FUNCTIONAL TECHNIQUES FOR DATA ANALYSIS

by

JOHN R. TOMLINSON, B.S.Phys., M.A.

A DISSERTATION

IN

MATHEMATICS

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for
the Degree of

DOCTOR OF PHILOSOPHY

Approved

Chairperson of the Committee

Accepted

Dean of the Graduate School

December, 1997

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support and assistance of Clyde Martin, the thesis advisor, and the committee members: W. P. Dayawansa, James Dunyak, Lawrence Schovanec, and Zhimin Zhang. The author would also like to thank the Texas Tech Graduate School for the support they provided in the form of a Summer Research Fellowship in 1997.

CONTENTS

ABSTRACT

This dissertation develops a new general method of solving Prony's problem. Two special cases of this new method have been developed previously. They are the Matrix Pencil and the Osculatory Interpolation. The dissertation shows that they are instances of a more general solution type which allows a wide ranging class of linear functional to be used in the solution of the problem. This class provides a continuum of functionals which provide new methods that can be used to solve Prony's problem.

LIST OF TABLES

CHAPTER 1
INTRODUCTION

Many real-world processes obey constant coefficient differential equations. The measurable quantities from these processes are described by exponential functions. Certain of these processes, such as radioactive decay, demand to be modeled using real decay factors.

The scientist is often presented with the task of identifying the material in question using whatever measurement capabilities are at his disposal. The rate of radiation from a mass is one of the measureable quantities available. This leads to the task of determining the real decay factors from measured data. The problem then, is to identify the real coefficients b_k in a real-valued function of the form

$$f(t) = \sum_{k=1}^{p} a_k e^{b_k t}, \tag{1.1}$$

from data. In order to make the problem more tractable, the data are acquired at uniform time intervals. This problem was first considered and a solution presented by Baron R. deProny, a Parisian engineer, in 1795 [85].

Prony had limited success in solving the problem. This is because the problem is highly nonlinear. In addition, unlike the problem of identifying a function with sums of sines and cosines, the basis functions of this problem, the real exponentials, are not orthogonal [56]. This has resulted in repeated attempts at solving the problem since Prony [89, 56, 29, 38, 73, 2, 39, 9, 83].

In recent years, this problem has generated considerable interest. Until recently, the solutions have fallen into two categories: nonlinear gradient descent methods, and autoregressive methods. The gradient descent methods are very costly and don't guarantee global convergence. Thus, most interest has focussed on the autoregressive or AR model. The use of the AR model stems from the observation that functions of the form (1.1) are, in the case of integer time intervals, solutions to some difference equation with constant coefficients. Only recently has this type of equation been

called autoregressive. However, a full theory has grown up around the AR problem. Fundamental to the theory is the assumption that the process is stationary. Disturbances, known and unknown, are allowed in the process. Various levels of success have been reported for the AR approach to the problem.

Recently, Ammar et al. [2] and Hua and Sarkar [39] developed a matrix pencil method for determining the coefficients b_k. The matrix pencil method performs well at finding the coefficients when noise levels are low, but is not as tolerant of noise as many of the AR methods. It is, however, directly related to the AR method proposed by Prony [79].

The pencil method and the osculatory method developed by Martin et al. [65] are both specialized examples of a new class of methods for solving this problem. This new class of solutions is termed the linear functional method. By applying a linear functional to the data, the data are modified in such a way that the inverse of the functional applied to the eigenvalues of a certain generalized eigenvalue problem yield the desired coefficients. For the functionals tried so far, the functional method doesn't perform well with noisy data.

1.1 Background

A time varying quantity is measured in many physical problems, for instance the temperature of a heated object at room temperature, or the rate of radiation from a radioactive object.

In the case of a radioactive object, the rate of radiation is determined by the mixture of materials.

Assumption 1. *The mixtures of materials are such that the radiation rates remain in the linear region.*

Each type of radioactive material loses mass at a rate which is different than that of others. One task of interest then, is the determination of the different rates of decay of matter in the object.

Assume that the rate of decay of a homogenous mass is related to the amount of radioactive material in the object. Designate the rate of decay at a particular time, t, by $\rho(t)$, the mass of remaining radioactive material by $M(t)$ and let γ be a constant, then

$$\rho(t) = \gamma M(t),$$

and

$$\rho(t) = \dot{M}(t),$$

so

$$\dot{M}(t) = \gamma M(t).$$

Thus

$$M(t) = M_0 e^{\gamma t}, \tag{1.2}$$

where M_0 is the mass of the radioactive material at time $t = 0$.

Now, if the radioactive object is composed of more than one type of radioactive material, equation (1.2) is no longer theoretically sound. Assuming that the radiation from one material does not affect the rate of radiation from another material, the rates will be additive and the mass of the object at time t becomes

$$M(t) = \sum_{k=1}^{p} a_k e^{b_k t}, \tag{1.3}$$

where the a_k's represent the initial mass of each of the component materials, and the b_k's represent the decay rates for each of the materials in the mix. Of course the assumption that the mixture of materials does not affect the decay rates of the matter in the mix is false. However, in the mix, new decay rates will exist for each type of matter and if the constants are adjusted, the formula (1.3) is still representative of the behaviour of the object. By differentiation,

$$\rho(t) = \sum_{k=1}^{p} a_k^* e^{b_k t},$$

so that the problem is the same whether the mass or radiation is being measured. In both cases, the desired quantities are the b_k's.

1.2 Roadmap

The rest of the dissertation is organized as follows. The second chapter presents some fundamental results. These results establish the basic theory and notation which is used throughout the dissertation. Chapter 3 presents the linear functional method. This method is the primary result of the dissertation. Once it is in place, Chapter 4 gives details on some possibilities for using the functional method. Chapter 5 provides some numerical comparisons between some methods based on the functional technique and some standard methods which appear elsewhere.

CHAPTER 2
FOUNDATION

This chapter presents some basic results and prepares the groundwork for the rest of the dissertation. Some notational shortcuts are presented to make the later developments shorter.

2.1 The Vandermonde

Lemma 1. *If $z_1, z_2, z_3, \ldots, z_p$ are distinct complex numbers, the Vandermonde matrix*

$$V = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ z_1 & z_2 & z_3 & \ldots & z_p \\ z_1^2 & z_2^2 & z_3^2 & \ldots & z_p^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ z_1^{p-1} & z_2^{p-1} & z_3^{p-1} & \ldots & z_p^{p-1} \end{bmatrix}$$

is nonsingular.

Proof. This standard result can be found in Atkinson [4]. ■

Lemma 2. *If $z_1, z_2, z_3, \ldots, z_p$ are distinct and $n \geq p$, the Vandermonde matrix*

$$V = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ z_1 & z_2 & z_3 & \ldots & z_p \\ z_1^2 & z_2^2 & z_3^2 & \ldots & z_p^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ z_1^n & z_2^n & z_3^n & \ldots & z_p^n \end{bmatrix}$$

has full column rank.

Proof. This follows from the previous lemma since V has rank p. ■

This dissertation focuses on numbers that can be represented by exponentials. Since any complex number except zero can be represented as an exponential, the z_i's in the previous two lemmas are restricted to be nonzero complex numbers. Under this restriction, the lemmas still hold.

2.2 The Difference Equation

Consider the function

$$f(t) = \sum_{k=1}^{p} a_k e^{b_k t}. \tag{2.1}$$

Assumption 2. *The b_k's are distinct.*

Lemma 3. *The functions $f(t)$ of the form (2.1) satisfy a difference equation of the form*

$$f(t+n) + \alpha_{p-1} f(t+n-1) + \cdots + \alpha_0 f(t+n-p) = 0. \tag{2.2}$$

Furthermore, the roots, λ_k, of the polynomial

$$\lambda^p + \alpha_{p-1}\lambda^{p-1} + \cdots + \alpha_0 = g(\lambda) \tag{2.3}$$

are the exponential factors in (2.1). In other words,

$$\lambda_k = e^{b_k}.$$

Proof. At time $t+m$,

$$f(t+m) = \sum_{k=1}^{p} a_k e^{b_k t} e^{b_k m} \tag{2.4}$$

and (2.1) becomes

$$f(t) = [1 \quad 1 \quad \dots \quad 1] \begin{bmatrix} e^{b_1 t} & & & \\ & e^{b_2 t} & & \\ & & \ddots & \\ & & & e^{b_p t} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}. \tag{2.5}$$

For $f(t), f(t+1), \dots, f(t+p-1)$,

$$\begin{bmatrix} f(t) \\ f(t+1) \\ \vdots \\ f(t+p-1) \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{b_1} & e^{b_2} & \dots & e^{b_p} \\ \vdots & \vdots & \vdots & \vdots \\ e^{b_1(p-1)} & e^{b_2(p-1)} & \dots & e^{b_p(p-1)} \end{bmatrix} \begin{bmatrix} e^{b_1 t} & & & \\ & e^{b_2 t} & & \\ & & \ddots & \\ & & & e^{b_p t} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

$$\boldsymbol{f}(t) = V\Gamma(t)\boldsymbol{a}. \tag{2.6}$$

Note that V is a Vandermonde matrix, and, as a direct result of the assumption, nonsingular. Thus, the vector $\boldsymbol{a}$ can be expressed as

$$\boldsymbol{a} = \Gamma(t)^{-1} V^{-1} \boldsymbol{f}(t).$$

Then, according to (2.4) with the notation introduced in (2.6),

$$\boldsymbol{f}(t+1) = V\Gamma(t+1)\boldsymbol{a} = V\Gamma(t+1)\Gamma(t)^{-1}V^{-1}\boldsymbol{f}(t).$$

But using the definition of $\Gamma(t)$ as a diagonal matrix,

$$\Gamma(t+1)\Gamma(t)^{-1} = \Gamma(1) = \begin{bmatrix} e^{b_1} & & & \\ & e^{b_2} & & \\ & & \ddots & \\ & & & e^{b_p} \end{bmatrix}$$

so

$$\boldsymbol{f}(t+1) = V\Gamma(1)V^{-1}\boldsymbol{f}(t). \tag{2.7}$$

Indeed, the characteristic equation of the matrix $\Gamma(1)$ is given by

$$g(\lambda) = \prod_{k=1}^{p}(\lambda - e^{b_k}) = \lambda^p + \sum_{k=0}^{p-1}\alpha_k\lambda^k. \tag{2.8}$$

Let

$$A = V\Gamma(1)V^{-1}.$$

Then

$$A \sim \Gamma(1),$$

and has the same characteristic equation. Writing out (2.7):

$$\begin{bmatrix} f(t+1) \\ f(t+2) \\ \vdots \\ f(t+p) \end{bmatrix} = A \begin{bmatrix} f(t) \\ f(t+1) \\ \vdots \\ f(t+p-1) \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \\ -a_0^* & -a_1^* & \dots & & -a_{p-1}^* \end{bmatrix} \begin{bmatrix} f(t) \\ f(t+1) \\ \vdots \\ f(t+p-1) \end{bmatrix}. \tag{2.9}$$

Observe that the matrix A is in companion form; thus the factors, a_k^*, in the matrix A are equal to the coefficients, α_k, of its characteristic polynomial. Designate $f(n) = y_n$, then

$$y_n + \alpha_{p-1} y_{n-1} + \cdots + \alpha_0 y_{n-p} = 0, \tag{2.10}$$

which is a difference equation. Thus, the functions (2.1) satisfy the difference equations (2.10). In addition, from (2.9), the coefficients b_k of the function (2.1) are related to the coefficients, α_k, of the difference equation by the relation (2.8). That is, if λ_i is a root of the polynomial $g(\lambda)$, then

$$\lambda_i = e^{b_k} \tag{2.11}$$

for one of the b_k's. ■

2.3 The Prony Step

The preceding proof has a few details worth mentioning. First, the step of the derivation in which (2.6) was solved for $\boldsymbol{a}$ and the result put into (2.7) is very common in the derivations of this thesis. It is called the "Prony Step" because it effectively transforms a nonlinear problem into a linear one. The second aspect of the proof worth

noticing is that a diagonalization of the companion matrix is obtained provided the eigenvalues are distinct. If they are not distinct, then the Vandermonde is singular and this diagonalization does not hold.

