an unstable polynomial. This problem arises in the SPR problem because the denominator polynomial $A(x)$ in an ARMA time series model is obtained from either estimated covariances or true covariances of a higher order system; in either case, this denominator polynomial may be unstable. We developed an efficient algorithm for finding the closest stable polynomial to a given unstable one. The closeness is measured by weighted Euclidean distance in the polynomial coefficient space. Since the stability set in terms of polynomial coefficients is not convex, the optimal solution is not unique in general. The essence of this method is to perform the nonlinear optimization in the Schur parameter space. By working in the Schur parameter space we can easily handle the stability requirement and we can perform the optimization in a computationally very efficient manner.

Second, we considered the problem of optimally adjusting a spectral estimate $g(\omega)$ which is not nonnegative to one which is nonnegative. We considered trigonometric polynomial $g(\omega)$ functions which correspond to a moving aver-
age spectral estimate. This problem arises in MA spectral estimation, time series modeling, and in some two step ARMA spectral estimation methods where the second step consists essentially of an MA spectral estimation procedure. In these applications, it is possible that the given covariance sequence corresponding to $g(\omega)$ is not NND; we found the closest NND covariance sequence to the given one. We developed an algorithm which is based on a set of constrained minimization problems, each parameterized by the zero frequencies of the spectral density function corresponding to the optimal solution. As a result, the nonlinear optimization is performed on a smaller dimension comparing to the other methods. In addition, by considering the problem in this space, we characterized some of the geometrical properties of the optimal solution in terms of the location of its zero frequencies. Future work should be concentrated on the tighter locations of the zero frequencies based on the characteristics of the given estimate $g(\omega)$. The first and second derivatives of $g(\omega)$ should give more information about the zero frequency locations; this information is not used in the present work.

Third, we considered the problem of minimal order stochastic partial realization. Here we are give a partial covariance sequence $C^{(n)}$, and one objective is to find all extensions of $C^{(n)}$ such that the stochastic realization is of minimal order. In the case where the extension is not unique, we wish to find all possible minimal order extensions. The minimal order requirement was handled in our case by working on two separate sets, namely, the stability set and the NND set. We developed a set of necessary conditions for the existence of covariance extensions of minimal order by separately considering necessary conditions for stable extensions and for NND extensions. In addition, for the
simplest case (namely, when $n = 2m - 1$ and the minimal order is $m$), we were able to provide a complete solution to the SPR problem. Even in this case, though, the solution procedure was tedious. Future work in this area could be concentrated on finding necessary and sufficient conditions for the general case.
REFERENCES


