7-1996

Parameter Estimation for Superimposed Weighted Exponentials

Edwards A. Ingham

PARAMETER ESTIMATION FOR SUPERIMPOSED
WEIGHTED EXPONENTIALS

DISSERTATION
Edward A. Ingham
Major, USAF

AFIT/DS/ENG/96-06

19960716 008

Approved for public release; distribution unlimited

AFIT/DS/ENG/96-06

PARAMETER ESTIMATION FOR SUPERIMPOSED WEIGHTED EXPONENTIALS

DISSERTATION

Presented to the Faculty of the Graduate School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Edward A. Ingham, B.S., M.S.

Major, USAF

July, 1996

PARAMETER ESTIMATION FOR SUPERIMPOSED WEIGHTED EXPONENTIALS

Edward A. Ingham, B.S., M.S.

Major, USAF

Approved:

_____   _____

Dr. Joseph J. Sacchini, Committee Chairman

_____   _____

Dr. Meir Pachter, Committee Member

_____   _____

Dr. Steven K. Rogers, Committee Member

_____   _____

Dr. Byron M. Welsh, Committee Member

_____   _____

Dr. Brahamanand Nagarsenker, Committee Member

_____   _____

Dr. George K. Haritos, Dean's Representative

_____

Dr. Robert A. Calico, Jr., Dean

Table of Contents

List of Figures

Acknowledgements

My list of acknowledgments is long but genuine. First thanks must go to Dr. Rogers for convincing the electrical engineering department that a masters student from the aeronautical engineering department could stand a chance in their Ph.D. program. Thanks also goes to my advisor, Dr. Sacchini, and the department head, Dr. D'Azzo, for accepting the endorsement. The participation and guidance of all three faculty members throughout my program has been greatly appreciated, especially the periodic "sanity" checks with Dr. Rogers.

After the first year of the program, I was given the opportunity to work under Dr. Pachter. Until then, I had reserved the descriptor "professional" for a mythical person to aspire to. Everything I have learned about good engineering, work ethic, and citizenship, I have witnessed in Dr. Pachter. I am thankful and indebted for his instruction.

I would also like to thank Dr. Welsh for his enthusiastic and thorough participation in my research, and Dr. Nagarsenker for grounding me in the universally applicable subject of statistics. Additionally, I must thank my cohorts: Dan Zahirniak for the many hours we spent figuring out how things really work, Matt Pepin for his enormous knowledge base and selflessness, and Jim Scofield for the critical assistance and instruction he provided in the application phase of my research.

Finally, and most importantly, I would like to thank my wife, Renee, for supporting me in what was suppose to be "her" assignment.

<div align="right">Edward A. Ingham</div>

AFIT/DS/ENG/96-06

Abstract

The approach of modeling measured signals as superimposed exponentials in white Gaussian noise is popular and effective. However, estimating the parameters of the assumed model is challenging, especially when the data record length is short, the signal strength is low, or the parameters are closely spaced.

In this dissertation, we first review the most effective parameter estimation scheme for the superimposed exponential model: maximum likelihood. We then provide a historical review of the linear prediction approach to parameter estimation for the same model. After identifying the improvements made to linear prediction and demonstrating their weaknesses, we introduce a completely tractable and statistically sound modification to linear prediction that we call iterative generalized least squares. It is shown, that our algorithm works to minimize the exact maximum likelihood cost function for the superimposed exponential problem and is therefore, equivalent to the previously developed maximum likelihood approach. However, our algorithm is indeed linear prediction, and thus revives a methodology previously categorized as inferior to maximum likelihood.

With our modification, the insight provided by linear prediction can be carried to actual applications. We demonstrate this by developing an effective algorithm for deep level transient spectroscopy analysis. The signal of deep level transient spectroscopy is not a straight forward superposition of exponentials. However, with our methodology, an estimator, based on the exact maximum likelihood cost function for the actual signal, is quickly derived. At the end of the dissertation, we verify that our estimator extends the current capabilities of deep level transient spectroscopy analysis.

x

PARAMETER ESTIMATION FOR SUPERIMPOSED WEIGHTED EXPONENTIALS

I. Introduction

1.1 Problem

This dissertation is focused on the superimposed weighted exponential parameter estimation problem. Many processes throughout applied science and engineering are modeled by this paradigm, *e.g.* (59, 23, 2, 32, 45). For example, in high frequency radar applications, a reflecting body of interest can be approximated as a collection of independent scattering centers. Under this assumption, the electromagnetic fields associated with each scattering center are modeled in the frequency domain as complex exponentials, and the net scattering from the body is considered to be the phasor sum of all the individual scattering centers (59). Estimates of the parameters of this model, when applied to actual radar signals, are used for radar imaging and analysis. Improved accuracy in the estimates directly improves resolution and increases capabilities.

Another use of the superimposed exponential model is seen in the analysis of signals associated with the free induction decay of nuclei in nuclear magnetic resonance (NMR) experiments. Both the concepts of NMR spectroscopy and magnetic resonance imaging (MRI) are predicated on the assumption that a volume of sample contains a finite number of different types of nuclei. Each type of nucleus is associated with the unique parameters of a complex exponential, and the total signal emanating from the volume is considered to be a finite sum of these complex expo-

nentials (23:342). Estimating the model parameters is essential to identifying the composition of samples in NMR spectroscopy or rendering an image in MRI.

Our research, in particular, is motivated by a class of applications interested in identifying the parameters of signals modeled by the sum of real decaying exponentials. Signals of this nature arise in many physics, chemistry, biophysics, and biochemistry experiments (2). Accurate parameter estimates can convey important information about molecular structure, reaction rates, and physical make up of the specimen.

Typically, exponential or multi-exponential signals are measured in experiments involving some type of perturbation. An external forcing parameter is pulsed or stepped on to a system, previously in an equilibrium state, and the resulting transient is monitored. For example, in the case of some fluorescence decay experiments, an exciting lamp flash is used to perturb proteins, protein conjugates, or nucleic acid conjugates for fluorimetric observation (32:1090). In relaxation kinetics, a sudden change in a chemical system's temperature, pressure, or electric field is used to invoke a concentration change in one or several species of the system (2:178). Spectrophotometric, fluorimetric, polarimetric, or conductometric detection methods are used to monitor the transient immediately after the perturbation.

A specific physics experiment that requires real exponential analysis is deep level transient spectroscopy (DLTS). DLTS is a capacitance transient thermal scanning technique used to characterize defects present in semiconductors (45:3023). Good characterization of the defects in semiconductors directly impacts semiconductor manufacturing, quality assurance, and technical advancement in general. For these reasons, accurate parameter estimation capability in the DLTS experiment is critical. In the latter chapters of this dissertation, we concentrate on the real ex-

ponential parameter estimation problem and the DLTS experiment. Ultimately, we identify improved parameter estimation algorithms created for DLTS.

Before developing the improved algorithms, we review the applicable contributions that have already been made for superimposed exponential parameter estimation. To keep the review as general as possible, the derivations and examples presented are application independent. In this vein, the model notation is now introduced.

1.2 Model

We assume $y[m]$ is an observed, complex signal plus noise element from a sequence of length M produced by N superimposed weighted exponentials with additive white Gaussian measurement noise,

$$(1) \qquad y[m] = s[m] + w[m], \qquad m = 0, 1, ..., M - 1$$

where

$$(2) \qquad s[m] = c_1\lambda_1^m + c_2\lambda_2^m + ... + c_N\lambda_N^m.$$

Unless specified otherwise, the model order (or number of modes), N, is assumed to be known and always less than the data record length, M. The complex exponentials, λ_n, are not equal to zero and are not repeated, $i.$ $e.$, $\lambda_i \neq \lambda_j$ for $i \neq j$, $i, j = 1, 2, ... N$. The complex amplitude coefficients, c_n, are also not equal to zero. Finally, the additive measurement noise contributions, $w[m]$, for each $y[m]$ are complex Gaussian random variables uncorrelated across all m with uncorrelated real and imaginary components, each of mean zero and variance $\sigma^2/2$.

Note, the Gaussian noise assumption is employed to scope the research. Although the assumption is often reasonable, we recognize the potential for it to be inappropriate. Many of the methodologies developed in this research may be extended for alternative measurement noise assumptions, but we defer that subject to future research.

Also note, when the purely real signal, $y[m]$, is considered, we assume the parameters λ_n and c_n are real. Additionally, the additive noise contributions, $w[m]$, are assumed to be real Gaussian random variables uncorrelated across all samples with mean zero and variance σ^2.

In vector notation, the observed signal and noise are written as

$$(3) \qquad\qquad \bar{y} = V\bar{c} + \bar{w}$$

where

$$
\begin{aligned}
\bar{y} &= \begin{bmatrix} y[0] & y[1] & \cdots & y[M-1] \end{bmatrix}^T \\
V &= \begin{bmatrix} \bar{v}(\lambda_1) & \bar{v}(\lambda_2) & \cdots & \bar{v}(\lambda_N) \end{bmatrix} \\
\bar{v}(\lambda_n) &= \begin{bmatrix} 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{M-1} \end{bmatrix}^T \\
\bar{c} &= \begin{bmatrix} c_1 & c_2 & \cdots & c_N \end{bmatrix}^T \\
\bar{w} &= \begin{bmatrix} w[0] & w[1] & \cdots & w[M-1] \end{bmatrix}^T .
\end{aligned}
$$

(4)

The $M \times N$ matrix V has Vandermonde structure. In Appendix A we show that with distinct exponentials, λ_n, the columns of V are linearly independent. Therefore, with $M > N$, the matrix V has full column rank, N.

4

1.3 Overview

In Chapters II and III, we identify two competing methodologies for estimating the parameters of superimposed weighted exponentials: maximum likelihood and linear prediction. Except at the end of Chapter III, the work developed in Chapters II and III is background information already presented by other researchers. In Chapter II, we review the maximum likelihood methodology and exalt it by identifying its relationship to the Cramer-Rao bound. In Chapter III, we review an alternative methodology: linear prediction. At the end of the Chapter, we introduce an extension to the linear prediction methodology that allows us to create an estimator equivalent to the estimator developed from maximum likelihood. This original finding allows us to simultaneously utilize the insights acquired during the development of linear prediction and, in general, the flexibility of the linear prediction method, while obtaining an estimator with performance capabilities approaching the Cramer-Rao bound. The equivalence of the two estimators is shown through a common cost function which is subject to minimization. In Chapter IV, a detailed review of techniques for minimizing the common cost function is given. In the review, we introduce new insight into the minimization problem by identifying equivalent minimization approaches and conditioning issues that were previously overlooked. In Chapter V, we transition to an analysis for the restricted case of superimposed *real* exponentials. The lessons learned in the previous chapters are applied, and the intricacies of the real exponential problem are considered. Our analysis for the superimposed real exponential problem is completely original work and can not be found in the literature.

With these subjects thoroughly investigated, our knowledge is applied to the DLTS application. The development and results are presented in Chapter VI. Be-

cause of our preparation, we are able to develop estimators for DLTS that significantly outperform the estimators currently in use. This contribution to DLTS has been accepted for publishing (31).

In Chapter VII, we conclude the dissertation by reviewing the primary and ancillary contributions of our research and addressing areas for immediate additional work.

II. Maximum Likelihood

2.1 Chapter Introduction

A parameter estimation method is coined maximum likelihood (ML) when it works to maximize the probability density function (PDF), $f(\bar{y}; \bar{\theta})$, associated with the observed realization, $\bar{y}$, and the assumed signal and noise model. The signal and noise model is parameterized by the vector $\bar{\theta}$. An ML estimate of $\bar{\theta}$ is rendered by finding the parameters that maximizes $f(\bar{y}; \bar{\theta})$ for a given realization of $\bar{y}$. The premise is that after observing $\bar{y}$, the maximizing parameters, $\bar{\theta}$, are the most likely parameters of the signal and noise model. When a PDF, $f(\bar{y}; \bar{\theta})$, is considered in this fashion, with a known $\bar{y}$ and unknown $\bar{\theta}$, it is called a likelihood function.

Estimating parameters with the ML approach is motivated by its relation to the Cramer-Rao bound (CRB). The Cramer-Rao bound identifies a lower bound for the variance of all unbiased estimators that is unique to the signal and noise model (35:27). Therefore, it provides a benchmark against which all unbiased estimation approaches can compete. An estimator derived from the ML approach is not guaranteed to attain the CRB, but it has been proven that if an unbiased estimator can attain the CRB, the ML approach will provide it (35:186–187). The theory of the CRB is presented at the end of this chapter and developed in Appendix C. First, we follow the methodology of Hogg and Craig (24:80–89), Kay (35:500–508), Ziskind and Wax (74:1554–1555), Lanczos (44:272–280), and Bresler and Macovski (4:1082–1084) to develop an ML estimator for the superimposed weighted exponential problem. All the work discussed in this chapter is review. It is presented to provide background and motivation for our research.