Lemma 4. *Let*

$$G = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \\ -\alpha_0 & -\alpha_1 & \dots & & -\alpha_{p-1} \end{bmatrix}.$$

If the eigenvalues, $\lambda_1, \lambda_2, \dots, \lambda_p$, of G are distinct, then $G = V\Lambda V^{-1}$, where

$$V = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_p \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{p-1} & \lambda_2^{p-1} & \dots & \lambda_p^{p-1} \end{bmatrix},$$

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix},$$

and the columns of V are eigenvectors of G.

Proof. Consider GV. Without loss of generality consider the first column of the result.

$$\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \\ -\alpha_0 & -\alpha_1 & \dots & & -\alpha_{p-1} \end{bmatrix} \begin{bmatrix} 1 \\ \lambda_1 \\ \lambda_1^2 \\ \vdots \\ \lambda_1^{p-1} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_1^2 \\ \lambda_1^3 \\ \vdots \\ \lambda_1^{p-1} \\ -\alpha_0 - \alpha_1\lambda_1 - \alpha_2\lambda_1^2 - \dots - \alpha_{p-1}\lambda_1^{p-1} \end{bmatrix}$$

But, since λ_1 is a root of the characteristic equation of G,

$$-\alpha_0 - \alpha_1\lambda_1 - \alpha_2\lambda_1^2 - \cdots - \alpha_{p-1}\lambda_1^{p-1} = \lambda_1^p.$$

Thus

$$\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \\ -\alpha_0 & -\alpha_1 & \dots & & -\alpha_{p-1} \end{bmatrix} \begin{bmatrix} 1 \\ \lambda_1 \\ \lambda_1^2 \\ \vdots \\ \lambda_1^{p-1} \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_1^2 \\ \lambda_1^3 \\ \vdots \\ \lambda_1^p \end{bmatrix} = \lambda_1 \begin{bmatrix} 1 \\ \lambda_1 \\ \lambda_1^2 \\ \vdots \\ \lambda_1^{p-1} \end{bmatrix}.$$

Then the first column of GV is the first column of $V\Lambda$. The proof for the other columns is similar. Thus

$$GV = V\Lambda.$$

■

2.4 Notation

With these results in place, the standard notation to be used throughout this dissertation is presented.

2.4.1 Model

The model for the analysis is

$$f(t) = \sum_{k=1}^{p} a_k e^{b_k t}. \tag{2.12}$$

The t are taken to occur at the integers $0, 1, 2, 3, \ldots$. Denote the function values at the times $t = n$ by

$$y_n = f(n).$$

2.4.2 Functionals

In the following, examples will be given in which different functionals are used. The functionals will typically be given by a rule. In order to shorten the presentation, whenever there is no ambiguity, the values which result from applying the functionals to the data will be represented by $\tilde{y}$. For example, if the functional $\mathcal{F}$ is applied to the data,

$$\tilde{y}_n = \mathcal{F}(y_n).$$

2.4.3 Hankel Matrices

The following is the template for solution of the inverse problem of determining the coefficients α_k of the difference equation

$$y_n + \alpha_{p-1} y_{n-1} + \cdots + \alpha_0 y_{n-p} = 0.$$

Let

$$H_0 = \begin{bmatrix} y_0 & y_1 & \cdots & y_{p-1} \\ y_1 & y_2 & \cdots & y_p \\ \vdots & \vdots & \vdots & \vdots \\ y_{p-1} & y_p & \cdots & y_{2p-2} \end{bmatrix}$$

be the square Hankel matrix formed from the data. Then, according to the difference equation, if

$$\boldsymbol{\alpha} = \begin{bmatrix} -\alpha_0 \\ -\alpha_1 \\ \vdots \\ -\alpha_{p-1} \end{bmatrix},$$

and

$$\boldsymbol{y}_p = \begin{bmatrix} y_p \\ y_{p+1} \\ \vdots \\ y_{2p-1} \end{bmatrix},$$

then

$$H_0\boldsymbol{\alpha} = \boldsymbol{y}_p.$$

H_0 is always the designation for the Hankel matrix of the data. It will sometimes be convenient to use H_0 for the corresponding nonsquare Hankel matrix when there are more than $2p$ data values. When needed, H_1 represents the Hankel matrix made of values other than the data.

For example, the Hankel matrix of data values starting at the second value, y_1, instead of y_0 is

$$H_1 = \begin{bmatrix} y_1 & y_2 & \cdots & y_p \\ y_2 & y_3 & \cdots & y_{p+1} \\ \vdots & \vdots & \vdots & \vdots \\ y_p & y_{p+1} & \cdots & y_{2p-1} \end{bmatrix}.$$

Now, consider the solution problem which appears repeatedly in this dissertation. For some matrix H_1, solve the equation

$$H_0 G = H_1, \tag{2.13}$$

and determine the eigenvalues of the matrix G. To make this process a little clearer, consider the result of multiplying the equation on the right by an eigenvector of G.

$$\begin{aligned} H_0 G\boldsymbol{x} &= H_1\boldsymbol{x} \\ H_0\lambda\boldsymbol{x} &= H_1\boldsymbol{x} \\ (H_0\lambda - H_1)\boldsymbol{x} &= 0, \end{aligned} \tag{2.14}$$

which implies that $\boldsymbol{x}$ and λ are eigenvector and eigenvalue of the generalized eigenvalue equation (2.14).

Now suppose that equation (2.14) has been solved for the eigenvalue and eigenvector. Then

$$H_0\lambda\boldsymbol{x} = H_1\boldsymbol{x}.$$

If H_0 is nonsingular, then this equation has the solution

$$\lambda \boldsymbol{x} = H_0^{-1} H_1 \boldsymbol{x}.$$

But, in this case, (2.13) gives

$$G = H_0^{-1} H_1.$$

Thus, the lemma:

Lemma 5. *If the matrix H_0 is nonsingular, then $(\lambda, \boldsymbol{x})$ are an eigenpair of the generalized eigenvalue equation*

$$(H_0 \lambda - H_1)\boldsymbol{x} = 0 \qquad \boldsymbol{x} \neq 0$$

if and only if $(\lambda, \boldsymbol{x})$ are an eigenpair of the matrix

$$G = H_0^{-1} H_1.$$

With this representation, it is obvious that it will not be necessary to determine the matrix G in order to find its eigenvalues. Similar results can be proven for the overdetermined case [106]. In that case the appropriate generalized eigenvalue equation is

$$(H_0^{\mathsf{T}} H_0 \lambda - H_0^{\mathsf{T}} H_1)\boldsymbol{x} = 0 \qquad \boldsymbol{x} \neq 0.$$

CHAPTER 3
THE LINEAR FUNCTIONAL METHOD

3.1 Introduction

This chapter develops the main result of the dissertation. The functional method is a level of abstraction which permits complicated analyses of data with minimal effort. To make the presentation more straightforward, assume the data perfectly match the model

$$f(t) = \sum_{k=1}^{p} a_k e^{b_k t},$$

with the samples taken at integer times. This allows focussing on the problem as an interpolation problem even though more than $2p$ data values may be used to determine $2p$ unknown parameters. A later section presents methods which permit relaxing the specification so that the overdetermined inconsistent and the exactly determined interpolation cases can be solved.

3.2 Theory

Let Φ be the set of all functions $\psi(t) : \mathbb{R} \rightarrow \mathbb{C}\backslash 0$, of the form

$$\psi(t) = e^{bt},$$

which have the property that $\psi(t+1) = \psi(t)\psi(1)$. Let $\mathcal{S}$ be the space of all finite linear combinations of elements of Φ. i.e.

$$f(t) \in \mathcal{S} \Rightarrow f(t) = \sum_{k=1}^{p} a_k \phi_k(t) \tag{3.1}$$

where $p \in \mathbb{Z}$ is finite, $a_k \in \mathbb{C}$, and $\phi_k(t) \in \Phi$.