7

2.2 Maximum Likelihood Estimator

Temporarily let a signal plus noise element, $y[m]$, take the complex representation $y[m] = y_r + jy_i$. Also let the signal model take the complex representation $s[m] = s_r + js_i$, where y_r, y_i, s_r, and s_i are real. Because the measurement noise real and imaginary components are Gaussian distributed with mean zero and variance $\sigma^2/2$, and because $y[m]$ is a linear combination of the deterministic $s[m]$ and stochastic $w[m]$, y_r and y_i are distributed as

$$(5) \qquad f(y_r; s_r, \sigma^2) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2}}} \exp\left[-\frac{1}{2\left(\frac{\sigma^2}{2}\right)}(y_r - s_r)^2\right]$$

and

$$(6) \qquad f(y_i; s_i, \sigma^2) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2}}} \exp\left[-\frac{1}{2\left(\frac{\sigma^2}{2}\right)}(y_i - s_i)^2\right].$$

Because the measurement noise real and imaginary components are uncorrelated and Gaussian, they are stochastically independent. Therefore, the PDF $f(y_r, y_i; s_r, s_i, \sigma^2)$ is the product of $f(y_r; s_r, \sigma^2)$ and $f(y_i; s_i, \sigma^2)$ so that

$$(7) \qquad f(y_r, y_i; s_r, s_i, \sigma^2) = \frac{1}{\pi\sigma^2} \exp\left[-\frac{1}{\sigma^2}\left((y_r - s_r)^2 + (y_i - s_i)^2\right)\right].$$

Incorporating the definitions of $y[m]$ and $s[m]$, Equation 7 becomes

$$(8) \qquad f(y[m]; s[m], \sigma^2) = \frac{1}{\pi\sigma^2} \exp\left[-\frac{1}{\sigma^2}\|y[m] - s[m]\|_2^2\right]$$

where $\| \bullet \|_2$ represents the 2-norm.

Likewise, because each element, $y[m]$, is uncorrelated and Gaussian—therefore stochastically independent—the PDF of the vector $\bar{y}$ is the product of the PDFs for

each $y[m]$ so that

$$
\begin{aligned}
f(\bar{y}; V, \bar{c}, \sigma^2) &= \prod_{m=0}^{M-1} f(y[m]; s[m], \sigma^2) \\
&= \frac{1}{\pi^M \sigma^{2M}} \exp\left[-\frac{1}{\sigma^2} \sum_{m=0}^{M-1} \|y[m] - s[m]\|_2^2 \right].
\end{aligned}
$$
(9)

In equivalent vector notation,

$$
f(\bar{y}; V, \bar{c}, \sigma^2) = \frac{1}{\pi^M \sigma^{2M}} \exp\left[-\frac{1}{\sigma^2} (\bar{y} - V\bar{c})^H (\bar{y} - V\bar{c}) \right]
$$
(10)

where H is the conjugate (Hermitian) transpose.

In abbreviated notation,

$$
\bar{y} \sim CN(V\bar{c}, \sigma^2 \mathrm{I})
$$
(11)

implies $\bar{y}$ is distributed complex Gaussian (complex normal) with mean vector $V\bar{c}$ and covariance matrix $\sigma^2 \mathrm{I}$. The identity matrix, I, is of dimension $M \times M$. Notice that the common and uncorrelated variance, σ^2, affects a covariance matrix with constant diagonal only terms.

Returning to the ML task of finding the V, $\bar{c}$, and σ^2 that maximize $f(\bar{y}; V, \bar{c}, \sigma^2)$ for the observation $\bar{y}$, we find it easier to first take the natural logarithm of $f(\bar{y}; V, \bar{c}, \sigma^2)$ so that

$$
\begin{aligned}
L(V, \bar{c}, \sigma^2) &= \ln f(\bar{y}; V, \bar{c}, \sigma^2) \\
&= -M \ln \pi - M \ln \sigma^2 - \frac{1}{\sigma^2} (\bar{y} - V\bar{c})^H (\bar{y} - V\bar{c}).
\end{aligned}
$$
(12)

The function L is known as the log-likelihood function, and because the natural logarithm is a monotonic function, L will maximize at the same parameters as the PDF. Additionally, since $M \ln \pi$ is not a function of V, $\bar{c}$, or σ^2, it is equivalent to maximize

$$(13) \qquad L(V, \bar{c}, \sigma^2) = -M \ln \sigma^2 - \frac{1}{\sigma^2}(\bar{y} - V\bar{c})^H(\bar{y} - V\bar{c}).$$

We also know at the maximum,

$$(14) \qquad \frac{\partial L(V, \bar{c}, \sigma^2)}{\partial \sigma^2} = 0$$

where

$$(15) \qquad \frac{\partial L(V, \bar{c}, \sigma^2)}{\partial \sigma^2} = -\frac{M}{\sigma^2} + \frac{1}{(\sigma^2)^2}(\bar{y} - V\bar{c})^H(\bar{y} - V\bar{c}).$$

Therefore, at the maximum of L,

$$(16) \qquad \begin{aligned} &-\tfrac{M}{\sigma^2} + \tfrac{1}{(\sigma^2)^2}(\bar{y} - V\bar{c})^H(\bar{y} - V\bar{c}) = 0 \\ \Rightarrow \quad & \sigma^2 M = (\bar{y} - V\bar{c})^H(\bar{y} - V\bar{c}) \\ \Rightarrow \quad & \sigma^2 = \tfrac{1}{M}(\bar{y} - V\bar{c})^H(\bar{y} - V\bar{c}). \end{aligned}$$

If we substitute the expression for σ^2 back into L, we can now say we need to maximize

$$(17) \qquad L(V, \bar{c}) = -M \ln \left(\frac{1}{M}(\bar{y} - V\bar{c})^H(\bar{y} - V\bar{c}) \right) - M.$$

10

Again, because M is a constant and the natural logarithm is a monotonic function, maximizing L is equivalent to minimizing

$$(18) \qquad L(V, \bar{c}) = (\bar{y} - V\bar{c})^H (\bar{y} - V\bar{c}).$$

To simplify L further, we can remove the dependence on $\bar{c}$ by considering the current expression for L at its minimum, but before doing so, we need to define real and complex vector gradients. Our definitions are consistent with Kay (35:517-527) and Therrien (70:686-690). We define the gradient of a scalar L with respect to a real parameter vector $\bar{\theta}_r$ as

$$(19) \qquad \frac{\partial L}{\partial \bar{\theta}_r} = \begin{bmatrix} \frac{\partial L}{\partial \theta_r 1} \\ \frac{\partial L}{\partial \theta_r 2} \\ \vdots \\ \frac{\partial L}{\partial \theta_r N} \end{bmatrix}.$$

At a stationary point,

$$(20) \qquad \frac{\partial L}{\partial \bar{\theta}_r} = \bar{0}$$

where the notation $\bar{0}$ defines a vector of zeros. If we define the complex parameter vector

$$(21) \qquad \bar{\theta} = \bar{\theta}_r + j\bar{\theta}_i,$$

where $\bar{\theta}_r$ and $\bar{\theta}_i$ are both real vectors, we can also define the gradient of any scalar L with respect to a complex parameter vector $\bar{\theta}$ as

$$(22) \qquad \frac{\partial L}{\partial \bar{\theta}} = \frac{1}{2} \left(\frac{\partial L}{\partial \bar{\theta}_r} - j \frac{\partial L}{\partial \bar{\theta}_i} \right)$$

and

$$(23) \qquad \frac{\partial L}{\partial \bar{\theta}^*} = \frac{1}{2} \left(\frac{\partial L}{\partial \bar{\theta}_r} + j \frac{\partial L}{\partial \bar{\theta}_i} \right)$$

where * denotes the complex conjugate.

Notice that the log-likelihood function is a real scalar function. These gradient definitions allow us to calculate a stationary point for a real scalar function with complex parameters. Alternative definitions do not allow this because a real scalar function with complex parameters is not analytic (35:518). At a stationary point,

$$(24) \qquad \frac{\partial L}{\partial \bar{\theta}_r} = \frac{\partial L}{\partial \bar{\theta}_i} = \bar{0}.$$

Therefore,

$$(25) \qquad \frac{\partial L}{\partial \bar{\theta}} = \bar{0}$$

and

$$(26) \qquad \frac{\partial L}{\partial \bar{\theta}^*} = \bar{0}$$

are both valid gradient expression for determining stationary points.

Returning to the ML estimator development, we know

$$(27) \qquad \frac{\partial L(V, \bar{c})}{\partial \bar{c}^*} = \frac{\partial(\bar{y}^H \bar{y} - \bar{c}^H V^H \bar{y} - \bar{y}^H V \bar{c} + \bar{c}^H V^H V \bar{c})}{\partial \bar{c}^*}.$$

Under the gradient definition just presented,

$$
\begin{aligned}
\frac{\partial(\bar{y}^H \bar{y})}{\partial \bar{c}^*} &= \bar{0} \\
\frac{\partial(\bar{c}^H V^H \bar{y})}{\partial \bar{c}^*} &= \bar{0} \\
\frac{\partial(\bar{y}^H V \bar{c})}{\partial \bar{c}^*} &= V^H \bar{y} \\
(28) \qquad \frac{\partial(\bar{c}^H V^H V \bar{c})}{\partial \bar{c}^*} &= V^H V \bar{c}.
\end{aligned}
$$

This implies

$$(29) \qquad \frac{\partial L(V, \bar{c})}{\partial \bar{c}^*} = -V^H \bar{y} + V^H V \bar{c}.$$

The last three gradients of Equation 28 are derived in Appendix B. Therefore, at the minimum of L,

$$
\begin{aligned}
(30) \qquad & V^H V \bar{c} = V^H \bar{y} \\
\Rightarrow \quad & \bar{c} = (V^H V)^{-1} V^H \bar{y}.
\end{aligned}
$$

Recall that V is of full rank N with $M > N$. Therefore, $V^H V$ is also of full rank N, and $(V^H V)^{-1}$ exists.

13

This expression for $\bar{c}$ can be substituted into L so that minimizing L is equivalent to minimizing

$$
\begin{aligned}
L(V) &= (\bar{y} - V(V^H V)^{-1} V^H \bar{y})^H (\bar{y} - V(V^H V)^{-1} V^H \bar{y}) \\
&= \left((I - V(V^H V)^{-1} V^H) \bar{y}\right)^H \left((I - V(V^H V)^{-1} V^H) \bar{y}\right) \\
&= \bar{y}^H (I - V(V^H V)^{-1} V^H)^H (I - V(V^H V)^{-1} V^H) \bar{y} \\
&= \bar{y}^H (I - V(V^H V)^{-1} V^H) \bar{y}.
\end{aligned}
\tag{31}
$$

Equation 31 is the exact maximum likelihood cost function for the superimposed exponential problem reduced to just the N λ_n parameters in V. The expression

$$
(I - V(V^H V)^{-1} V^H) \bar{y}
\tag{32}
$$

is recognized as the projection of the observation vector $\bar{y}$ onto the orthogonal compliment of the range space of the columns of V. Under the maximum likelihood approach, we desire the exponential parameters, λ_n, for the columns of V that minimize this projection. If the minimizing λ_n can be accurately estimated, we can then estimate the amplitude coefficients, $\bar{c}$, by solving

$$
\bar{c} = (V^H V)^{-1} V^H \bar{y}.
\tag{33}
$$

With V and $\bar{c}$ estimated, we can even estimate the measurement noise variance, σ^2, with the earlier expression

$$
\sigma^2 = \frac{1}{M} (\bar{y} - V\bar{c})^H (\bar{y} - V\bar{c}).
\tag{34}
$$

14

The range space of the columns of V is referred to as the signal subspace, and its orthogonal compliment is referred to as the annihilator subspace. The maximum likelihood solution seeks to minimize the projection of the observation vector onto the annihilator subspace. To accomplish this by finding the λ_n that minimize the cost function L is a highly nonlinear task. With some astute insight, the process can be simplified by building an alternative projection matrix specifically for the annihilator subspace. What follows is the derivation of such a projection matrix.

We know from differential equation theory, the noiseless signal model

$$(35) \qquad s[m] = c_1 \lambda_1^m + c_2 \lambda_2^m + \ldots + c_N \lambda_N^m$$

is a solution to the homogeneous, constant coefficient, linear difference or linear prediction (LP) equation

$$(36) \qquad b_0 s[m] + b_1 s[m-1] + \ldots + b_N s[m-N] = 0.$$

To see this, substitute the solution into the LP equation so that

$$
\begin{aligned}
& b_0 \left[c_1 \lambda_1^m + c_2 \lambda_2^m + \ldots + c_N \lambda_N^m \right] + \\
& b_1 \left[c_1 \lambda_1^{m-1} + c_2 \lambda_2^{m-1} + \ldots + c_N \lambda_N^{m-1} \right] + \ldots + \\
& b_N \left[c_1 \lambda_1^{m-N} + c_2 \lambda_2^{m-N} + \ldots + c_N \lambda_N^{m-N} \right] = 0 \\
\Rightarrow \ & \left[b_0 c_1 \lambda_1^m + b_1 c_1 \lambda_1^{m-1} + \ldots + b_N c_1 \lambda_1^{m-N} \right] + \\
& \left[b_0 c_2 \lambda_2^m + b_1 c_2 \lambda_2^{m-1} + \ldots + b_N c_2 \lambda_2^{m-N} \right] + \ldots + \\
& \left[b_0 c_N \lambda_N^m + b_1 c_N \lambda_N^{m-1} + \ldots + b_N c_N \lambda_N^{m-N} \right] = 0 \\
\Rightarrow \ & c_1 \lambda_1^{m-N} \left[b_0 \lambda_1^N + b_1 \lambda_1^{N-1} + \ldots + b_N \right] + \\
& c_2 \lambda_2^{m-N} \left[b_0 \lambda_2^N + b_1 \lambda_2^{N-1} + \ldots + b_N \right] + \ldots + \\
& c_N \lambda_N^{m-N} \left[b_0 \lambda_N^N + b_1 \lambda_N^{N-1} + \ldots + b_N \right] = 0.
\end{aligned}
$$

(37)

Because c_n and λ_n are non zero, the LP coefficients, $b_0 \ldots b_N$, must be such that

$$(38) \qquad \begin{bmatrix} b_N & b_{N-1} & \cdots & b_0 \end{bmatrix} \begin{bmatrix} 1 \\ \lambda_n \\ \lambda_n^2 \\ \vdots \\ \lambda_n^N \end{bmatrix} = 0$$

for all n. Also, the roots z of the characteristic polynomial formed from the LP coefficients

$$(39) \qquad b_0 z^N + b_1 z^{N-1} + \ldots + b_N$$

are equal to the exponentials, λ_n.

If we propose the $(M - N) \times M$ Toeplitz matrix

$$(40) \qquad B = \begin{bmatrix} b_N & b_{N-1} & \cdots & b_0 & 0 & \cdots & 0 \\ 0 & b_N & b_{N-1} & \cdots & b_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & b_N & b_{N-1} & \cdots & b_0 \end{bmatrix},$$

we can see

$$(41) \qquad BV = 0.$$

Thus, every row of B is orthogonal to every column of the Vandermonde matrix V described earlier. Also, because each row of B is linearly independent, the rank of B is $M - N$. Therefore, the rows of B form a basis for the orthogonal compliment

16

of the range space of the columns of V, and through the projection matrix, we know

$$(42) \qquad B^H(BB^H)^{-1}B = I - V(V^HV)^{-1}V^H.$$

When this information is brought back to the maximum likelihood derivation, minimizing L is equivalent to minimizing the alternative maximum likelihood cost function

$$
\begin{aligned}
L(\bar{b}) &= \bar{y}^H B^H (BB^H)^{-1} B \bar{y} \\
&= \bar{b}^H Y^H (BB^H)^{-1} Y \bar{b}
\end{aligned}
$$

(43)

where

$$
\begin{aligned}
Y &= \begin{bmatrix} \bar{y}_0 & \bar{y}_1 & \cdots & \bar{y}_N \end{bmatrix} \\
\bar{y}_n &= \begin{bmatrix} y[N-n] & y[N+1-n] & \cdots & y[M-1-n] \end{bmatrix}^T \\
\bar{b} &= \begin{bmatrix} b_0 & b_1 & \cdots & b_N \end{bmatrix}^T .
\end{aligned}
$$

(44)

The data matrix Y is $(M-N) \times (N+1)$ with Toeplitz structure. If we let

$$(45) \qquad R(\bar{b}) = BB^H,$$

we can rewrite L as

$$(46) \qquad L(\bar{b}) = \bar{b}^H Y^H R(\bar{b})^{-1} Y \bar{b}.$$

Minimizing the alternative L is still a nonlinear task, but if we use a previous estimate of $\bar{b}$, labeled $\bar{b}_{i-1}$, to create a fixed $R(\bar{b}_{i-1})$ we attain the quadratic expression in $\bar{b}_i$

$$(47) \qquad \bar{b}_i^H Y^H R(\bar{b}_{i-1})^{-1} Y \bar{b}_i$$

where $\bar{b}_i$ identifies the current unknown. Techniques for finding the $\bar{b}_i$ that minimizes a quadratic are well known and are examined further in Chapters III and IV. An iterative approach that utilizes this quadratic expression naturally follows and is known as the iterative quadratic maximum likelihood (IQML) algorithm. The algorithm is most concisely presented in a paper by Bresler and Macovski (4), but its origins and alternative deliveries can be found in papers by Evans and Fischl (15) and Kumaresan, Scharf, and Shaw (39).

It is important to emphasize the IQML algorithm does not directly minimize the maximum likelihood cost function but rather a quadratic approximation of it. Nevertheless, the IQML algorithm is the most popular and accurate of all parameter estimation schemes for the superimposed exponential problem.

2.3 Cramer-Rao Bound

Confidence in all ML based estimators is gained when the relationship between maximum likelihood and the Cramer-Rao bound is understood. In this section, key properties of the theory of the CRB are presented to show the relationship. In Appendix C the theory necessary to justify the properties is developed. For both this section and the appendix, the books of Scharf (61:221–227), Kay (35:30–45), and Papoulis (54:260–265) were the primary references.

Ideally, we desire a function of the data, $g(\bar{y})$, that provides the best estimate of the parameters $\bar{\theta}$ where, in our case, $\bar{\theta}$ is a vector representation of all the parameters in V, $\bar{c}$ and σ^2. Recall the log-likelihood function, initially defined in Equation 12, where

$$(48) \qquad\qquad L(\bar{\theta}) = \ln f(\bar{y}; \bar{\theta}).$$

In deriving the alternative maximum likelihood cost function

$$(49) \qquad\qquad L(\bar{b}) = \bar{b}^H Y^H R(\bar{b})^{-1} Y \bar{b},$$

we have shown that finding the $\bar{b}$ that minimizes $L(\bar{b})$ is equivalent to finding the $\bar{\theta}$ that maximizes $L(\bar{\theta})$. Therefore, the IQML algorithm serves as a function of the data, $g(\bar{y})$, that estimates θ for maximizing $L(\bar{\theta})$.

In developing the IQML algorithm, and all other ML based estimators, we exploit the requirement that at the maximum,

$$(50) \qquad\qquad \frac{\partial L(\bar{\theta})}{\partial \bar{\theta}} = \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} = 0.$$

The gradient of the log-likelihood function

$$(51) \qquad\qquad \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}}$$

is a key concept for the theory of the CRB and links maximum likelihood with the CRB.

Before identifying the significant properties of the CRB, we need to review and develop a few more concepts. If $g(\bar{y})$ is an unbiased estimator of $\bar{\theta}$, then the mean

vector or expected value of $g(\bar{y})$ is equal to $\bar{\theta}$. Symbolically,

$$(52) \qquad E\left\{g(\bar{y})\right\} = \int_{-\infty}^{\infty} g(\bar{y})f(\bar{y};\bar{\theta})d\bar{y} = \bar{\theta}.$$

The covariance matrix of $g(\bar{y})$ is defined as

$$(53) \qquad C = E\left\{(g(\bar{y}) - E\{g(\bar{y})\})(g(\bar{y}) - E\{g(\bar{y})\})^H\right\},$$

but for the unbiased estimator

$$(54) \qquad C = E\left\{(g(\bar{y}) - \bar{\theta})(g(\bar{y}) - \bar{\theta})^H\right\}.$$

The expression

$$(55) \qquad g(\bar{y}) - \bar{\theta}$$

is known as the estimator error vector.

Returning to the gradient of the log-likelihood function, L, we find that if a "regularity" condition is satisfied,

$$(56) \qquad \frac{\partial \ln f(\bar{y};\bar{\theta})}{\partial \bar{\theta}}$$

has statistical properties of its own. The "regularity" condition assumes

$$(57) \qquad \frac{\partial}{\partial \bar{\theta}} \int_{-\infty}^{\infty} f(\bar{y};\bar{\theta})d\bar{y} = \int_{-\infty}^{\infty} \frac{\partial f(\bar{y};\bar{\theta})}{\partial \bar{\theta}}d\bar{y}.$$

This is generally true whenever the domain of the non-zero portion of the PDF is not a function of $\bar{\theta}$ and is definitely true in our application. The need for this condition

20

is critical to the derivation of the CRB in Appendix C and is recognized here to satisfy the last concept required for identifying the following significant properties of the theory of the CRB:

Property One. If a PDF, $f(\bar{y}; \bar{\theta})$, satisfies the "regularity" condition, then the error covariance matrix, C, of any unbiased estimator, $g(\bar{y})$, must satisfy the condition that $C - F(\bar{\theta})^{-1}$ is positive semi-definite. The matrix, $F(\bar{\theta})$, is the Fisher information matrix and is defined as the covariance matrix of the gradient of the log-likelihood function

$$(58) \qquad F(\bar{\theta}) = E\left\{ \left(\frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \right) \left(\frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \right)^{H} \right\}.$$

For the difference of matrices, $C - F(\bar{\theta})^{-1}$, to be positive semi-definite, each element of the diagonal of C must be greater than or equal to each element of the diagonal of $F(\bar{\theta})^{-1}$. This implies

$$(59) \qquad C_{nn} = E\left\{ (g(\bar{y})_n - \bar{\theta}_n)(g(\bar{y})_n - \bar{\theta}_n)^{H} \right\} \geq F(\bar{\theta})_{nn}^{-1}.$$

In words, the nnth element of the inverse of the Fisher information matrix is a lower bound on the variance of the nth parameter estimated by the unbiased estimator $g(\bar{y})$. Calculating the Fisher information matrix for a known PDF to test the performance of an unbiased estimator is straight forward. For the complex superimposed exponential problem, closed form expressions for the calculations are presented by Steedly and Moses in (67).

Property Two. An unbiased estimator, $g(\bar{y})$, may be found that attains the CRB, in that $C = F(\bar{\theta})^{-1}$, if and only if

$$(60) \qquad \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} = F(\bar{\theta}) \left(g(\bar{y}) - \bar{\theta} \right).$$

The resulting estimator, $g(\bar{y})$, is the minimum variance unbiased (MVU) estimator of $\bar{\theta}$.

With regards to the second property, the ability to manipulate the log-likelihood function into the form

$$(61) \qquad F(\bar{\theta}) \left(g(\bar{y}) - \bar{\theta} \right)$$

may be difficult. This is definitely the case for our superimposed exponential problem, but we can demonstrate the capability by considering a simplified version of the same problem. Assume the exponentials, λ_n in V, and the noise variance, σ^2, of the superimposed exponential PDF, $f(\bar{y}; V, \bar{c}, \sigma^2)$, are known. Therefore, the amplitude coefficients, $\bar{c}$, are the unknown parameters to be estimated. From the development in the last section, the log-likelihood function of the superimposed exponential PDF is

$$(62) \qquad L(\bar{c}) = -M \ln \pi - M \ln \sigma^2 - \frac{1}{\sigma^2} (\bar{y} - V\bar{c})^H (\bar{y} - V\bar{c}).$$

Since L in this case is only a function of $\bar{c}$, we can maximize L by solving for $\bar{c}$ at the stationary point $\frac{\partial L}{\partial \bar{c}} = \bar{0}$ or $\frac{\partial L}{\partial \bar{c}^*} = \bar{0}$. From Appendix B we know

$$
\begin{aligned}
\frac{\partial L(\bar{c})}{\partial \bar{c}^*} &= -\frac{1}{\sigma^2}\left(-V^H \bar{y} + V^H V \bar{c}\right) \\
&= \frac{1}{\sigma^2}\left(V^H \bar{y} - V^H V \bar{c}\right) \\
\text{(63)} \qquad &= \frac{V^H V}{\sigma^2}\left((V^H V)^{-1} V^H \bar{y} - \bar{c}\right).
\end{aligned}
$$

From the second property of the theory of the CRB,

$$
\text{(64)} \qquad\qquad F^{-1}(\bar{c}) = C = \sigma^2 (V^H V)^{-1}
$$

and

$$
\text{(65)} \qquad\qquad g(\bar{y}) = (V^H V)^{-1} V^H \bar{y}.
$$

Recall from the development of the ML estimator in the previous section, after accurately estimating λ_n, the amplitude coefficients, $\bar{c}$, are estimated by solving the same expression

$$
\text{(66)} \qquad\qquad \bar{c} = (V^H V)^{-1} V^H \bar{y} = g(\bar{y}).
$$

Furthermore, knowing $\bar{y}$ is distributed

$$
\text{(67)} \qquad\qquad \bar{y} \sim CN(V\bar{c}, \sigma^2 \mathrm{I}),
$$

we can use the linear transformation or pseudo-inverse, $(V^H V)^{-1} V^H$ from Equation 66 to determine the distribution of the estimator $g(\bar{y})$ (61:59). The result is

$$
(68) \quad
\begin{aligned}
g(\bar{y}) &\sim \mathcal{N}\left(\left(V^H V\right)^{-1} V^H V \bar{c}, \left(V^H V\right)^{-1} V^H \sigma^2 I V \left(V^H V\right)^{-1} \right) \\
\Rightarrow \quad g(\bar{y}) &\sim \mathcal{N}\left(\bar{c}, \sigma^2 \left(V^H V\right)^{-1} \right).
\end{aligned}
$$

Indeed, $g(\bar{y})$ is an unbiased estimator with mean vector $\bar{c}$ and covariance matrix $\sigma^2 (V^H V)^{-1}$ as identified from the second property of the theory of the CRB.