Lemma 6. *If $y_i \in \mathcal{S}$ of the form in (3.1), and $\mathcal{F}$ is a linear functional on $\mathcal{S}$, then*

$$\mathcal{F}(y_i) = \sum_{k=1}^{p} a_k \phi_k(i) \left[\mathcal{F}\phi_k\right](0)$$

or, in matrix form with $[\mathcal{F}\phi_k](0) = [\mathcal{F}\phi_k](i)|_{i=0} = \lambda_k$

$$\mathcal{F}(y(i)) = [\phi_1(i) \quad \phi_2(i) \quad \dots \quad \phi_p(i)] \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}.$$

Proof. First note that

$$y_{i+1} = a_1\phi_1(i)\phi_1(1) + a_2\phi_2(i)\phi_2(1) + \cdots + a_p\phi_p(i)\phi_p(1),$$

and because $\phi_k(1)$ is constant,

$$\mathcal{F}(y_i) = a_1 [\mathcal{F}\phi_1](i) + a_2 [\mathcal{F}\phi_2](i) + \cdots + a_p [\mathcal{F}\phi_p](i)$$

implies

$$\mathcal{F}(y_{i+1}) = a_1 [\mathcal{F}\phi_1](i)\phi_1(1) + a_2 [\mathcal{F}\phi_2](i)\phi_2(1) + \cdots + a_p [\mathcal{F}\phi_p](i)\phi_p(1).$$

So that if

$$\mathcal{F}(y_i) = \begin{bmatrix} [\mathcal{F}\phi_1](i) & [\mathcal{F}\phi_2](i) & \dots & [\mathcal{F}(\phi_p](i)) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix},$$

then

$$\mathcal{F}(y_{i+1}) = \begin{bmatrix} [\mathcal{F}\phi_1](i) & [\mathcal{F}\phi_2](i) & \dots & [\mathcal{F}\phi_p](i) \end{bmatrix} \begin{bmatrix} \phi_1(1) & & & \\ & \phi_2(1) & & \\ & & \ddots & \\ & & & \phi_p(1) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}.$$

Now, induct from $i = 0$

$$\mathcal{F}(y_0) = \begin{bmatrix} [\mathcal{F}\phi_1](0) & [\mathcal{F}\phi_2](0) & \dots & [\mathcal{F}\phi_p](0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

$$= [\lambda_1 \quad \lambda_2 \quad \dots \quad \lambda_p] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} \tag{3.2}$$

so that for any i, from the above,

$$\mathcal{F}(y_i) = [\lambda_1 \quad \lambda_2 \quad \dots \quad \lambda_p] \begin{bmatrix} \phi_1(1) & & & \\ & \phi_2(1) & & \\ & & \ddots & \\ & & & \phi_p(1) \end{bmatrix}^i \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix},$$

$$= [\phi_1(i) \quad \phi_2(i) \quad \dots \quad \phi_p(i)] \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}. \tag{3.3}$$

■

Lemma 7. *The matrix*

$$\begin{bmatrix} \phi_1(0) & \phi_2(0) & \dots & \phi_p(0) \\ \phi_1(1) & \phi_2(1) & \dots & \phi_p(1) \\ \vdots & & & \\ \phi_1(p-1) & \phi_2(p-1) & \dots & \phi_p(p-1) \end{bmatrix}$$

is a Vandermonde matrix and can be represented as

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ z_1 & z_2 & \dots & z_p \\ \vdots & & & \\ z_1^{p-1} & z_2^{p-1} & \dots & z_p^{p-1} \end{bmatrix}$$

or

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{b_1} & e^{b_2} & \dots & e^{b_p} \\ \vdots & & & \\ e^{b_1(p-1)} & e^{b_2(p-1)} & \dots & e^{b_p(p-1)} \end{bmatrix}$$

for some $z_k \in \mathbb{C}$ and $b_k \in \mathbb{C}$. Then, provided $z_k \neq z_j$, when $k \neq j$, the matrix is nonsingular.

Lemma 8. *If $b_1, b_2, \dots, b_p$ are distinct, all $a_1, a_2, \dots, a_p$ are nonzero, and if*

$$y_k = \sum_{j=1}^{p} a_j e^{b_j k}$$

for $k = 1, 2, \dots$, then the matrix

$$H_0 = \begin{bmatrix} y_0 & y_1 & \dots & y_{p-1} \\ y_1 & y_2 & \dots & y_p \\ \vdots & & & \\ y_{p-1} & y_p & \dots & y_{2p-2} \end{bmatrix}$$

is nonsingular.

Proof. Let

$$\boldsymbol{y}_k = \begin{bmatrix} y_k \\ y_{k+1} \\ \vdots \\ y_{k+p-1} \end{bmatrix}.$$

Then, the companion matrix G which relates $\boldsymbol{y}_k = G\boldsymbol{y}_{k-1}$, is nonsingular. Note that

$$\boldsymbol{y}_k = G\boldsymbol{y}_{k-1}$$

implies

$$\boldsymbol{y}_k = G^k\boldsymbol{y}_0.$$

Now, if H_0 is singular, one of the columns can be written as a combination of the remaining columns. Assume that

$$\boldsymbol{y}_k = \sum_{j=0, j\neq k}^{p-1} a_j \boldsymbol{y}_j.$$

Then

$$G^k\boldsymbol{y}_0 = \sum_{j=0, j\neq k}^{p-1} a_j G^j \boldsymbol{y}_0$$

which, with $a_k = -1$, gives

$$0 = \sum_{j=0}^{p-1} a_j G^j.$$

This last result is a polynomial in G of order less than p which is equal to zero. According to the Cayley-Hamilton theorem, zero is one of the roots of the characteristic polynomial for G. But since the eigenvalues of G are $\{e^{b_1}, e^{b_2}, \ldots, e^{b_p}\}$, none of its eigenvalues are zero. This is a contradiction. Therefore, H_0 is nonsingular. ■

Theorem 9. *Given a set of data $y_0, y_1, \ldots, y_n, \ldots$, generated at integer times by a sum of p distinct elements in $\mathcal{S}$, then the matrix*

$$H = \begin{bmatrix} y_0 & y_1 & \cdots & y_{p-1} \\ y_1 & y_2 & \cdots & y_p \\ \vdots & \vdots & \vdots & \vdots \\ y_{p-1} & y_p & \cdots & y_{2p-1} \end{bmatrix}$$

is nonsingular, and the matrix G, given by

$$G = H^{-1} \begin{bmatrix} \mathcal{F}(y_0) & \mathcal{F}(y_1) & \dots & \mathcal{F}(y_{p-1}) \\ \mathcal{F}(y_1) & \mathcal{F}(y_2) & \dots & \mathcal{F}(y_p) \\ \vdots & \vdots & \vdots & \vdots \\ \mathcal{F}(y_{p-1}) & \mathcal{F}(y_p) & \dots & \mathcal{F}(y_{2p-1}) \end{bmatrix}$$

has the the eigenvalues $\{\mathcal{F}(\phi_1(0)), \mathcal{F}(\phi_2(0)), \dots, \mathcal{F}(\phi_p(0))\}$.

Proof. Consider the data vector

$$\boldsymbol{y}_0 = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{p-1} \end{bmatrix}.$$

From the definition of the function which generated the data $\boldsymbol{y}_0$, there is a vector

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix},$$

so that

$$\begin{bmatrix} \phi_1(0) & \phi_2(0) & \dots & \phi_p(0) \\ \phi_1(1) & \phi_2(1) & \dots & \phi_p(1) \\ \vdots & & & \\ \phi_1(p-1) & \phi_2(p-1) & \dots & \phi_p(p-1) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{p-1} \end{bmatrix}$$

$$V\boldsymbol{a} = \boldsymbol{y}_0. \tag{3.4}$$

Then since V is nonsingular,

$$\boldsymbol{a} = V^{-1}\boldsymbol{y}_0.$$

Also, from Lemma 6, if

$$\tilde{\boldsymbol{y}}_0 = \begin{bmatrix} \mathcal{F}(y_0) \\ \mathcal{F}(y_1) \\ \vdots \\ \mathcal{F}(y_{p-1}) \end{bmatrix},$$

then

$$V \begin{bmatrix} \mathcal{F}(\phi_1(0)) & & & \\ & \mathcal{F}(\phi_2(0)) & & \\ & & \ddots & \\ & & & \mathcal{F}(\phi_p(0)) \end{bmatrix} \boldsymbol{a} = \tilde{\boldsymbol{y}}_0 \tag{3.5}$$

or

$$V \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix} \boldsymbol{a} = \tilde{\boldsymbol{y}}_0. \tag{3.6}$$

Where, as before: $\lambda_k = \mathcal{F}(\phi_k(0))$. Then substituting in the "Prony Step,"

$$V \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix} V^{-1} \boldsymbol{y}_0 = \tilde{\boldsymbol{y}}_0. \tag{3.7}$$

Now, designate

$$\boldsymbol{y}_i = \begin{bmatrix} y_i \\ y_{i+1} \\ \vdots \\ y_{i+p-1} \end{bmatrix}, \quad \text{and} \quad \tilde{\boldsymbol{y}}_i = \begin{bmatrix} \mathcal{F}(y_i) \\ \mathcal{F}(y_{i+1}) \\ \vdots \\ \mathcal{F}(y_{i+p-1}) \end{bmatrix}.$$

Claim: (3.7) holds for all i.

Let

$$\Gamma = \begin{bmatrix} \phi_1(1) & & & \\ & \phi_2(1) & & \\ & & \ddots & \\ & & & \phi_p(1) \end{bmatrix},$$

and

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{bmatrix}.$$

Then, note

$$\begin{bmatrix} \phi_1(i) & \phi_2(i) & \dots & \phi_p(i) \\ \phi_1(i+1) & \phi_2(i+1) & \dots & \phi_p(i+1) \\ \vdots & & & \\ \phi_1(i+p-1) & \phi_2(i+p-1) & \dots & \phi_p(i+p-1) \end{bmatrix} = V\Gamma^i,$$

so that,

$$V\Gamma^i \boldsymbol{a} = \boldsymbol{y}_i,$$

and

$$V\Gamma^i \Lambda \boldsymbol{a} = \tilde{\boldsymbol{y}}_i.$$

Then perform the "Prony Step" again:

$$V\Gamma^i \Lambda \Gamma^{-i} V^{-1} \boldsymbol{y}_i = \tilde{\boldsymbol{y}}_i.$$

But since Γ^i and Λ are diagonal,

$$\Gamma^i \Lambda \Gamma^{-i} = \Lambda \Gamma^i \Gamma^{-i} = \Lambda,$$

so

$$V\Lambda V^{-1}\boldsymbol{y}_i = \tilde{\boldsymbol{y}}_i.$$

Let

$$G = V\Lambda V^{-1}.$$

Then,

$$G\boldsymbol{y}_i = \tilde{\boldsymbol{y}}_i$$

for each $i \in \{0, 1, \ldots, p-1\}$. Now, G takes each column of

$$H_0 = \begin{bmatrix} y_0 & y_1 & \cdots & y_{p-1} \\ y_1 & y_2 & \cdots & y_p \\ \vdots & \vdots & \vdots & \vdots \\ y_{p-1} & y_p & \cdots & y_{2p-1} \end{bmatrix}$$

to the corresponding column of

$$H_1 = \begin{bmatrix} \mathcal{F}(y_0) & \mathcal{F}(y_1) & \ldots & \mathcal{F}(y_{p-1}) \\ \mathcal{F}(y_1) & \mathcal{F}(y_2) & \ldots & \mathcal{F}(y_p) \\ \vdots & \vdots & \vdots & \vdots \\ \mathcal{F}(y_{p-1}) & \mathcal{F}(y_p) & \ldots & \mathcal{F}(y_{2p-1}) \end{bmatrix}$$

so

$$GH_0 = H_1.$$

Since $G \sim \Lambda$, the eigenvalues of G are the values

$$\{\mathcal{F}(\phi_1(0)), \mathcal{F}(\phi_2(0)), \ldots, \mathcal{F}(\phi_p(0))\}.$$

■

Corollary 10. *Given a set of data* $y_0, y_1, \ldots, y_n, \ldots$, *generated at integer times by a sum of* p *distinct elements in* $\mathcal{S}$, *the matrix*

$$H = \begin{bmatrix} y_0 & y_1 & \cdots & y_{p-1} \\ y_1 & y_2 & \cdots & y_p \\ \vdots & \vdots & \vdots & \vdots \\ y_{p-1} & y_p & \cdots & y_{2p-1} \end{bmatrix}$$

and the matrix

$$H_1 = \begin{bmatrix} \mathcal{F}(y_0) & \mathcal{F}(y_1) & \ldots & \mathcal{F}(y_{p-1}) \\ \mathcal{F}(y_1) & \mathcal{F}(y_2) & \ldots & \mathcal{F}(y_p) \\ \vdots & \vdots & \vdots & \vdots \\ \mathcal{F}(y_{p-1}) & \mathcal{F}(y_p) & \ldots & \mathcal{F}(y_{2p-1}) \end{bmatrix}.$$

The generalized eigenvalue problem

$$(H_0\lambda - H_1)\boldsymbol{x} = 0$$

has the eigenvalues $\{\mathcal{F}(\phi_1(0)), \mathcal{F}(\phi_2(0)), \ldots, \mathcal{F}(\phi_p(0))\}$.