This exercise highlights the relationship between maximum likelihood estimators and the CRB. Although we did not use the second property of the CRB for developing the ML estimator of the completely unknown superimposed exponential problem, we do use the first property to calculate the CRB for testing all the unbiased estimators discussed in this dissertation. Additionally, the second property is implicitly referenced anytime we claim an estimator to be MVU.

2.4   Chapter Conclusion

In this review chapter, we developed the maximum likelihood cost function for the superimposed exponential parameter estimation problem. We showed equivalent relationships for representing the ML cost function in terms of all the parameters, V, $\bar{c}$, and σ^2; just the exponentials, V; and even the linear prediction coefficients, $\bar{b}$. With a previous estimate of the LP coefficients, we identified how the ML cost function can be approximated by a quadratic for a new $\bar{b}$ estimate.

The impetus for developing a maximum likelihood estimator comes form its relationship to the Cramer-Rao bound. We showed this relationship and identified two significant properties of the theory of the CRB. At this point, we can conclude that the ML approach is a logical methodology for finding the estimator that is, or

gets as close as possible to being, the MVU parameter estimator for the superimposed exponential problem. In the next chapter, we present an alternative methodology, solely based on linear prediction, for comparison.

III. Linear Prediction

3.1 Chapter Introduction

Linear prediction (LP) based parameter estimators for the superimposed exponential problem date back to 1795 and were originally proposed by le Baron de Prony (9). Although initially inadequate, they have been continually improved upon. In this chapter, we begin with the original Prony estimator and highlight the significant improvements made to LP estimators over their history. We culminate the chapter by deriving our own linear prediction based estimator that possesses a cost function identical to that of the maximum likelihood (ML) cost function developed in Chapter II. With an identical cost function, the LP estimator acquires all the desirable attributes associated with ML and the theory of the Cramer-Rao bound (CRB). Establishing an equivalence between the linear prediction based methodology and the maximum likelihood based methodology is an original contribution of this dissertation. Giving linear prediction the capability to possibly obtain the minimum variance unbiased (MVU) estimator provides an alternative to the ML methodology and allows us to retain the critical insight gained during linear prediction's long and arduous development. In Chapter VI, this insight leads to a significantly improved estimator for deep level transient spectroscopy (DLTS) applications. The same estimator would be extremely difficult to directly develop from maximum likelihood.

3.2 Least Squares

Consider the noiseless superimposed exponential signal

$$
(69) \qquad\qquad s[m] = c_1 \lambda_1^m + c_2 \lambda_2^m + \ldots + c_N \lambda_N^m.
$$

From the equation, it is apparent that each element, $s[m]$, is a linear weighted combination of the λ_n exponentials. If, once again, we assume the exponentials are known and the weights or amplitude coefficients, c_n, are the parameters to be estimated, we are inclined to solve for $c_1 \ldots c_N$ with a system of N linear equations and N unknowns. Only N realizations of $s[m]$ are necessary to construct an exactly determined system of linear equations, implying that, without noise, the solutions for $c_1 \ldots c_N$ are exact.

When the exponentials are also unknown, the linear relationship in Equation 69 is no longer apparent. Prony has been credited with realizing that the summed exponential signal of Equation 69 is the solution to a linear difference or linear prediction equation (61:489)

$$(70) \qquad b_0 s[m] + b_1 s[m-1] + \ldots + b_N s[m-N] = 0.$$

This fact was demonstrated in Chapter II. Recall that the exponentials, λ_n, are the roots of the polynomial formed by the linear prediction coefficients, $b_0 \ldots b_N$. The LP polynomial takes the form

$$(71) \qquad b_0 z^N + b_1 z^{N-1} + \ldots + b_N.$$

Notice that Equation 70 is a linear weighted combination of consecutive realizations of $s[m]$. Once again, we are inclined to construct an exactly determined system of linear equations to solve for the LP coefficients $b_0 \ldots b_N$. If we assume b_0 normalizes to 1, the soluble roots of the LP polynomial are unaffected, and only $2N$ realizations of $s[m]$ are necessary to develop N equation for the N unknowns, $b_1 \ldots b_N$.

After solving $b_1 \ldots b_N$ and rooting the exponentials, λ_n, from the LP polynomial, the amplitude coefficients, c_n, can be exactly determined from the previous system of linear equations based on the noiseless signal model Equation 69. This methodology is known as the original Prony method. It conveys the fundamental concepts of linear prediction in the context of parameter estimation for the superimposed exponential problem. The elegance of this estimation technique is easily overlooked. It embodies the classic approach of decomposing a nonlinear problem into separate problems which are solved using linear methods, and as long as the data is completely modeled and noiseless, an exact solution is obtained.

To emphasize further, we will apply the original Prony method to four data points from a noiseless two mode signal. Step One of the original Prony method—and all other LP based methodologies for estimating the parameters of a superimposed exponential signal—requires the construction of the system of linear equations

$$
\begin{aligned}
s[2] + b_1 s[1] + b_2 s[0] &= 0 \\
s[3] + b_1 s[2] + b_2 s[1] &= 0
\end{aligned}
\tag{72}
$$

or

$$
\begin{bmatrix} s[1] & s[0] \\ s[2] & s[1] \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} -s[2] \\ -s[3] \end{bmatrix}
\tag{73}
$$

for solving for the LP coefficients b_1 and b_2. After solving for b_1 and b_2 we root λ_1 and λ_2 from the LP polynomial

$$
z^2 + b_1 z + b_2.
\tag{74}
$$

Step Two requires the construction of the system of linear equations

$$s[0] = c_1\lambda_1^0 + c_2\lambda_2^0$$

(75)
$$s[1] = c_1\lambda_1^1 + c_2\lambda_2^1$$

or

(76)
$$\begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} s[0] \\ s[1] \end{bmatrix}$$

for solving for the amplitude coefficients c_1 and c_2.

When noise is added to the signal, neither the system of linear equations for the LP coefficients nor the system of linear equations for the amplitude coefficients yield exact solutions. Because the LP coefficient estimator is developed solely from a noiseless LP equation, its estimates are easily corrupted by noise (44:275). To make matters worse, in most LP methodologies, the amplitude coefficient estimates from Step Two are based on exponential estimates from the roots of the LP polynomial formed from LP coefficient estimates in Step One. Inaccuracies in the LP coefficient estimates from Step One are compounded throughout the algorithm.

To further analyze the effects of noise and introduce improvements made to linear prediction, consider all the data points of the noisy signal in the vector notation

(77)
$$\bar{y} = V\bar{c} + \bar{w}$$

or

(78)
$$\bar{y} - V\bar{c} = \bar{w}.$$

Even though we have an assumed distribution for $\bar{w}$, the assumption is unnecessary for the development that follows. The vector $\bar{w}$ is simply the error vector between the observed noisy signal, $\bar{y}$, and the assumed model, $V\bar{c}$. One possible approach to estimating the parameters of the assumed model is to search for the λ_n and c_n that minimizes the sum of the squared error

$$
\begin{aligned}
J(V, \bar{c}) &= \bar{w}^H \bar{w} \\
&= (\bar{y} - V\bar{c})^H (\bar{y} - V\bar{c}).
\end{aligned}
$$

(79)

This approach is known as least squares (LS), and in this case, the least squares cost function, $J(V, \bar{c})$, is identical to the maximum likelihood cost function, $L(V, \bar{c})$ developed in Chapter II.

With unknown V and $\bar{c}$, minimizing $J(V, \bar{c})$ is a non-linear least squares problem, but for further insight, once again assume the exponentials are known. Minimizing $J(\bar{c})$ is a linear least squares problem and in Chapter II we developed the minimum solution to be the pseudo-inverse estimator

$$
\bar{c} = (V^H V)^{-1} V^H \bar{y}.
$$

(80)

We can prove the pseudo-inverse estimator obtains a unique minimal solution by completion of squares. Let $\bar{c}_0 = (V^H V)^{-1} V^H \bar{y}$ and expand $J(\bar{c})$ so that

$$
\begin{aligned}
J(\bar{c}) &= (\bar{y}_{\bar{c}} - V(\bar{c} + \bar{c}_0 - \bar{c}_0))^H (\bar{y}_{\bar{c}} - V(\bar{c} + \bar{c}_0 - \bar{c}_0)) \\
&= (\bar{y}_{\bar{c}} - V\bar{c}_0)^H (\bar{y}_{\bar{c}} - V\bar{c}_0) + (\bar{c} - \bar{c}_0)^V V^H V(\bar{c} - \bar{c}_0).
\end{aligned}
$$

(81)

The left term of Equation 81 is not a function of $\bar{c}$ and is therefore fixed by the data. The right term of Equation 81 is of quadratic form in $\bar{c}$. In Appendix A we show V has full column rank. Because V has more rows, M, than columns, N, we also know $V^H V$ is positive definite (33:26). This implies

$$(82) \qquad\qquad (\bar{c} - \bar{c}_0)^V V^H V (\bar{c} - \bar{c}_0) > 0$$

for all $\bar{c} \neq \bar{c}_0$. Therefore, the unique minimum $J(\bar{c})$ occurs when $\bar{c} = \bar{c}_0 = (V^H V)^{-1} V^H \bar{y}$. Also, recall that when the uncorrelated zero mean Gaussian distributional assumption about $\bar{w}$ is made,

$$(83) \qquad\qquad \bar{c} = (V^H V)^{-1} V^H \bar{y}$$

is the MVU estimator for the simplified superimposed exponential problem.

Therefore, when considering the completely unknown parameter estimation problem, if we can accurately estimate the exponentials in Step One of the LP methodology, we can utilize the pseudo-inverse estimator to attain the best possible estimate of the amplitude coefficients, even in noise. Thus, linear prediction research for the superimposed exponential parameter estimation problem is focused on Step One of the LP methodology: estimating the linear prediction coefficients.

As mentioned before, the LP equation

$$(84) \qquad\qquad b_0 s[m] + b_1 s[m-1] + \ldots + b_N s[m-N] = 0$$

is only exact under the noiseless signal assumption. If we substitute the noisy realizations, $y[m]$, for the noiseless realizations, $s[m]$, we can say

$$(85) \qquad b_0 y[m] + b_1 y[m-1] + \ldots + b_N y[m-N] = e[m]$$

where $e[m]$ is the unknown error invoked by the substitution. Using all the available data, we can construct the overdetermined system of linear equations

$$(86) \qquad Y\bar{b} = \bar{e}$$

where

$$
\begin{aligned}
Y &= \begin{bmatrix} \bar{y}_0 & \bar{y}_1 & \cdots & \bar{y}_N \end{bmatrix} \\
\bar{y}_n &= \begin{bmatrix} y[N-n] & y[N+1-n] & \cdots & y[M-1-n] \end{bmatrix}^T \\
\bar{b} &= \begin{bmatrix} b_0 & b_1 & \cdots & b_N \end{bmatrix}^T \\
(87) \qquad \bar{e} &= \begin{bmatrix} e[N] & e[N+1] & \cdots & e[M-1] \end{bmatrix}^T.
\end{aligned}
$$

The data matrix, Y, has dimensions $(M-N) \times (N+1)$ and Toeplitz structure. With the previous least squares development, it is logical to assume we can improve our LP coefficient estimates by minimizing the sum of the squared error

$$
\begin{aligned}
J(\bar{b}) &= \bar{e}^H \bar{e} \\
&= (Y\bar{b})^H (Y\bar{b}) \\
(88) \qquad &= \bar{b}^H Y^H Y \bar{b}.
\end{aligned}
$$

Before continuing further, note that unlike $J(\bar{c})$, $J(\bar{b})$ is not identical to its counterpart, $L(\bar{b})$, developed with the maximum likelihood methodology in Chapter II. Under maximum likelihood,

$$(89) \qquad L(\bar{b}) = \bar{b}^H Y^H (BB^H)^{-1} Y \bar{b}.$$

Therefore, the $\bar{b}$ that minimizes $J(\bar{b})$ is not a maximum likelihood estimate.

In the iterative quadratic maximum likelihood (IQML) algorithm, we addressed this by using a previous estimate of $\bar{b}$ to calculate and fix $(BB^H)^{-1}$ for a new estimate of $\bar{b}$. If an initial estimate of

$$(90) \qquad \bar{b} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$$

is used for the IQML algorithm, the resulting matrix

$$(91) \qquad (BB^H)^{-1} = I$$

shows that the initial estimate is equivalent to starting the IQML algorithm with a least squares estimate. This is a popular technique for initializing the IQML algorithm and is discussed in detail in Chapter V.

Historically, the least squares improvement was introduced to linear prediction by Hildebrand in the 1950s (22:457-462). At the time, the relationship between the least squares cost function and the maximum likelihood cost function was unknown. Dubbed the extended Prony method, the least squares modification offers considerable improvement over the original Prony method (33:225). The covariance method presented by Makhoul in 1975 is identical to the extended Prony method but is de-

rived under entirely different assumptions (50:564). In this dissertation, we find it informative to reference the extended Prony method as the least squares algorithm.

Before demonstrating the improved performance gained by the least squares algorithm over the original Prony method, we will further define the least squares LP coefficient cost function $J(\bar{b})$. Let

$$(92) \qquad \tilde{Y} = \left[\begin{array}{cccc} \bar{y}_1 & \bar{y}_2 & \cdots & \bar{y}_N \end{array} \right]$$

and

$$(93) \qquad \tilde{b} = \left[\begin{array}{cccc} b_1 & b_2 & \cdots & b_N \end{array} \right]^T .$$

so that

$$
\begin{array}{rcl}
J(\bar{b}) & = & (Y\bar{b})^H (Y\bar{b}) \\[2mm]
& = & (\bar{y}_0 + \tilde{Y}\tilde{b})^H (\bar{y}_0 + \tilde{Y}\tilde{b}) \\[2mm]
(94) & = & J(\tilde{b}).
\end{array}
$$

Again we have assumed b_0 normalizes to 1. Also, $\tilde{Y}$ is an $(M - N) \times N$ data matrix with Toeplitz structure. In Appendix A we describe how $\tilde{Y}$ is assumed to have full column rank N. Therefore, we assume $\tilde{Y}^H \tilde{Y}$ is invertible and like $J(\bar{c})$, we can minimize $J(\tilde{b})$ with the unique pseudo-inverse solution

$$(95) \qquad \tilde{b} = -(\tilde{Y}^H \tilde{Y})^{-1} \tilde{Y}^H \bar{y}_0.$$

In review, the least squares algorithm, for estimating the parameters of superimposed exponential signals, follows the same LP methodology as the original Prony

method. However, in both the LP coefficient estimation portion and the amplitude coefficient estimation portion, the exactly determined solutions of the original Prony method are replaced with overdetermined, least squares, solutions.

To show the superior performance of the LS algorithm over the Prony method, simulated noisy signals were analyzed in a Monte-Carlo experiment. A complex, 25 data point, two mode, noiseless signal was created with parameters $c_1 = c_2 = 1$, $\lambda_1 = e^{j2\pi(.20)}$, and $\lambda_2 = e^{j2\pi(.22)}$. Two randomly generated 25 element vectors were then added to the real and imaginary elements of the noiseless signal, respectively. The randomly generated elements of each vector were Gaussian distributed with mean zero and variance $\sigma^2/2$. The variance was governed by SNR, in dB, under the formula

$$(96) \qquad \sigma^2 = \frac{|c_1|^2}{10^{\frac{SNR}{10}}}.$$

so that

$$(97) \qquad SNR = 10\log_{10}\left(\frac{|c_1|^2}{\sigma^2}\right).$$

Two hundred realizations of the noisy signal were created for each SNR. The parameters of each realization were estimated with both the LS algorithm and the original Prony method. For each parameter, the error between each estimate and the actual parameter was calculated and recorded in both magnitude and phase. After two hundred realizations, the magnitude and phase bias was calculated for each parameter by summing the error and dividing by the number of realizations. Also, the mean square error (MSE) was calculated by summing the square of each

35

Figure 1. The λ_1 phase inverse MSE of the LS algorithm and the original Prony
method over a range of SNRs. Also plotted is the CRB.

error and dividing by the number of realizations. This process was repeated for every

SNR ranging form 0 dB to 100 dB in 2 dB intervals.

The results of the experiment are displayed in plots of magnitude and phase

MSE versus SNR and magnitude and phase bias versus SNR for each parameter

estimated. The MSE plots include the CRB for reference and the ordinate of the

MSE plots is scaled to $10\log_{10}(1/\text{MSE})$. The $10\log_{10}$ scaling is justified because

the CRB obeys a log-linear relationship with respect to SNR. In the plot, the linear

bound assists with visualization. The inverse of MSE is justified by the convention

of plotting superior performance above inferior performance.

The performance of both estimators can be assessed in Figures 1, 2, 3, and 4.

The first two figures pertain to the MSE and bias of phase errors in the λ_1 parameter

estimates. As described earlier, we consider λ_n estimates to be the primary indicators

of estimator performance because of their influence on the remainder of most LP

36

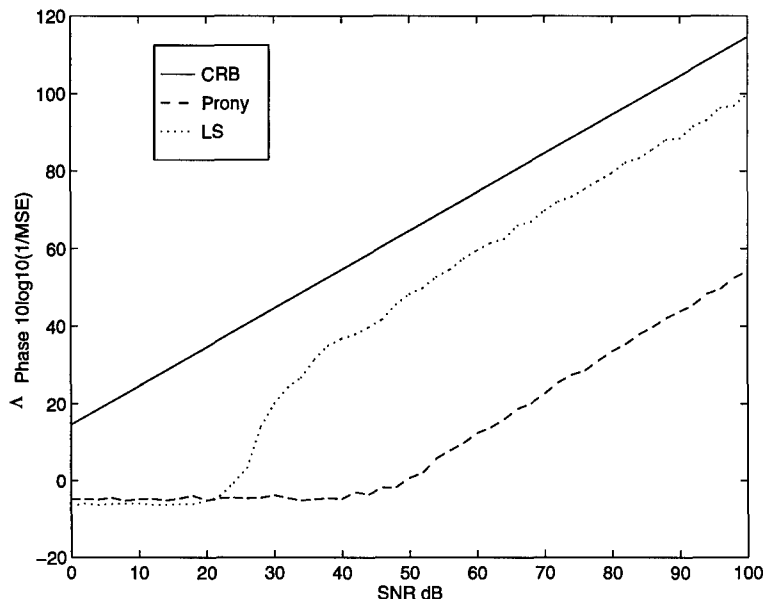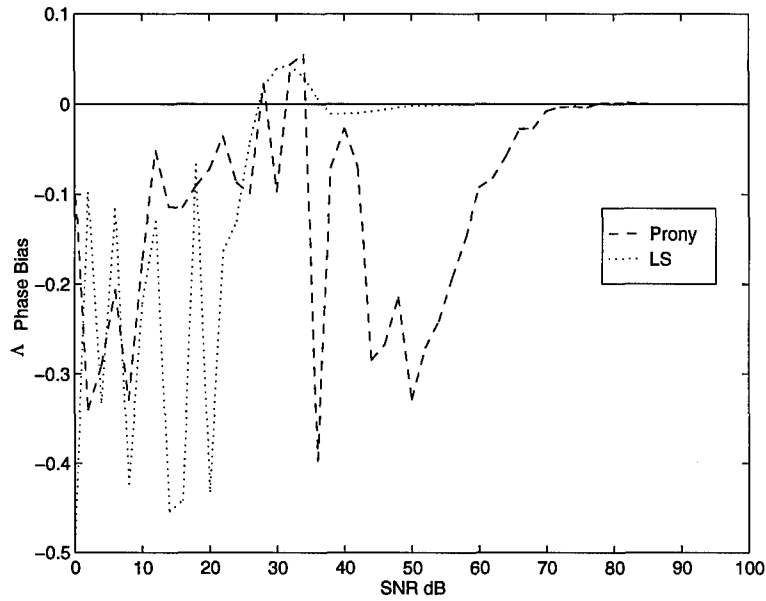Figure 2. The λ_1 phase bias of the LS algorithm and the original Prony method over a range of SNRs.
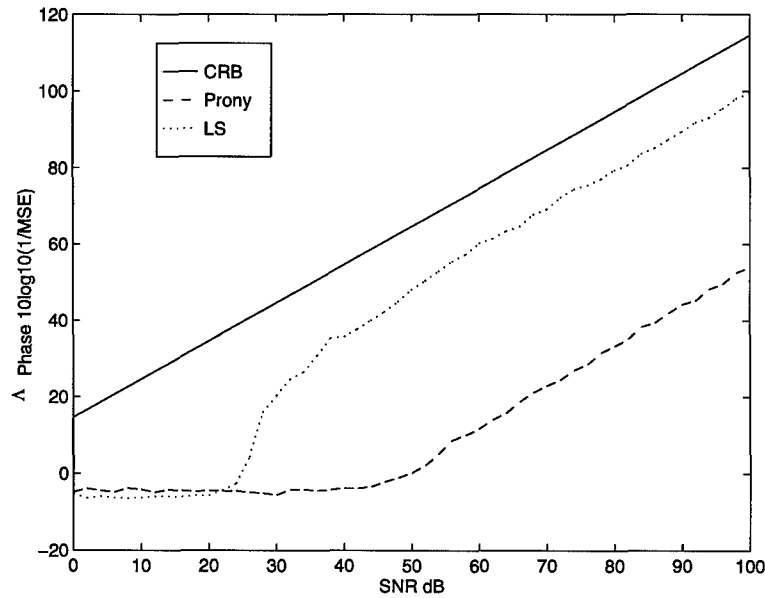
Figure 3. The λ_2 phase inverse MSE of the LS algorithm and the original Prony method over a range of SNRs. Also plotted is the CRB.

37
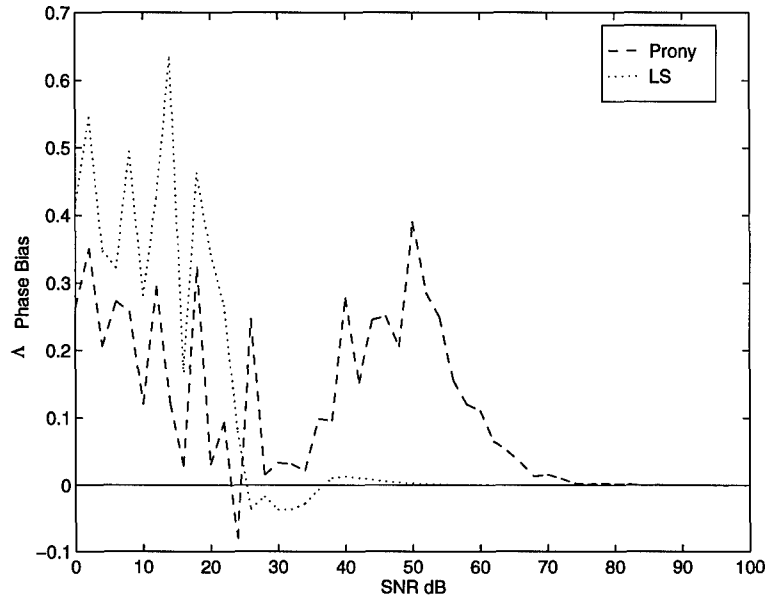
Figure 4. The λ_2 phase bias of the LS algorithm and the original Prony method over a range of SNRs.

algorithms. For this complex experiment, performance assessment in the phase of λ_n is appropriate because the phase of λ_n is the only discriminator between both modes of the underlying signal. In the λ_1 phase inverse MSE plot, Figure 1 we see that above 30 dB the LS algorithm estimates are consistently closer to the CRB than the original Prony method estimates. From the λ_1 phase bias plot, Figure 2, we see that the bias becomes erratic below 70 dB for the original Prony method and below 30 dB for the LS algorithm. Above those SNRs, the estimators appear to be unbiased. Consequently, their comparison with the CRB in the inverse MSE plot is relevant. Below those SNRs, the erratic bias makes comparison with the CRB inappropriate. The rapid increase in MSE (portrayed as a decrease in the inverse MSE plot) of the LS algorithm estimator at 30 dB is called the noise threshold. The noise threshold is prevalent in many of the algorithms considered in the dissertation.

Because the associated erratic bias, few conclusions can be drawn below the noise threshold from the inverse MSE plots.

Figures 3 and 4 are the phase inverse MSE and bias plots for estimates on the λ_2 parameter. The performance of both estimators on the λ_2 parameter is very similar to the performance on the λ_1 parameter. The ultimate direction of the bias is the only real difference and because it occurs below the noise threshold we give it no attention.