CHAPTER 4
IMPLEMENTATION

The previous chapter developed a functional theory which may be useful in finding the coefficients of the model equation. This chapter expands on this theme by developing some methods for using the functional approach to determine the coefficients b_k of the model equation. The first section describes the three basic categories of problem: the exact problem in which data are assumed to be perfect, the interpolation problem in which p data values and p functional values are used to determine an interpolating solution, and the overdetermined case in which there is an overabundance of data and functional values. The second section describes various types of functionals which have been identified to date. The third section connects some of the described functionals to methods which have appear elsewhere.

4.1 Eigenvalues

In this section, three different data situations are considered and approaches to using functionals to solve them are presented. The first situation is the theoretical one. In that situation, the data are assumed to be generated from the model equation and the data values are assumed perfect. Thus, it is called the "Perfect Data" case. Even in the case of perfect data, though, the numerical methods used to determine the coefficients, b_k, may introduce anomalies, so two different methods of using the data to determine the coefficients are given.

The second situation is the case in which there is "just enough" data to solve the problem. The developments in this case follow the method presented in Martin et al. [65]. The difficulty presented in this case is that, for some functionals, the matrices H_0 and H_1 are not full. Some means is needed to fill in the missing information to provide a complete solution to the problem. The results of this section do not have theoretical backing, however, the numerical results presented in Martin's article indicate promise.

The overdetermined case is presented third. Standard linear least squares ideas are used to develop methods for coping with excess data. The theoretical foundations for using linear least squares for this type of approach are shaky however, and numerical results presented in a later section urge caution.

4.1.1 Perfect Data

Let $\boldsymbol{x}$ be an eigenvector of G^{T}. Then, consider the equation

$$H_0 G^{\mathsf{T}} = H_1,$$

where H_0 and H_1 are two Hankel matrices related by a linear functional. Apply $\boldsymbol{x}$ on the right to obtain

$$\begin{aligned} H_0 G^{\mathsf{T}} \boldsymbol{x} &= H_1 \boldsymbol{x} \\ H_0 \lambda \boldsymbol{x} &= H_1 \boldsymbol{x} \end{aligned} \tag{4.1}$$

or

$$(H_0 \lambda - H_1)\, \boldsymbol{x} = 0. \tag{4.2}$$

Thus, there are two methods of determining the eigenvalues of the matrix G.

Method 1. *Solve*

$$G H_0 = H_1$$

for G and determine the eigenvalues.

Method 2. *Solve the generalized eigenvalue equation*

$$(H_0 \lambda - H_1) x = 0.$$

4.1.2 Interpolation

Consider the situation in which there are p data values, y_k, and p functional values, $\mathcal{F}(y_k) = \tilde{y}_k$. Then, to form the matrices H_0 and H_1, $p-1$ unknown data values and $p-1$ unknown functional values must be determined. The desired matrices are then

$$H_0 = \begin{bmatrix} y_0 & y_1 & \cdots & y_{p-2} & y_{p-1} \\ y_1 & y_2 & \cdots & y_{p-1} & * \\ y_2 & y_3 & \cdots & * & * \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{p-2} & y_{p-1} & \cdots & * & * \\ y_{p-1} & * & \cdots & * & * \end{bmatrix},$$

and

$$H_1 = \begin{bmatrix} \tilde{y}_0 & \tilde{y}_1 & \cdots & \tilde{y}_{p-2} & \tilde{y}_{p-1} \\ \tilde{y}_1 & \tilde{y}_2 & \cdots & \tilde{y}_{p-1} & * \\ \tilde{y}_2 & \tilde{y}_3 & \cdots & * & * \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{y}_{p-2} & \tilde{y}_{p-1} & \cdots & * & * \\ \tilde{y}_{p-1} & * & \cdots & * & * \end{bmatrix},$$

where the $*$ represent unknown values. Then the solutions, λ, to the eigenproblem

$$(H_0\lambda - H_1)\boldsymbol{x} = 0,$$

are the values $\mathcal{F}(\phi_k(0))$, as usual. The difficulty lies in determining the unknown parts of the matrices.

It may be possible to create an iterative method which will converge to the correct matrices H_i, $i = 0, 1$. Suppose for a moment that the values $y_p, y_{p+1}, \ldots, y_{2p-2}$ and

$\tilde{y}_p, \tilde{y}_{p+1}, \ldots, \tilde{y}_{2p-2}$ are assumed to be zero. Then let

$$H_0^{(0)} = \begin{bmatrix} y_0 & y_1 & \cdots & y_{p-2} & y_{p-1} \\ y_1 & y_2 & \cdots & y_{p-1} & 0 \\ y_2 & y_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{p-2} & y_{p-1} & \cdots & 0 & 0 \\ y_{p-1} & 0 & \cdots & 0 & 0 \end{bmatrix},$$

and

$$H_1^{(0)} = \begin{bmatrix} \tilde{y}_0 & \tilde{y}_1 & \cdots & \tilde{y}_{p-2} & \tilde{y}_{p-1} \\ \tilde{y}_1 & \tilde{y}_2 & \cdots & \tilde{y}_{p-1} & 0 \\ \tilde{y}_2 & \tilde{y}_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{y}_{p-2} & \tilde{y}_{p-1} & \cdots & 0 & 0 \\ \tilde{y}_{p-1} & 0 & \cdots & 0 & 0 \end{bmatrix},$$

and solve the eigenproblem

$$(H_0\lambda - H_1)\boldsymbol{x} = 0,$$

for the values $\lambda_k^{(0)}$. Solve the functional equation

$$\mathcal{F}(\phi_k(0)) = \lambda_k^{(0)}$$

for $\phi_k(1) = z_k$ and form the polynomial

$$l^{(0)}(z) = \prod_{k=1}^{p}(z - z_k) = z^p + \sum_{j=0}^{p-1} a_j z^j$$

and form the companion matrix

$$A^{(0)} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & & & \ddots & \\ -a_0 & -a_1 & -a_2 & \ldots & -a_{p-1} \end{bmatrix}.$$

Then, since both y_k and $\tilde{y}_k$ are sequences which obey the same difference equation[1], use the matrix A to fill in a first approximation to the missing values of H_0 and H_1. Call the new matrices $H_0^{(1)}$ and $H_1^{(1)}$. Repeat the process until convergence is achieved.

This method has been presented without proof. However, in experiments conducted by Martin et al. [65], the method converged in only 2 or 3 iterations for the osculatory (derivative) functional. An example of this method in use is shown in section 5.3.

4.1.3 Approximation

The formulation of the functional approach assumed that there was exactly enough data to create square matrices which could be inverted. As it happens, if there is an excess of data, and the matrix G represents a valid relationship between the data and the transformed data, then its eigenvalues still represent the functional values. Then the only concern is in determining the eigenvalues of G in a mathematically valid manner.

However, in the case where the system

$$GH_0^\mathsf{T} = H_1^\mathsf{T}$$

is overdetermined and inconsistent, some means must be found which offers a reasonable approximation of the eigenvalues of the problem. The most obvious method of determining the eigenvalues is to determine the G which minimizes the norm of

$$GH_0^\mathsf{T} - H_1^\mathsf{T},$$

and then to determine the eigenvalues of G. One way to do this is to use linear least squares to determine the rows of G. The use of the normal equations gives

$$H_0^\mathsf{T} H_0 G^\mathsf{T} = H_0^\mathsf{T} H_1.$$

[1]The linear functionals pass across the difference equations.

The solution, G, is unique if the matrix $H_0^\top H_0$ is nonsingular.

The numerical results reported in a later section, page 37, suggest that this formulation of the problem, for any method of solution, is inconsistent. A much better formulation of the problem is to use the generalized eigenvalue problem.

Method 3. *Let*

$$H_0 = \begin{bmatrix} y_0 & y_1 & y_2 & \cdots & y_{p-1} \\ y_1 & y_2 & y_3 & \cdots & y_p \\ \vdots & & & & \\ y_{n-p-r+1} & y_{n-p-r+2} & y_{n-p-r+3} & \cdots & y_{n-p-r} \end{bmatrix}$$

$$H_1 = \begin{bmatrix} \tilde{y}_0 & \tilde{y}_1 & \tilde{y}_2 & \cdots & \tilde{y}_{p-1} \\ \tilde{y}_1 & \tilde{y}_2 & \tilde{y}_3 & \cdots & \tilde{y}_p \\ \vdots & & & & \\ \tilde{y}_{n-p-r+1} & \tilde{y}_{n-p-r+2} & \tilde{y}_{n-p-r+3} & \cdots & \tilde{y}_{n-p-r} \end{bmatrix}.$$

Then solve

$$(H_0^\top H_0 \lambda - H_0^\top H_1)\boldsymbol{x} = 0$$

as a general eigenvalue problem.

4.2 Functionals

In this section, three different types of functionals are presented. The shift functionals are seen to be variations on the shift operator. The discussion shows that there are many different ways to use shifts to obtain solvable linear systems. The derivative operator is also a linear functional and will be seen to offer a simple way of determining the coefficients. Finally, the anti-derivative or integral operator is another linear functional which offers some simple solution methods.