The significant contribution from the experiment is verification of the least squares improvement on the linear prediction methodology. In the next section, we consider another modification to linear prediction that realized improved estimator performance but ultimately stifled further improvement until our introduction of a more appropriate alternative.

3.3 Overmodeled Least Squares

Over history, numerous researchers have realized that applying LP methodologies with model orders in excess of the assumed underlying model order normally leads to improved parameter estimates, *e.g.*, (47, 72, 71, 41, 42, 57). Bresler and Macovski categorized these approaches as "heuristic modifications of algorithms that yield exact results when there is no noise or the amount of available data is infinite (4:1081)." They support this statement by recalling that the cornerstone of all LP methodologies, the LP equation

$$(98) \qquad b_0 s[m] + b_1 s[m-1] + \ldots + b_N s[m-N] = 0,$$

is only valid in the noiseless case. Nevertheless, the practice of overmodeling was and still is accepted by many. Methodologies that incorporate overmodeling avoid

iterations, and when the SNR is sufficient, obtain estimates that are comparable to maximum likelihood (38:1208-1209).

The premise of overmodeling is that the extraneous modes model the noise component of the observed data while the N actual modes appropriately model the signal component of the observed data. The artificial model order, P, has been dubbed the prediction order. Modifying the LS algorithm to incorporate overmodeling is not difficult. We refer to the new algorithm as overmodeled least square (OLS). Instead of $N + 1$ coefficients, the LP equation in OLS has as $P + 1$ coefficients. Therefore, the LP polynomial roots to P exponentials, λ_n. At this point in the algorithm, a requirement exists to sort the N actual λ_n from the $P - N$ extraneous λ_n estimates. After sorting, the algorithm resumes the previous LS methodology for estimating the amplitude coefficients.

Numerous sorting schemes have been suggested. Kumaresan "qualitatively argued" that all extraneous λ_n are distributed uniformly around the inside of the unit circle (42:218-220). Empirically, this appears to be the case for a good range of SNRs, but exceptions occur as SNR decreases. Kumaresan used this argument as a sorting criterion for parameter estimation problems where the exponentials are constrained to the unit circle (sinusoids). In his algorithm, the $P - N$ λ_n estimates farthest inside the unit circle are deemed extraneous. Kumaresan and Tufts also applied the criterion to complex signals with decaying exponentials (41). They realized that, in noiseless conditions, if the data matrix Y is created with the data in reverse order, the resulting LP coefficients root to estimates that are the inverse of the N actual λ_n values. Therefore, they lie outside the unit circle. However, despite the reverse order of Y, the $P - N$ extraneous λ_n estimates still typically lie inside the unit circle.

Consequently, Kumaresan and Tufts developed a sorting scheme that capitalized on this behavior.

We have found an alternative criterion, known as the energy sort, to be reliable and in our opinion, more tractable. Although heuristic as well, the energy sort has been successfully applied in radar applications (60). It is predicated on the calculation of the amplitude coefficients with the unique pseudo-inverse solution

$$(99) \qquad \bar{c} = (V^T V)^{-1} V^T \bar{y}$$

derived earlier. The $\bar{c}$ estimator will also utilize the signal vector, $\bar{y}$, and an overmodeled matrix V_{over} to appropriately represent the overmodeled amplitude coefficients in an overmodeled vector $\bar{c}_{over}$. When an exponential, λ_n, is extraneous to the observed signal, $\bar{y}$, its corresponding amplitude coefficient, c_n, is typically small. If we estimate the overmodeled c_n and λ_n with the LS algorithm, the extraneous mode estimates can be sorted and separated by mode energy. The energy of a single mode is defined as

$$(100) \qquad \begin{aligned} E_n &= \sum_{m=0}^{M-1} |c_n \lambda_n^m|^2 \\ &= \begin{cases} |c_n|^2 \left(\frac{1-|\lambda_n|^{2M}}{1-|\lambda_n|^2} \right), & |\lambda_n| \neq 1 \\ |c_n|^2 M, & |\lambda_n| = 1 \end{cases} \end{aligned}.$$

Under the energy sort criterion, the N modes with the highest energy, E_n, are considered to be the actual modes of the signal. With N actual modes selected, the amplitude coefficient estimator is reapplied to acquire N amplitude coefficient estimates.
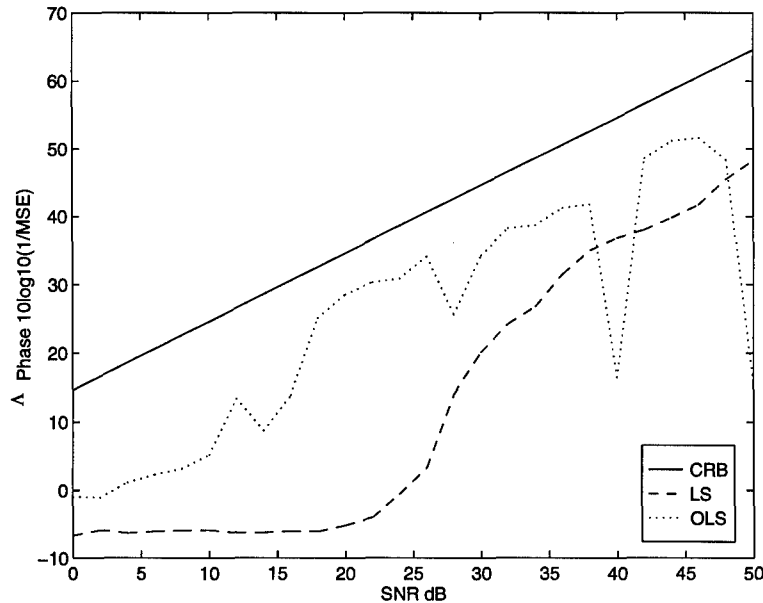
Figure 5.   The λ_1 phase inverse MSE of the OLS and LS estimators over a range of SNRs. Also plotted is the CRB.

Figures 5, 6, 7, and 8 are λ_n phase inverse MSE and bias plots for comparing the performance of the OLS and LS estimators.   The Monte-Carlo experiment performed to create the plots was identical to the previous Monte-Carlo experiment for comparing the LS algorithm and the original Prony method. A prediction order, P, of 12 was used on every OLS execution.

In the phase bias plots, we see an extended SNR range in the unbiased region for the OLS estimator. In the phase inverse MSE plots we see a corresponding SNR range where the OLS estimator performance is closer to the CRB than the LS estimator performance, but as SNR increases, the OLS estimator performance is inconsistent. This is not a desirable feature for parameter estimators in low noise scenarios.

The erratic behavior as SNR increases can be attributed to a near rank deficiency in the overmodeled data matrix. For explanation purposes, let Y_{over} and
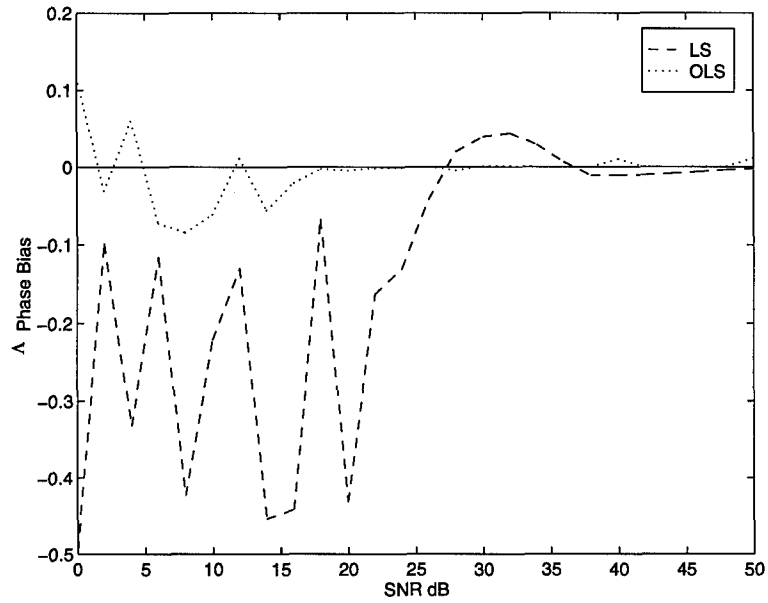
Figure 6. The λ_1 phase bias of the OLS and LS estimators over a range of SNRs.
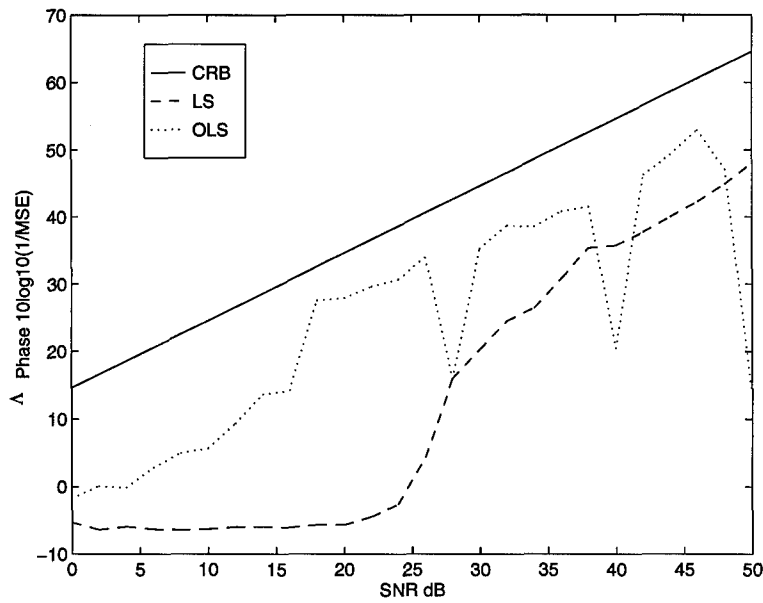
Figure 7. The λ_2 phase inverse MSE of the OLS and LS estimators over a range of SNRs. Also plotted is the CRB.

43
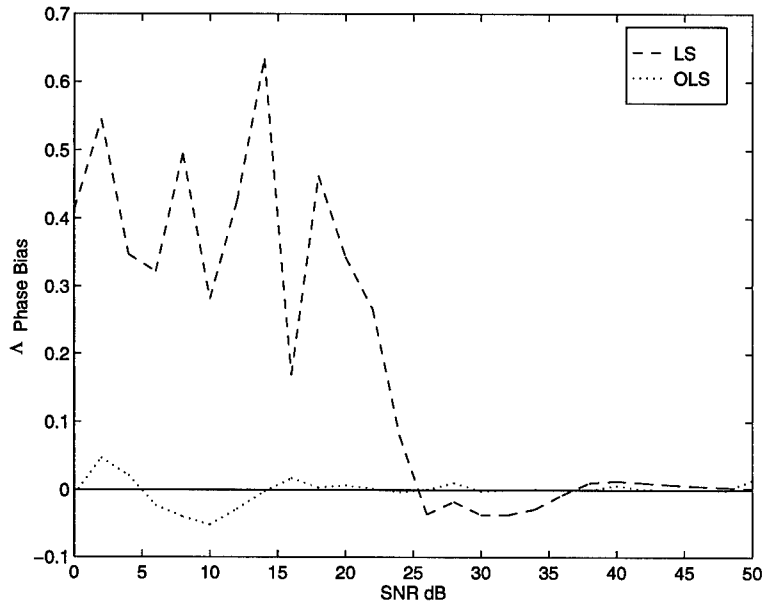
Figure 8. The λ_2 phase bias of the OLS and LS estimators over a range of SNRs.

$\tilde{Y}_{over}$ represent overmodeled data matrices identical to Y and $\tilde{Y}$ except that their dimensions are expanded from the model order N to the prediction order P. In Appendix A, we discussed how the theoretical rank of $\tilde{Y}_{over}$ is P—due to the stochastic nature of its columns—even though the underling signal matrix rank is N. As SNR increases and the noise decreases, the contrast in rank begins to influence the condition of $\tilde{Y}_{over}^H \tilde{Y}_{over}$ in the pseudo-inverse estimator

$$(101) \qquad \tilde{b}_{over} = -\left(\tilde{Y}_{over}^H \tilde{Y}_{over}\right)^{-1} \tilde{Y}_{over}^H \bar{y}_0.$$

Further insight into this problem is gained by observing the pseudo-inverse estimator in terms of the singular value decomposition (SVD). It has been shown (25:414) that for any $M \times P$ matrix Y, there exists an $M \times M$ unitary matrix

$$(102) \qquad U = [\bar{u}_1, \bar{u}_2, \cdots, \bar{u}_M]$$

44

and a $P \times P$ unitary matrix

$$(103) \qquad W = [\bar{w}_1, \bar{w}_2, \cdots, \bar{w}_P]$$

such that

$$(104) \qquad U^H Y W = \Sigma$$

where Σ is an $M \times P$ matrix of zeros except on the diagonal. The diagonal entries of Σ are real and non-negative. When $M \geq P$ they are ordered

$$(105) \qquad \sigma_{11} \geq \sigma_{22} \geq \cdots \geq \sigma_{PP} \geq 0.$$

If the rank of Y is less than P, say N for example, then

$$(106) \qquad \sigma_{N+1,N+1} = \sigma_{N+2,N+2} = \cdots = \sigma_{PP} = 0.$$

Hereafter, we reference the diagonal elements of Σ as the singular values and use the abbreviated index, σ_n. Since the singular values are ordered by size, the columns of U and W have order as well and are referenced as the left and right singular vectors, $\bar{u}_n$ and $\bar{v}_n$, respectively.

With the SVD defined, we can return to the pseudo-inverse estimator for $\tilde{b}_{over}$. With the guidance of Golub and Van Loan (19:242), we introduce the SVD to this problem. Recall that we desire the $\tilde{b}_{over}$ that minimizes the sum of the squared error

$$
\begin{aligned}
J(\tilde{b}_{over}) &= (\bar{y}_0 + \tilde{Y}_{over}\tilde{b}_{over})^H (\bar{y}_0 + \tilde{Y}_{over}\tilde{b}_{over}) \\
(107) &= \left\| \bar{y}_0 + \tilde{Y}_{over}\tilde{b}_{over} \right\|_2^2.
\end{aligned}
$$

45

Let

$$(108) \qquad U^H \tilde{Y}_{over} W = \tilde{\Sigma}_{over}$$

as defined by the SVD. Because orthogonal transformations do not change the 2-norm of a vector,

$$
\begin{aligned}
J(\tilde{b}_{over}) &= \left\| U^H \bar{y}_0 + \left(U^H \tilde{Y}_{over} W \right) \left(W^H \tilde{b}_{over} \right) \right\|_2^2 \\
&= \left\| U^H \bar{y}_0 + \tilde{\Sigma}_{over} \left(W^H \tilde{b}_{over} \right) \right\|_2^2 .
\end{aligned}
$$

(109)

Also let $\bar{\alpha} = W^H \tilde{b}_{over}$, so that

$$
\begin{aligned}
J(\tilde{b}_{over}) &= \left\| U^H \bar{y}_0 + \tilde{\Sigma}_{over} \bar{\alpha} \right\|_2^2 \\
&= \sum_{n=1}^{P} \left| \bar{u}_n^H \bar{y}_0 + \sigma_n \alpha_n \right|^2 .
\end{aligned}
$$

(110)

From the last expression in Equation 110, we see that $\alpha_n = -\frac{1}{\sigma_n} \bar{u}_n^H \bar{y}_0$ is the minimal solution to the LS problem. Therefore, since $\tilde{b}_{over} = W \bar{\alpha}$, the unique minimum estimate of $\tilde{b}_{over}$ is

$$(111) \qquad \tilde{b}_{over} = -W \tilde{\Sigma}_{over}^+ U^H \bar{y}_0$$

where $\tilde{\Sigma}_{over}^+$ is a $P \times (M - N)$ matrix of zeros except on the diagonal where $\sigma_{nn}^+ = \frac{1}{\sigma_n}$. The matrix $W \tilde{\Sigma}_{over}^+ U^H$ is also the pseudo inverse of $\tilde{Y}_{over}$, and when $\tilde{Y}_{over}$ is full rank,

$$
\begin{aligned}
\tilde{b}_{over} &= -W \tilde{\Sigma}_{over}^+ U^H \bar{y}_0 \\
&= -\left(\tilde{Y}_{over}^H \tilde{Y}_{over} \right)^{-1} \tilde{Y}_{over}^H \bar{y}_0 .
\end{aligned}
$$

(112)

Since the underlying signal matrix of $\tilde{Y}_{over}$ has a rank of N, we know as SNR increases, the smallest $P - N$ singular values approach zero. Therefore, the inverse of $\sigma_{N+1} \ldots \sigma_P$, on the diagonal of $\tilde{\Sigma}_{over}^+$, approach ∞. This action corrupts the estimate of $\tilde{b}_{over}$ and leads to an erroneous but unique solution. To alleviate this problem, Tufts and Kumaresan (72) propose intervening at the $\tilde{b}_{over}$ estimate and forcing the last $P - N$ singular values to zero. The effect of this action is to create a low rank, N, approximation of $\tilde{Y}_{over}$, call it $\hat{\tilde{Y}}_{over}$, where

$$(113) \qquad U^H \hat{\tilde{Y}}_{over} W = \hat{\tilde{\Sigma}}_{over}.$$

The unitary matrices, U and W, are the original unitary matrices from the SVD of $\tilde{Y}_{over}$, and the matrix $\hat{\tilde{\Sigma}}_{over}$ is the original $\tilde{\Sigma}_{over}$ with the last $P - N$ singular values set to zero. Although, with a rank deficient data matrix, a unique solution no longer exists, the SVD can be utilized to obtain a minimum 2-norm solution for $\tilde{b}_{over}$. For this reason, a rank deficient data matrix is preferred over a near rank deficient data matrix (3:510–511).

A proof of the minimal 2-norm solution follows the same derivation from Golub and Van Loan (19:242) as the full rank LS problem with the SVD. In the rank deficient scenario,

$$
(114) \qquad
\begin{aligned}
J(\tilde{b}_{over}) &= \left\| U^H \bar{y}_0 + \hat{\tilde{\Sigma}}_{over} \bar{\alpha} \right\|_2^2 \\
&= \sum_{n=1}^{N} |\bar{u}_n^H \bar{y}_0 + \hat{\sigma}_n \alpha_n|^2 + \sum_{n=N+1}^{P} |\bar{u}_n^H \bar{y}_0|^2
\end{aligned}
$$

because $\hat{\sigma}_{N+1} = \hat{\sigma}_{N+2} = \ldots = \hat{\sigma}_P = 0$. Again, $\alpha_n = -\frac{1}{\hat{\sigma}_n}\bar{u}_n^H \bar{y}_0$ obtains the minimal solution to $J(\tilde{b}_{over})$. Therefore,

$$(115) \qquad \tilde{b}_{over} = -W\hat{\tilde{\Sigma}}_{over}^+ U^H \bar{y}_0$$

is a minimal 2-norm solution for $\tilde{b}_{over}$. Note that with $W\hat{\tilde{\Sigma}}_{over}^+ U^H$, the pseudo-inverse is calculated even with a rank deficient $\hat{\tilde{Y}}_{over}$.

In 1982, Tufts and Kumaresan used the SVD and a low rank approximation of $\tilde{Y}_{over}$ for the superimposed exponential parameter estimation problem to effectively remove the erratic behavior seen in the OLS estimator at high SNR (72). In 1987, Rahman and Yu (57) used the SVD and a low rank approximation in a total least squares (TLS), rather than a least squares, methodology for estimating the LP coefficients. With this adjustment, they were able to attain slightly better parameter estimates. Ironically, four years before Rahman and Yu, Tufts and Kumaresan incorporated the same adjustment into their algorithm (36), but the equivalence between their new algorithm and Rahman and Yu's algorithm was not realized until 1991 (13). We will now introduce total least squares as a subtle extension to the SVD enhanced OLS algorithm just developed in this dissertation. Our delivery is unique and tractable.

Conceptually, TLS dates back to the early 1900s. Wide spread use of TLS did not occur until the 1980s when Golub and Van Loan utilized the SVD to increase the method's ease of implementation and understanding (18). For the superimposed

exponential problem, TLS can be applied to minimize the sum of the squared error

$$
\begin{aligned}
J(\tilde{b}_{over}) &= \left\| \bar{y}_0 + \tilde{Y}_{over}\tilde{b}_{over} \right\|_2^2 \\
&= \left\| Y_{over} \begin{bmatrix} 1 \\ \tilde{b}_{over} \end{bmatrix} \right\|_2^2 .
\end{aligned}
$$

(116)

From Appendix A, we know the rank of Y_{over} is $P+1$ even if it is near rank deficient. Therefore, there is no solution, other than a vector of zeros, that will cause $J(\tilde{b}_{over})$ to equal 0. In other words, a vector of zeros is the only vector in the null space of Y_{over}. The minimizing problem of Equation 116 can be recast, equivalently, as a search for the minimum perturbation of Y_{over} so that $[\ 1 \quad \tilde{b}_{over}^T\ ]^T$ is a solution to

(117)
$$
(Y_{over} - \Delta Y_{over}) \begin{bmatrix} 1 \\ \tilde{b}_{over} \end{bmatrix} = \bar{0}
$$

where ΔY_{over} is the perturbation (28:33). For the nonzero solution, $[\ 1 \quad \tilde{b}_{over}^T\ ]^T$, to exist, the perturbation matrix must be such that $(Y_{over} - \Delta Y_{over})$ has a rank less than $P + 1$. As shown previously, the SVD can be used to create a low rank approximation of Y_{over}, call it $\hat{Y}_{over}$, by forcing the smallest singular values of Y_{over} to equal zero. The Eckart-Young-Mirsky matrix approximation theorem (51) proves, in the matrix 2-norm and the Frobenius norm, that a low rank matrix approximation created with the SVD accomplishes the rank reduction with the minimum possible perturbation. Consequently, the SVD facilitates the TLS methodology.

The simple least squares approach can also be recast into a search for the minimum perturbation of $\bar{y}_0$ so that

$$(118) \qquad (\bar{y}_0 - \Delta\bar{y}_0) + \tilde{Y}_{over}\tilde{b}_{over} = \bar{0}$$

where $\Delta\bar{y}_0$ is the perturbation (28:33). The prevailing argument for TLS is that the ability to perturb the entire data matrix instead of just one column of the data matrix obtains a better estimate of $\tilde{b}_{over}$, e.g., (57, 27, 13, 1, 6). We address this issue further in Chapter IV, but for the overmodeled LP approach, it appears TLS is slightly more effective than LS. We believe a portion of the improvement, if not all of the improvement, is due to a lesser exalted minimum norm solution provided by TLS and essential to the overmodeled application. The minimum norm solution was also identified by Golub and Van Loan in their introductory TLS article (18:891–892). To develop it we need to review another feature of the SVD.

We know for any $M \times P$ matrix Y with rank less than P, say N,

$$
\begin{aligned}
U^H Y W &= \Sigma \\
\Rightarrow \quad Y W &= U\Sigma \\
\Rightarrow \quad Y\bar{w}_n &= \sigma_n \bar{u}_n.
\end{aligned}
$$

(119)

Therefore, the right most linearly independent vectors of W, $\bar{w}_{N+1} \ldots \bar{w}_P$, span the space where $\sigma_{N+1} \ldots \sigma_P$ equals zero. Consequently, $Y\bar{w}_{N+1} \ldots Y\bar{w}_P$ equals $\bar{0}$. Thus, the right most $(N+1) \ldots P$ singular vectors of W are a basis for the null space of Y. In other words, the range space of $\bar{w}_{N+1} \ldots \bar{w}_P$ is the null space of Y.

In our application, if we only require the low rank approximation $\hat{Y}_{over}$ to have a rank 1 less then the rank of Y_{over}, there is only one vector, $\bar{w}_{P+1}$, in the null space of $\hat{Y}_{over}$. If we normalize that vector by dividing the first element, $\bar{w}_{1,P+1}$, into all the

elements of $\bar{w}_{P+1}$, we have the unique minimum solution for $[\ 1\ \ \tilde{b}_{over}^{T}\ ]^{T}$. Typically, in the overmodeled LP coefficient estimation problem, we desire to approximate Y_{over}, of rank $P+1$, with a matrix $\hat{Y}_{over}$, of rank N, where N is significantly less than P. Therefore, the null space of $\hat{Y}_{over}$ is the range space of $\bar{w}_{N+1}\ldots\bar{w}_P$ and an infinite number of solutions exists.

To develop Golub and Van Loan's technique for determining the minimum norm solution with a multi-dimensional null space, let the $P\times1$ vector

$$(120) \qquad\qquad \tilde{b}_{TLS} = \frac{1}{\gamma}[\bar{x}]$$

symbolize the minimum norm solution. Also let

$$(121) \qquad\qquad W_2 = \left[\ \bar{w}_{N+1}\ \ \bar{w}_{N+2}\ \ \ldots\ \ \bar{w}_{P+1}\ \right]$$

represent the basis of all solutions. We desire a linear transformation, H, such that

$$(122) \qquad\qquad W_2 H = \left[\begin{array}{c|c} 0\cdots0 & \gamma \\ \hline \diagdown\diagup & \bar{x} \end{array}\right]_{(P+1)\times(P-N+1)}$$

Thus, the goal of H is to rotate the energy of the top row of W_2 into γ so that one column of $W_2 H$ has the largest γ in the range space of W_2. With the maximum γ, the solution, $\tilde{b}_{TLS} = \frac{1}{\gamma}[\bar{x}]$, is minimized for any norm (18:892). By definition, a readily available Householder matrix can be computed to accomplish this task. Additionally, since only the last column of $W_2 H$ is of interest, only a portion of H is necessary for the transformation. (57:1443).