4.2.1 Shift Functionals

Consider the shift operator $qy(t) = y(t+1)$. Then, it is a small step to create a linear functional from the shift functional by creating a polynomial of shifts,

$$\mathcal{F}(y) = (c_0 + c_1 q + c_2 q^2 + \cdots + c_r q^r) y.$$

Let

$$\tilde{y}_n = \mathcal{F}(y_n) = (c_0 + c_1 q + c_2 q^2 + \cdots + c_r q^r) y_n.$$

Form H_0 from the values y_n, and H_1 from the values $\tilde{y}_n$. Then the solutions, $\lambda_1, \lambda_2, \ldots, \lambda_p$, to

$$(H_0 \lambda - H_1)\boldsymbol{x} = 0 \tag{4.3}$$

are such that

$$\begin{aligned} \lambda_k &= c_0 + c_1 e^{b_k} + c_2 e^{b_k 2} + \cdots + c_r e^{b_k r} \\ \lambda_k &= c_0 + c_1 z_k + c_2 z_k^2 + \cdots + c_r z_k^r \\ 0 &= c_0 - \lambda_k + c_1 z_k + c_2 z_k^2 + \cdots + c_r z_k^r, \end{aligned} \tag{4.4}$$

where b_k is one of the true coefficients. The polynomial (4.4) has r roots, of which there is only a guarantee that at least one is the desired root z_k. Thus there is some ambiguity in choosing the coefficient b_k based on the polynomial (4.4). However, there is some hope. Under the assumption that the coefficients b_k are negative real, then the desired root, z_k, must be in the interval $(0, 1)$. Moreover, the following lemma and corollary show that if λ obeys certain easily tested criteria, and the polynomial is of a certain form, then it is easy to see that there is exactly one root of (4.4) in $(0, 1)$.

Lemma 11. *If $0 < \alpha < r$, the polynomial equation*

$$\alpha = z + z^2 + \cdots + z^r$$

has exactly one solution in $(0, 1)$.

Proof. Let

$$p(z) = z + z^2 + \cdots + z^r$$

then $p(0) = 0$, $p(1) = r$ and $z \geq 0 \Rightarrow p'(z) > 0$.
Therefore, $p(z) : (0,1) \rightarrow (0,r)$ is one-to-one and onto. ■

Corollary 12. *If* $1 < \alpha < r + 1$,

$$\alpha = 1 + z + z^2 + \cdots + z^r$$

has exactly one solution in $(0,1)$.

Lemma 11 can be extended to polynomials with all positive coefficients if desired. This alleviates some of the difficulties associated with the use of shift polynomials for determining coefficients, but requires certain a priori assumptions to be made about the nature of the coefficients.

There are certain major advantages of using the shift functionals. Consider a system with coefficients $b_1 = -0.0202$ and $b_2 = -0.0101$ which correspond to $0.98 = e^{-0.0202}$ and $0.99 = e^{-0.0101}$ respectively. Then, if the coefficients are to be determined from data which is generated by the function

$$f(t) = c_1 e^{-0.0202t} + c_2 e^{-0.0101t},$$

the shift functional $\mathcal{F} = 1 + q + q^2 + \cdots + q^6$ provides a characteristic equation with roots 6.5937 and 6.7935. The solution of this equation is more numerically stable than the equation with roots 0.98 and 0.99. Since the objective is to solve the generalized eigenvalue problem

$$(H_0\lambda - H_1)\boldsymbol{x} = 0$$

for the roots, then the use of the functional to add separation to the roots is beneficial. The tradeoff is that the polynomial

$$0 = c_0 - \lambda_k + c_1 z_k + c_2 z_k^2 + \cdots + c_r z_k^r$$

is hard to solve. If this is a large consideration and enough data is available, then instead of polynomials of shift, the user can just use a shift of high degree. To see an example of one of these shift functionals in use see page 26. The functional approach allows for easy justification of the resulting method and easy calculation of the inverse function. For comparison with the above, the functional transformation $\mathcal{F}(y) = q^{12}y$ yields corresponding eigenvalues of .7847 and .8864.

4.2.2 Differentiation Functionals

Let

$$\mathcal{F}(y) = \frac{d}{dt}y.$$

Then, the eigenvalues of the matrix G are

$$\mathcal{F}(\phi_k(0)) = \frac{d}{dt}e^{b_k t}\Big|_{t=0} = b_k.$$

Then, provided the derivatives of the measured data are available, this formulation presents a very simple method of determining the decay factors. The theoretical groundwork laid in developing the functional technique, significantly reduces the algebraic overhead in describing the use of derivatives. Compare this to the development in [65] of osculatory interpolation.

Often the determination of derivatives from data is difficult. Noise in the data can significantly affect the computed derivatives, it is easier to compute the integral of the data. The functional technique can be used to justify this approach also.

4.2.3 Integration Functionals

Let

$$\mathcal{F}(y) = \int_0^t y d\tau - \int_0^\infty y d\tau = -\int_t^\infty y d\tau$$

The second integral is in place to allow the integrated process to be shifted to satisfy the stationary assumption of the AR process. Assume that Re $b_k < 0$. Then the

eigenvalues of G are

$$\mathcal{F}(\phi_k(0)) = -\int_t^\infty e^{b_k\tau}d\tau\Big|_{t=0} = \frac{1}{b_k}e^{b_k t}\Big|_{t=0} = \frac{1}{b_k}.$$

One possible advantage to using the integration functional is that integration is an inherently smoothing operation and there may be some beneficial effect on the noise content of the data by using numerical integration.

4.3 Comparison with Methods from Literature

The functional technique can also be viewed as a unifying theory for various methods that exist in the literature. Prony's method and the matrix pencil are the most obvious applications of the functional method. The osculatory method developed by Martin et al. is another method.

Let q be the standard shift of degree one. Then put

$$\mathcal{F}(y) = qy.$$

Let H_0 and H_1 be as usual. Solve the matrix equation

$$GH_0 = H_1. \tag{4.5}$$

Then, the matrix, G, is a companion matrix whose last row, $[-c_0, -c_1, \ldots, -c_p]$, is precisely the vector which results from Prony's method. The eigenvalues of G are the eigenvalues resulting from the matrix pencil and are the roots of the characteristic equation identified by Prony's method. Now, consider the problem as an eigenproblem. The eigenvalues of G are theoretically the same as those from

$$(H_0\lambda - H_1)\boldsymbol{x} = 0, \qquad x \neq 0.$$

The only time that a difference might occur in the solutions to the two problems is when the conditioning of the matrix H_0 is such that (4.5) is unstable [30, 106]. Therefore, the simple first order shift, $\mathcal{F}(y) = qy$, generates Prony's Method and the matrix pencil. Examples of these methods can be found in [2, 35, 63, 79].

Let $\mathcal{F}(y) = \frac{d}{dt}y$. Designate the derivative values by $\tilde{y}_n$. Then let H_0 be given as usual for data values $y_0, y_1, \ldots, y_{2p-2}$ and let H_1 be generated from the derivative values $\tilde{y}_0, \tilde{y}_1, \ldots, \tilde{y}_{2p-1}$. Then the solutions, λ, to the eigenproblem

$$(H_0\lambda - H_1)\boldsymbol{x} = 0, \qquad x \neq 0,$$

are precisely the coeffcients, b_k. This result is the same as that achieved by Martin et al. [65].

Thus, the standard AR formulation of the problem is the special case of the functional method resulting from the shift functional of order one. The new development of Martin et al. is the special case of the functional method resulting form the derivative functional.

4.4 Conclusion

The implementation of the functional method is simple. It only requires that a linear functional be chosen for which the inverse problem can be solved. Then, the data is transformed using the functional and the eigenvalue problem solved. Then the user solves the inverse linear functional problem for the coefficients of the system. The use of the derivatives in one case showed that if there is data available which is related to the measurements by some linear functional, that additional data can be used to aid the process of determining the coefficients.

This chapter showed some simple functionals which are easy to implement with measured data. In the case of the shift-polynomials, the inverse problem is not necessarily easy to solve, however, the division of difficulty between the eigenvalue problem and inverse functional problem may prove to be beneficial in certain instances.

The integration technique has promise in the area of reducing the sensitivity of the method to noise. The comparisons in the next chapter between methods based on different functionals shows this promise clearly.

CHAPTER 5
NUMERICAL RESULTS

The last chapter demonstrated some simple linear functionals which can be used as bases for solution methods. This chapter compares some of those functional based extensions of Prony's method to solution methods which are reported in literature. The results reported here show that the functional methods are not uniformly better than methods which already exist. However, some of the methods presented here show promise in certain situations. The lesson to be learned from these examples is that the functional approach offers the user the ability to try different methods until he is satisfied with the results. The addition of many new methods can only result in better analysis capability.

5.1 A Simulation

This section presents the results a simulation comparison of several AR methods versus several different methods based on linear functionals. First, consider the problem of identifying the nonlinear coefficients of the function

$$f(t) = e^{-t} + e^{-2t} + e^{-3t} + e^{-4t} + e^{-5t}$$

from data obtained by simulating the function at the evenly spaced times, $t_i = ih$ for $i \in \{0, 1, \ldots, 100\}$ and $h = 0.01$.

Several different methods are used to try to solve this problem. Pisarenko Harmonic Decomposition (PHD), standard least squares, and Osborne's Objective Reweighting Algorithm (ORA) are representative algorithms from the AR class of techniques. Several different methods based on the functional approach are used for the remainder of the comparison. For the Derivative functional, the actual function was used to generate derivative values with no added noise. For the integration functional used in the two examples "Integrate-1" and "Integrate-2", the eigenvalue problems were $H_0^\top H_0 \lambda - H_0^\top H_1$, and $H_1^\top H_1 \lambda - H_1^\top H_0$ respectively. The "Shift-n" functionals

are the simple shift of order n functionals. These provide easy inverse solution with no ambiguity in the solution. All of the functional methods were solved as eigenvalue problems with the QZ algorithm.

Table 5.1 shows the results of the simulation when the level of noise in the data is strictly a function of the numerical noise in the computer representation. For Table 5.2, measurement noise was simulated in the data at the level of $N(0, .00005^2)$. Certain data values represent a negative root to the AR equation. Some of these values are easy to spot and are designated with *. Note that for the Shift-18 functional, $17.5 \times 18 = 315$, which is π/h within rounding. The same is true for the Shift-21 functional, where $15 \times 21 = 315$.

5.2 Lanczos' Problem

Lanczos presented an example of using Prony's method [56]. In his example, he summed the data to create a smaller data set which could be solved directly using Prony's method. The data was generated by the function

$$f(t) = 0.0951e^{-t} + 0.8607e^{-3t} + 1.5576e^{-5t}$$

with equal length time intervals $h = 0.05$. Each data value was rounded to 2 decimal places to simulate measurement capabilities. The data are reproduced in Table 5.3 for reference.

Both Lanczos and Ruhe [89] report limited success determining the exponential coefficients. Table 5.4 reports the results obtained by Lanczos, Ruhe, and current experiments using Osborne's Objective-function Reweighting Algorithm(ORA) method, Linear least Squares Prony method, the Matrix Pencil, and several different functional methods. The factors, k, represent the model size used in the solution example. Note that in that table, several of the examples use oversize models. If the user is interested in only real decay factors, the use of an oversize model can increase accuracy in the real factors. This is most apparent in the results reported for $\mathcal{F}_1$, $k = 11$. Outside of the ORA method of Osborne, this functional returned the best results for this data

set. In the case of the linear least squares and matrix pencil methods, some of the returned coefficients were complex. These discarded answers are represented in the table by a dash.

The method developed by Osborne and reported in [42] as the ORA method is a simplification of the Gradient-function Reweighting Algorithm(GRA) which was not used for these comparisons. The ORA method appears to provide answers which are significantly better than those available from any other method. These results can be misleading however. ORA is a nonlinear method which requires a starting value for the autoregressive parameters. Global convergence is not guaranteed, so it is customary to start the algorithm with the true parameters. Alas, if the true parameters are not known, the algorithm may not converge to the correct values.

The result from the functional method $\mathcal{F}_1$ with model size k=11 is also good. For reference, the functionals used in Table 5.4 are listed here.

$$\mathcal{F}_1 = 1 + z.$$
$$\mathcal{F}_2 = 1 + z + z^2 + z^3 + z^4.$$

The functional $\mathcal{F}_1$ is invertible, but $\mathcal{F}_2$ is not. For any given solution, λ, of the eigenproblem, the polynomial equation

$$\lambda = 1 + z + z^2 + z^3 + z^4$$

has 4 solutions. The difficulty of the method is that it must distinguish among these solutions and report the one which is most likely. The heuristic used in this program is that the correct solution is the one which is closest to one on the complex plane. Imaginary parts are then discarded and duplicates merged. The zeros reported in the table represent failures to find any solution for at least one λ which was within 1 unit distance from 1 on the complex plane.

At then end of the table, results are reported for several different integration methods. All of the integration methods use the same type of integration. That is,

they add the values in sequence, generating a new data stream of cumulative sums. The entire sequence then has the last data value subtracted from it to simulate an integration from x to infinity (see page 33). This works well, as can be seen from the results. The different results stem from the method used to solve the eigenvalue problem. The problem is overdetermined so the projection method shown in section 4.1.3 is used. In the first two examples, Integrate-1 and Integrate-2, the eigenvalue problem is

$$H_1^\top H_0 G = H_1^\top H_1 \qquad \text{Integrate-1}$$

or

$$(H_1^\top H_0 \lambda - H_1^\top H_1)\boldsymbol{x} = 0 \qquad \boldsymbol{x} \neq 0 \qquad \text{Integrate-2.}$$

Notice that this differs from the usual projection in that the matrix from the modified data is used to define the projection space. The results from both solution methods are shown to be the same to within reported significance. The next two examples are made from the usual projection.

$$H_0^\top H_0 G = H_0^\top H_1 \qquad \text{Integrate-3}$$

and

$$(H_0^\top H_0 \lambda - H_0^\top H_1)\boldsymbol{x} = 0 \qquad \boldsymbol{x} \neq 0 \qquad \text{Integrate-4.}$$

In the case reported as Integrate-3, the equation is solved for G and the eigenvalues are found from G. In Integrate-4, the generalized eigenvalue problem is solved directly using the QZ algorithm described in [30]. Clearly, using H_1 to define the projection space greatly improves the conditioning of the problem. One possible explanation for this is that the process of summing the data reduces the influence of rounding errors. The matrix $H_1^\top H_1$ is thus better conditioned than $H_0^\top H_0$. However, the results for Integrate-3 and Integrate-4 clearly show that solving the eigenproblem

directly is much more stable than solving the matrix equation first. In the event that projection onto the space defined by H_1 is not justified (all the shift functionals), the eigenproblem should be solved directly to avoid numerical difficulties. The user may even want to experiment with the more expensive VZ algorithm reported in [106].

5.3 An Interpolation Example

Suppose data were collected from an experiment and due to some difficulty, the middle part of the data were deemed inadequate. For the sake of definiteness, suppose the data values y_3, y_4 of a total of 8 data values were missing. Then, the researcher has only $y_0, y_1, y_2, y_5, y_6, y_7$ from which to make a determination of the decay rate of the experiment. Since the data can be taken in pairs, y_k, y_{k+5}, the researcher chooses to use a shift functional of order 5 to determine the coefficients. in this case, the starting matrices look like

$$H_0 = \begin{bmatrix} y_0 & y_1 & y_2 \\ y_1 & y_2 & 0 \\ y_2 & 0 & 0 \end{bmatrix}$$

and

$$H_1 = \begin{bmatrix} y_5 & y_6 & y_7 \\ y_6 & y_7 & 0 \\ y_7 & 0 & 0 \end{bmatrix}.$$

To see how this example actually performs, test data were generated with

$$f(t) = e^{-t} + e^{-3t} + e^{-5t},$$

for the interval $[0, 1]$ at equal length time intervals $h = 0.05$. No noise was added. Table 5.5 contains the first ten data values to 4 decimal places. It also has the results of the interpolation after 6 iterations of the scheme. The data values used in the experiment were maintained in their full precision in Matlab on a PC. After three iterations the matrices are

$$H_0 = \begin{bmatrix} 3.0000 & 2.5907 & 2.2522 \\ 2.5907 & 2.2522 & 1.9717 \\ 2.2522 & 1.9709 & 1.7381 \end{bmatrix}$$

and

$$H_1 = \begin{bmatrix} 1.5377 & 1.3705 & 1.2284 \\ 1.3705 & 1.2284 & 1.1070 \\ 1.2284 & 1.1068 & 1.0026 \end{bmatrix}$$

and the coeficients, b_k, have been estimated as $\{-1.2609, -4.5082, -6.4569\}$. Note that after three iterations, the matrices are not true Hankel matrices and the proper values for the interpolated data cannot be determined. After three more iterations, the matrices are truly Hankel matrices to 4 decimal places and the estimated data stream is reproduced in the second column of Table 5.5. The values with * beside them represent the interpolated values. The estimated coefficients are now $\{-1.2638, -3.1055, -4.5049\}$. Thus the method appears to converge nicely and interpolates the missing data values to within 0.0003. Further iterations result in little to no improvement. The Matlab program which implements the iteration is listed in Figure 5.1.

This example shows that the functional method can be useful in interpolating data. Caution must be used; however, the same procedure was tried with the first ten values from Table 5.3. The iteration scheme never converged to Hankel matrices with 4 decimal places of accuracy (to be expected from data with only 2 decimal places of accuracy), although the returned coefficients were closer to the correct values than many of the examples in Table 5.4.

A note about implementation is in order. The algorithm in Figure 5.1 specifies that the columns in the matrices H_0 and H_1 be updated from the preceding columns in the previous iteration. In other words, if $\boldsymbol{r}$ is the $k-1$st column in $H_0^{(i)}$, then the kth column in $H_0^{(i+1)}$ should be

$$G^{(i)}\boldsymbol{r} \tag{5.1}$$

instead of

$$\left(G^{(i)}\right)^{k-1}\boldsymbol{y} \tag{5.2}$$

where $\boldsymbol{y}$ is the fully known first column of H_0. The update method (5.2) results in all of the coefficients converging to the same number rather than (in this case) three distinct ones. To gain some insight into why this is so, consider a step of the iteration using (5.2):

$$H_0^{(i+1)} = \begin{bmatrix} \boldsymbol{y} & G^{(i)}\boldsymbol{y} & \left(G^{(i)}\right)^2 \boldsymbol{y} \end{bmatrix} \tag{5.3}$$

and a step of the iteration using (5.1):

$$H_0^{(i+1)} = \begin{bmatrix} \boldsymbol{y} & G^{(i)}\boldsymbol{y} & G^{(i)}G^{(i-1)}\boldsymbol{y} \end{bmatrix} \tag{5.4}$$

It appears that slowing down the convergence rate by using a previous value for G keeps the algorithm from overconverging. It should also be noted that with the matrix G in companion form, the known parts of H_0 and H_1 never change.

Table 5.1: Results with no added noise

Method	Results				
PHD	$56 + 230i$	$56 - 230i$	-1.09	-2.77	-4.76
ORA	$-7.75 + 288i$	$90.9 - 130i$	-1.08	-2.74	-4.73
Derivative	-5.0000	-4.0003	-3.0005	-2.0001	-1.0000
Shift-1	-5.0016	-4.0161	-3.0223	-2.0060	-1.0002
Stand. LLS	-5.0016	-4.0163	-3.0226	-2.0060	-1.0002
Shift-3	-4.9995	-3.9952	-2.9929	-1.9981	$-.9999$

Table 5.2: Results with noise level $\varepsilon \sim N(0, 0.00005^2)$.

Method	Results				
PHD	-1.56	-5.31	-6.50	$-4.5 + 167i$	$-4.5 - 167i$
ORA	-1.08	-4.71	-2.71	$-5.98 - 308i$	$111 - 10i$
Derivative	-1.43	-4.21	-0.28	0.0000	0.0000
Integration-1	-1.09	-5.76	-2.25	$-173 - 68i$	$-173 + 68i$
Integration-2	-1.59	-4.26	0.23	$-142 + 52i$	$-142 - 52i$
Shift-1	-1.37	-4.18	$-60 + 314i^*$	$-47 + 184i$	$-47 - 184i$
Stand. LLS	-1.37	-4.18	$-60 + 314i^*$	$-47 - 184i$	$-47 + 184i$
Shift-3	-1.37	-4.17	-102	$-36 - 56i$	$-36 + 56i$
shift-6	-1.36	-4.15	-40.8	$-23.0 + 26i$	$-23.0 - 26i$
shift-9	-1.36	-4.14	-19.9	$-19.0 - 34.9i$	$-44.1 - 34.9i$
shift-12	-1.35	-4.13	-32.6	$-22.0 + 9.07i$	$-22.0 - 9.07i$
shift-15	-1.35	-4.11	-24.6	$-14.7 - 3.17i$	$-14.7 + 3.17i$
shift-18	-1.35	-4.11	$-15.5 + 17.5i^*$	$-8.89 + 16.4i$	$-8.89 - 16.4i$
shift-21	-1.34	-4.09	$-7.58 + 15.0i^*$	$-6.81 + 6.03i$	$-6.81 - 6.03i$
shift-24	-1.34	-4.06	-5.44	$-3.66 + 2.94i$	$-3.66 - 2.94i$

Table 5.3: Lanczos' data.

Data Values					
2.51	1.12	0.53	0.27	0.15	0.09
2.04	0.93	0.45	0.23	0.13	0.08
1.67	0.77	0.38	0.20	0.11	0.07
1.37	0.64	0.32	0.17	0.10	0.06

Table 5.4: Results from various methods.

Method	Result		
Ruhe (k=2)	-1.75	-4.55	
Lanczos (k=2)	-1.58	-4.45	
LLS (k=2)	-3.5	-8.5	
LLS (k=3)	-2.71	-5.1	—
LLS (k=4)	-1.97	-4.6	—
Matrix Pencil (k=2)	-3.53	-8.5	
Matrix Pencil (k=3)	-2.71	-5.08	—
ORA (k=2)	-1.81	-4.57	
ORA (k=3)	-0.97	-3.3	-5.1
PHD (k=3)	-0.26	-3.92	-16.9
$\mathcal{F}_1$ (k=3)	-2.71	-5.08	0
$\mathcal{F}_1$ (k=11)	-0.89	-3.35	-5.29
$\mathcal{F}_2$ (k=2)	-2.8	-5.3	
$\mathcal{F}_2$ (k=3)	-2.0	-4.7	0
Integrate-1 (k=3)	-.72	-3.2	-6.5
Integrate-2 (k=3)	-.72	-3.2	-6.5
Integrate-3 (k=3)	-.3.3	-16.2	-65.5
Integrate-4 (k=3)	-.3	-1.2	-6.0

Table 5.5: Comparison of true data with interpolated data

Actual	Interpolated
3.0000	3.0000
2.5907	2.5907
2.2522	2.2522
1.9707	1.9708*
1.7354	1.7357*
1.5377	1.5377
1.3705	1.3705
1.2284	1.2284
1.1068	1.1068*
1.0023	1.0020*

```
% assumes the polynomial p and the matrices H0, H1 already exist
G=[0 1 0
   0 0 1
   -p(4:-1:2)];
% it is very important to update H0, H1 from H0, H1
%  rather than from y, y1 and G
H0=[y G*H0(:,1) G*H0(:,2)]
H1=[y1 G*H1(:,1) G*H1(:,2)]
e=real(eig(H1,H0));
e1=log(e)/5;     % assumes y1 is shifted 5 from y
p=poly(exp(e1));
e2=e1/.05        % only necessary when reporting coefficients
```

Figure 5.1: Matlab Program for interpolation

CHAPTER 6
CONCLUSIONS

The functional technique described in this dissertation represents a great leap forward for the task of solving Prony's Problem. Before this technique was developed, the available methods depended on either nonlinear techniques or the autoregressive formulation. The AR formulation has been the method of choice because the problem is linear and responds to many different techniques. With the introduction of linear functionals, the arena has become much more vast.

This dissertation showed the theoretical basis for the functional techniques. It also showed how simple many problems become when the functional techniques are used. The justification for many different approaches to the problem becomes easier and the different transformations have been unified with one theory.

However, this dissertation could only scratch the surface of the new functional technique. There are still many questions to be answered. The iteration technique introduced in section 4.1.2 needs justification. The limitations on its convergence need to be identified. Since those convergence properties depend on the functional which is used, it is impossible to plumb those depths in this dissertation. Additionally, the properties of the overdetermined case need to be investigated. The method presented there is apparently irreconcilable with the covariance methods presented by Pisarenko, Osborne and others [83, 73]. Hopefully further research in this area will result using the functionals to create an orthogonalization technique for these problems. The existence of such a technique would result in vastly improved computational performance in these problems.

BIBLIOGRAPHY

[1] Akhiezer, N. L. *The classical moment problem and some related questions in analysis.* New York: Hafner, 1965.

[2] Ammar, G., W. Dayawansa, and C. Martin. "Exponential interpolation: theory and numerical algorithms." *Appl. Math. Comp.* 41 (1991): 189-232.

[3] Åström, K. J., and B. Wittenmark. *Adaptive control.* Reading, MA: Addison-Wesley, 1995.

[4] Atkinson, K. E. *An introduction to numerical analysis.* New York: Wiley, 1989.

[5] Baker, G. *Essentials of Padé approximants.* New York: Academic, 1975.

[6] Bates, D. M. and D. G. Watts. *Nonlinear regression analysis and its applications.* New York: John Wiley, 1988.

[7] Brockett, R. "The geometry of the partial realization problem." *Proceedings of the 1978 IEEE Conference on Decision and Control*, 1048-1052.

[8] Bunch, J. R. and L. Kaufmann. "Some stable methods for calculating inertia and solving symmetric linear systems." *Math. Comp.* 31(1977): 162-179.

[9] Barham, R. H. and W. Drane. "An algorithm for least estimation of nonlinear parameters when some of the parameters are linear." *Technometrics* 14 (1972): 757-766.

[10] Bellman, R. *Introduction to matrix analysis*, 2d ed. New York: McGraw-Hill, 1970.

[11] Benson, M. "Parameter fitting in dynamic models." *Ecol. Mod.* 6(1979): 97-115.

[12] Bradley, M. "Prony the bridge-builder: the life and times of Gaspard deProny, educator and scientist." *Centaurus* 37 (1994): 230-268.

[13] Bresler, Y. and A. Macovski. "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise." *IEEE Trans. Acoust. Speech Signal Process.* 34 (1986): 1081-1089.

[14] Box, G. E. P. , G. M. Jenkins, and G. C. Reinsel. *Time series analysis.* Englewood Cliffs, NJ: Prentice Hall, 1994.

[15] Caines, P. E. *Linear stochastic systems.* New York: John Wiley, 1988.

[16] Cantor, D. G., and J. W. Evans. "On approximation by positive sums of powers." *SIAM J. Appl. Math.* 18 (1970), pp. 380-388.

[17] Chebyshev, P. L. "Sur l'interpolation par la méthod des moindres carrés." *Mém. Acad. Impér. Sci.* (St. Péterbourg) 7 (1859): 1-24.

[18] Claeyssen, J. C. R., and L. A. Dos Santos Leal. "Diagonalization and spectral decomposition by factor block circulant matrices." *Linear Algebra Appl.* 99(1988): 41-44.

[19] Davis, P. J. *Circulant matrices.* New York: John Wiley, 1979.

[20] ——. *Interpolation and approximation*, New York: Blaisdell, 1963.

[21] Derin, H. "Estimating components of univariate Gaussian mixtures using Prony's method." *IEEE Trans. Pattern Analysis Machine Intelligence* 9 (1987): 142-148.

[22] Dongarra, J. J., J. R. Bunch, C. B. Moler, and G. W. Stewart. *LINPACK users' guide.* Philadelphia: Society for Industrial and Applied Mathematics, 1979.

[23] Evans, J. W., W. B. Gragg and R. J. LeVeque. "On least squares exponential sum approximation with positive coefficients." *Math. Comp.* 34 (1980): 203-211.

[24] Fletcher, R. *Practical methods of optimization vol 1: Unconstrained optimization.* Chichester, U. K.: Wiley, 1980.

[25] Folland, G. B. *Real analysis.* New York: Wiley, 1984.

[26] Gantmacher, F. R. *Matrix theory*, Vol.1, New York: Chelsea, 1977.

[27] Gautschi, W. "On generating orthogonal polynomials." *SIAM J. Sci. Statist. Comput.* 3 (1982): 289-317.

[28] Gilmour, A. E. "Circulant matrix methods for the numerical solution of partial differential equations by FFT convolutions." *Appl. Math. Modelling* 12 (1988): 44-50.

[29] Golub, G. H., and V. Pereyra. "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate." *SIAM J. Numer. Anal.* 10 (1973): 413-432.

[30] Golub, G. H., and C. F. Van Loan. *Matrix computations.* Baltimore: Johns Hopkins University Press, 1989.

[31] Gragg, W. "Matrix interpretations and applications of the continued fraction algorithm." *Rocky Mountain J. Math.* 4 (1974): 213-225.

[32] Gragg, W., and A. Lindquist. "On the partial realization problem." *Linear Algebra Appl.* 50 (1983): 277-319.

[33] Gragg, W., R. J. Leveque, and J. A. Trangenstein. "Numerically stable methods for updating regressions." *J. Amer Statist. Assoc.* 74, no. 365 (1979): 161-168.

[34] Hamming, R. W. *Numerical methods for scientists and engineers.* New York: Dover, 1986.

[35] Hildebrand, F. *Introduction to numerical analysis.* New York: McGraw-Hill, 1956.

[36] Hogg, R. V., and A. T. Craig, *Introduction to mathematical statistics.* New York: MacMillan, 1978.

[37] Householder, A. S. "On Prony's method of fitting exponential decay curves and multiple-hit survival curves." *Report ORNL-455.* Oak Ridge National Laboratory, 1950.

[38] ——. "Analyzing exponential decay curves." *Seminar on Scientific Computation: November, 1949.* New York: IBM (1950): 28-32.

[39] Hua, Y., and T. K. Sarkar. "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise." *IEEE Trans. Acoustics, Speech, and Signal Processing* 38, no. 5 (1990): 814-824.

[40] Huse, V. R. "Asymptotic properties for the sequential CUSUM procedure." *Statist. and Prob. Letters* 7 (1989): 73-80.

[41] Jennrich, R. I. "Asymptotic properties of non-linear least squares estimators." *Ann. Math. Statist.* 40 (1969): 633-643.

[42] Kahn, M., M. S. Mackisack, M. R. Osborne, and G. K. Smyth. "On the consistency of Prony's method and related algorithms." *J. Comput. Graph. Statist.* 1 (1992): 329-349.

[43] Kailath, T. *Linear systems.* Englewood Cliffs, NJ: Prentice-Hall, 1980.

[44] Kailath, T., A. Viera, and M. Morf. "Inverses of Toeplitz operators, innovations, and orthogonal polynomials." *SIAM Review* 20, no. 1 (1978): 106-119.

[45] Kalman, R. "On partial realizations, transfer functions and canonical forms." *Acta Polytech. Scand.* MA31 (1979): 9-32.

[46] Kammler, D. W. "Least squares approximation of completely monotonic functions by sums of exponentials." *SIAM J. Numer. Anal.* 16 (1979): 801-818.

[47] Kass, R. and G. K. Smyth. *Convergence of the scoring algorithm: a geometric interpretation.* Tech. Report. Department of Mathematics, University of Queensland, St. Lucia, Australia, 1989.

[48] Kaufmann, L. "A variable projection method for solving separable nonlinear least squares problems." *BIT* 15 (1975): 49-57.

[49] Kay, S. M. "Noise compensation for autoregressive spectral estimates." *IEEE Trans. Acoust. Speech Signal Proc.* 28, no. 3 (1980): 292-303.

[50] ——. "Robust detection by autoregressive spectrum analysis." *IEEE Trans. Acoust. Speech Signal Proc.* 30, no. 2 (1982): 256-269.

[51] ——. "Recursive maximum likelihood estimation of autoregressive processes." *IEEE Trans. Acoust. Speech Signal Proc.* 31, no. 1 (1983): 56-65.

[52] ——. *Modern spectral estimation: Theory and application.* Englewood Cliffs, NJ: Prentice-Hall, 1988.

[53] Kay, S. M. and S. L. Marple. "Spectrum analysis: a modern perspective." *Proceedings of the IEEE* 69, no. 11 (1981): 1380-1419.

[54] Kundu, D. "Estimating the parameters of undamped exponential signals." *Technometrics* 35 (1993): 215-218.

[55] Kung, S. Y. *Multivariable and multidimensional systems: analysis and design.* Ph.D. Thesis, Dept. of Electrical Engineering, Stanford Univ., June 1977.

[56] Lanczos, C. *Applied analysis.* Englewood Cliffs, NJ: Prentice-Hall, 1956.

[57] ——. "An iteration method for the solution of the eigenvalue problem of linear differential and integral operators." *J. Res. Nat. Bur. Standards* 45 (1950): 255-282.

[58] Lawton, W. H., and E. A. Sylvestre. "Elimination of linear parameters in nonlinear regression." *Technometrics* 13 (1971): 461-467.

[59] Lemmerling, P., B. DeMoor, and S. VanHuffel. "On the equivalence of constrained total least squares and structured total least squares." *IEEE Trans. Signal Proc.* 44, no. 11 (1996): 2908-2911.

[60] Ljung, G. M., and G. E. P. Box. "The likelihood function of stationary autoregressive-moving average models." *Biometrika* 66, no. 2 (1979): 265-270.

[61] Mackisack, M. S., M. R. Osborne, and G. K. Smyth."A modified Prony algorithm for estimating sinusoidal frequencies." *J. Statist. Comput. Simulation* 49 (1994): 111-124.

[62] Marple, S. L. "A new autoregressive spectrum analysis algorithm." *IEEE Trans. Acoust., Speech, Signal Proc.* 28, no. 4 (1980): 441-454.

[63] ——. *Digital spectral analysis with applications.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[64] Martin, C., and J. Smith. "Approximation, interpolation and sampling." *Contemp. Math.* 68 (1987): 227-252.

[65] Martin, C., W. Gragg, and Y. DeWoody. "Osculatory Interpolation for sums of exponentials." to be submitted.

[66] Mittleman, H. D., and H. Weber. "Multi-grid solution of bifurcation problems." *SIAM J. Sci. Statist. Comput.* 6 (1985): 49-60.

[67] Mulholland, R. J., J. R. Cruz, and J. Hill. "State-variable canonical forms for Prony's method." *Int. J. Systems Sci.* 17, no. 1 (1986): 55-64.

[68] Numerical Algorithms Group. *NAG FORTRAN library manual mark 10.* Oxford, U. K.: Numerical Algorithms Group, 1983.

[69] Nussbaum, R. D. "Circulant matrices and differential-delay equations." *J. Differential Equations* 60 (1985): 201-217.

[70] J. M. Ortega. *Matrix theory, a second course.* New York: Plenum, 1987.

[71] Ortega, J. M., and W. C. Rheinhold. *Iterative solution of nonlinear equations in several variables.* New York: Academic Press, 1970.

[72] Osborne, M. R. "A class of nonlinear regression problems." *Data Representation.* R. S. Anderssen and M. R. Osborne, eds. St Lucia: University of Queensland Press (1970): 94-101.

[73] ——. "Some special nonlinear least squares problems." *SIAM J. Numer. Anal.* 12 (1975): 571-592.

[74] ——. "Nonlinear least squares – the levenberg algorithm revisited." *J. Austral. Math. Soc.* Ser. B, 19 (1976), pp. 343-357.

[75] ——. "Inverse iteration, Newton's method, and nonlinear eigenvalue problems." *"The Contribution of Dr. J. H. Wilkinson to Numerical Analysis," Institute of Mathematics and Its Applications* 19 (1978): 21-53. London: Royal Society.

[76] Osborne, M. R., and G. K. Smyth. "An algorithm for exponential fitting revisited." *Essays in Time Series and Allied Processes: Papers in Honour of E. J. Hannan.* J. Gani and M. B. Priestly, eds. Sheffield: Applied Probability Trust (1986): 419-430.

[77] ——. "A modified Prony algorithm for fitting functions defined by difference equations." *SIAM J. Sci. Stat. Comput.* 12 (1991): 362-382.

[78] ——. "A modified Prony algorithm for exponential function fitting." *SIAM J. Sci. Comp.* 16, no. 1 (1995): 119-138.

[79] Ouibrahim, H. "Prony, Pisarenko, and the matrix pencil: A unified presentation." *IEEE Trans. Acoust. Speech Signal Process.* 37 (1989): 133-134.

[80] Ouibrahim, H., D. D. Weiner, and T. K. Sarkar. "A generalized approach to direction finding." *IEEE Trans. Acoust. Speech Signal Process.* 36, no. 4 (1988): 610-612.

[81] Pelikan, E. "Spectral analysis of ARMA processes by Prony's method." *Kybernetika* 20 (1984): 322-328.

[82] Pierre, D. A., D. J. Trudnowski, and J. F. Hauer. "Identifying linear reduced-order models for systems with arbitrary initial conditions using Prony signal analysis." *IEEE Trans. on Automatic Control* 37, no. 6 (1992): 831-835.

[83] Pisarenko, V. F. "The retrieval of harmonics from a covariance function." *Geophysical Journal of the Royal Astronomical Society* 33 (1973): 347-366.

[84] Pólya, G. "On the mean value theorem corresponding to a given linear homogeneous differential equation." *Bull. Amer. Math. Soc.* 24 (1922): 312-324.

[85] Prony, R. "Essai expérimental et analytique." *J. de L'École Polytechnique* 1 (1795): 24-76.

[86] Ratkowsky, D. A. *Nonlinear regression modelling: a unified practical approach.* New York: Marcel Dekker, 1953.

[87] Rissanen, J. "Recursive identification of linear sequences." *SIAM J. Control* 9: (1971): 420-430.

[88] Ross, G. J. S. *Nonlinear estimation.* New York: Springer, 1990.

[89] Ruhe, A. "Fitting empirical data by positive sums of exponentials." *SIAM J. Sci. Stat. Comput.* 1 (1980): 481-498.

[90] Rutishauser, H. *Der quotienten-differenzen-algorithmus.* Stuttgart: Birkhäuser, 1957.

[91] ——. *Lectures on numerical mathematics.* ed. M. Gutknecht, trans. W. Gautschi, Boston: Birhäuser, 1990.

[92] Sakai, H. "Statistical analysis of Pisarenko's method for sinusoidal frequency estimation." *IEEE Trans. Acoust. Speech Signal Process.* 32 (1984): 95-101.

[93] Seber, G. A. F., and C. J. Wild. *Nonlinear regression.* New York: Wiley, 1989.

[94] Siddiqui, M. M. "On the inversion of the sample covariance matrix in a stationary autoregressive process." *Annals of Mathematical Statistics* 28 (1958): 585-588.

[95] Smith, M. R., S. Cohn-Sfetcu, and H. A. Buckmaster. "Decomposition of multicomponent exponential decays by spectral analytic techniques." *Technometrics* 18 (1976): 467-482.

[96] Smyth, G. K. *Coupled and separable iterations in nonlinear estimation.* Ph.D. thesis, Canberra: Australian National University, 1985.

[97] ——. "Curvature and convergence." *Proc. Statistical Computing Section.* Alexandria: American Statistical Society (1987): 278-283.

[98] ——. "Partitioned algorithms for maximum likelihood and other nonlinear estimation." *Centre for Statistics, Research report 3.* St Lucia: University of Queensland, 1993.

[99] Stoer, J., and R. Bulirsch. *Introduction to numerical analysis.* New York: Springer, 1993.

[100] Stoica, P., and A. Nehorai. "Study of the statistical performance of the Pisarenko harmonic decomposition method." *Comm. Radar Signal Process.* 153 (1988): 161-168.

[101] Stout, W. F. "Some results on the complete and almost sure convergences of linear combinations of independent random variables and martingale differences." *Ann. Math. Statist.* 5 (1968): 1549-1562.

[102] ——. *Almost sure convergence.* New York: Academic Press, 1974.

[103] Thomas, G. B., and R. L. Finney. *Calculus and analytic geometry.* Reading, MA: Addison-Wesley, 1988.

[104] Turan, P. "On the theory of mechanical quadrature." *Acta Sci. Math.* 12 (1950): 30-37.

[105] VanHuffel, S., H. Park, and J. B. Rosen. "Formulation and solution of structured total least norm problems for parameter estimation." *IEEE Trans. Signal Proc.* 44, no. 10 (1996): 2464-2474.

[106] Van Loan, C. F. "A general matrix eigenvalue algorithm." *SIAM J. Numer. Anal.* 12, no. 6 (1975): 819-834.

[107] ——. "Generalizing the singular value decomposition." *SIAM J. Numer. Anal.* 13, no. 1 (1976): 76-83.

[108] Varah, J. M. "A spline least squares method for numerical parameter estimation in differential equations." *SIAM J. Sci. Stat. Comput.* 3 (1982): 28-46.

[109] ——. "On fitting exponentials by nonlinear least squares." *SIAM J. Sci. Stat. Comput.* 6 (1985): 30-44.

[110] Wallace, D., and C. Martin. "Observability and transcendental number theory." preprint, 1988.

[111] Walling, D. "Non-linear least squares curve fitting when some parameters are linear." *Texas J. Sci.* 20 (1968): 119-124.

[112] Whittaker, E., and G. Robinson. *The calculus of observations.* London: Blackie, 1926.

[113] Wilde, A. C. "Differential equations involving circulant matrices." *Rocky Mountain J. Math.* 13 (1983): 1-13.

[114] Wilkinson, J. H. *The algebraic eigenvalue problem.* Oxford: Clarendon Press, 1965.

[115] Wiscombe, W. J., and J. W. Evans. "Exponential-sum fitting of radiative transmission functions." *J. Comput. Physics* 24 (1977): 416-444.