In review, with the SVD and a minimum norm solution technique, it is easy to incorporate total least squares into the overmodeled LP methodology. We refer to
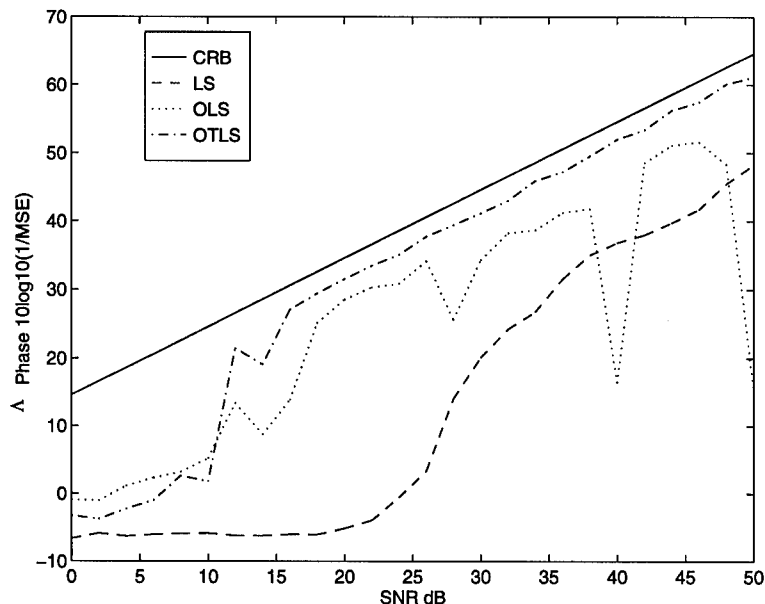
51

Figure 9. The λ_1 phase inverse MSE of the OTLS, OLS, and LS estimators over a range of SNRs. Also plotted is the CRB.

the improved algorithm as overmodeled total least squares (OTLS). In the algorithm, after the overmodeled LP coefficients are estimated in a total least squares sense, the energy sort is used to identify the extraneous exponentials, λ_n. After that, the pseudo-inverse is used to estimate the amplitude coefficients, c_n.

Figures 9, 10, 11, and 12 are phase inverse MSE and bias plots for comparing the performance of the OTLS estimator with the non SVD OLS estimator and the regular LS estimator derived earlier. Again, the same Monte-Carlo experiment as that for OLS, LS, and the original Prony method was accomplished. The prediction order, P, for OTLS was 12. From the inverse MSE plots, we see that the OTLS estimator almost achieves the CRB and performs consistently better than the previously derived estimators.

Despite the improved performance, there are significant drawbacks to over-modeling in general. For example, to attain the best results, the OTLS estimator
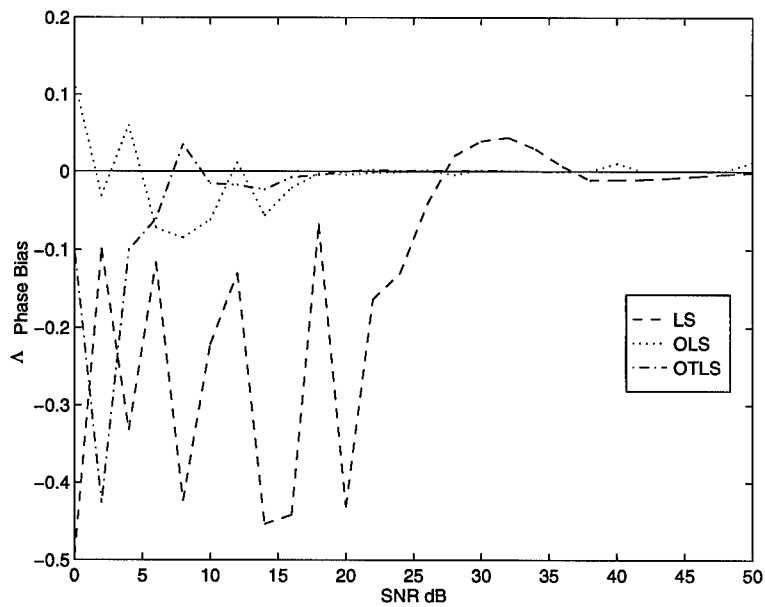
Figure 10. The λ_1 phase bias of the OTLS, OLS, and LS estimators over a range of SNRs.
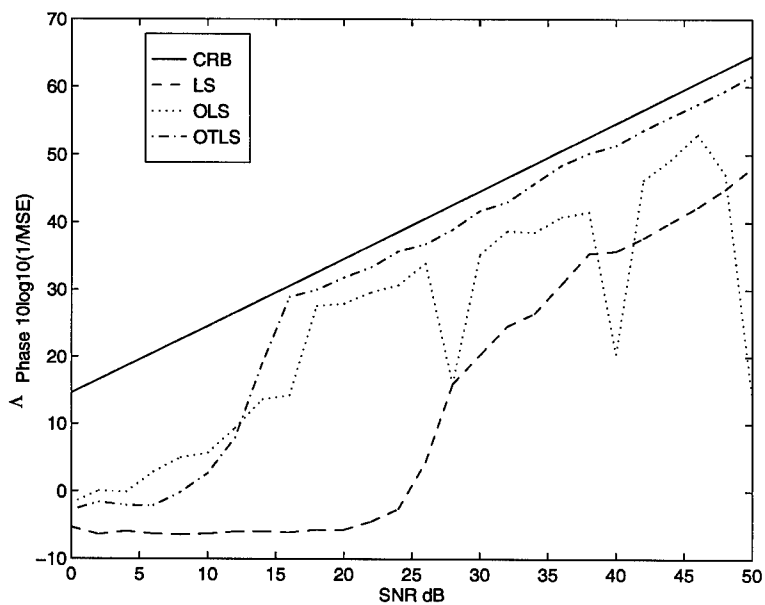
Figure 11. The λ_2 phase inverse MSE of the OTLS, OLS, and LS estimators over a range of SNRs. Also plotted is the CRB.
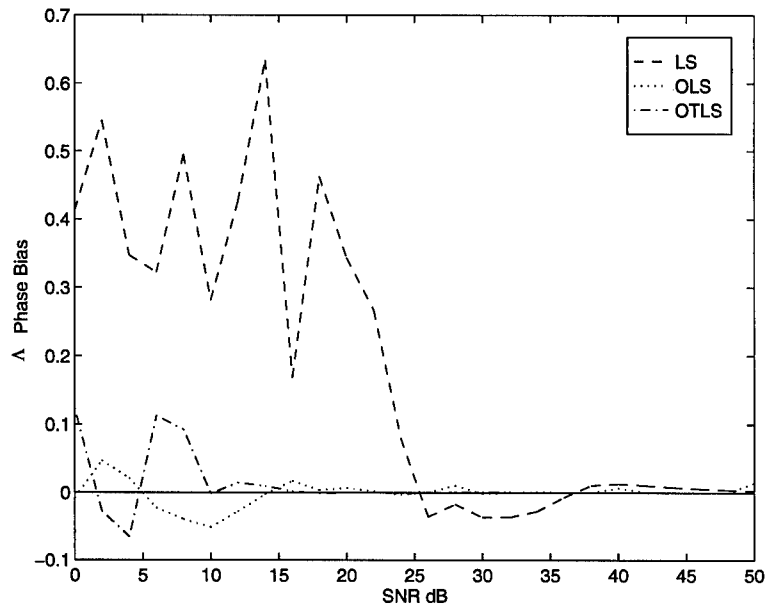
Figure 12. The λ_2 phase bias of the OTLS, OLS, and LS estimators over a range of SNRs.

was executed at various prediction orders. For the experiment, a prediction order of 12 obtained estimates nearest to the CRB. Recall that for the CRB to be plotted, knowledge of the underling signal parameters is required. With unknown parameters, the proper choice of P may be critical to the success of the algorithm and most likely will introduce another heuristic to the methodology. Also, the increased degree of the LP polynomial, brought on by the artificially high model order, is disturbing. Rooting the polynomial for the λ_n estimates is an inherently ill-conditioned numerical procedure and it is well known that rooting errors are greatly exacerbated by increasing the polynomial's degree. When these drawbacks are coupled with the fact that even under straight forward simulations, the estimator was unable to attain the CRB, we question the use of overmodeling in the superimposed exponential parameter estimation problem.

In the next section, we introduce our modification to linear prediction which is not heuristic and does not utilize overmodeling. The modification is completely couched in tractable statistics. In fact, the resulting cost function is identical to the cost function derived from maximum likelihood.

3.4   Iterative Generalized Least Squares

The development that follows is original work and extends on the LP equation introduced in the beginning of this chapter. Recall the equation

$$(123) \qquad b_0 y[m] + b_1 y[m - 1] + \ldots + b_N y[m - N] = e[m]$$

as a noisy modification to the noiseless LP equation

$$(124) \qquad b_0 s[m] + b_1 s[m - 1] + \ldots + b_N s[m - N] = 0.$$

Also, recall that

$$(125) \qquad \begin{aligned} y[m] &= s[m] + w[m] \\ \Rightarrow \quad s[m] &= y[m] - w[m]. \end{aligned}$$

If we substitute the last expression for $s[m]$ into the noiseless LP equation, we get the same noisy modified relationship as in Equation 123, only the LP equation error

$$(126) \qquad e[m] = b_0 w[m] + b_1 w[m - 1] + \ldots + b_N w[m - N].$$

Now, the LP error, $e[m]$, has known structure. It is the sum of weighted Gaussian random variables, and although each $w[m]$ is uncorrelated over all m, each $e[m]$ is correlated with the $\pm N$ adjoining elements of $e[m]$.

55

In vector notation,

$$(127) \qquad\qquad Y\bar{b} = \bar{e}.$$

With Equation 126, we know the statistics of the LP equation error vector, $\bar{e}$. Since each $w[m]$ in $e[m]$ has mean zero, the mean vector of $\bar{e}$ is $\bar{0}$. The LP error covariance matrix, $\sigma^2 R$, of $\bar{e}$ is defined as the $(M - N) \times (M - N)$ matrix

$$(128)$$
$$\sigma^2 R = \begin{bmatrix} E\{e[N]e^*[N]\} & E\{e[N]e^*[N+1]\} & \cdots & E\{e[N]e^*[M-1]\} \\ E\{e[N+1]e^*[N]\} & E\{e[N+1]e^*[N+1]\} & \cdots & E\{e[1]e^*[M-1]\} \\ \vdots & \vdots & \ddots & \vdots \\ E\{e[M-1]e^*[N]\} & E\{e[M-1]e^*[1]\} & \cdots & E\{e[M-1]e^*[M-1]\} \end{bmatrix}$$

where, from (126),

$$E\{e[m_1]e^*[m_2]\} =$$
$$(129) \qquad E\left\{ \left(\sum_{n=0}^{N} b_n w[m_1 - n] \right) \left(\sum_{n=0}^{N} b_n w[m_2 - n] \right)^* \right\}$$

and

$$(130) \qquad m_1 = N, N+1, \ldots, M-1, \qquad m_2 = N, N+1, \ldots, M-1.$$

When the product from (129) is carried out for each element of the matrix $\sigma^2 R$ and the uncorrelated assumption, $E\{w[m_1]w[m_2]\} = 0$ for $m_1 \neq m_2$, is invoked for

the cross terms, we calculate

$$
(131) \qquad R_{i,j} = \begin{cases} \sum_{k=0}^{N} b_k b_{k+j-i}^*, & i \le j \\[1mm] \sum_{k=0}^{N} b_k^* b_{k+i-j}, & i > j \\[1mm] 0, & |i-j| > N \end{cases} \qquad .
$$

Alternatively,

$$
(132) \qquad R = BB^H
$$

where, as defined in Chapter II, B is an $(M-N) \times M$ Toeplitz matrix of LP coefficients in the form

$$
(133) \qquad B = \begin{bmatrix} b_0 & b_1 & \cdots & b_N & 0 & \cdots & 0 \\ 0 & b_0 & b_1 & \cdots & b_N & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & b_0 & b_1 & \cdots & b_N \end{bmatrix} .
$$

The matrix B has full row rank $M - N$. Therefore, R has full rank $M - N$ and is positive definite. A Cholesky decomposition of R provides a nonsingular, lower triangular, $(M-N) \times (M-N)$, matrix R_l such that the LP equation error covariance matrix is factored as

$$
(134) \qquad R_l R_l^H = R.
$$

The LP coefficients in R are unknown, but for a moment assume we know R. If so, the inverse of R_l, R_l^{-1}, can be used to linearly transform the LP error vector, $\bar{e}$. A linear transformed Gaussian distributed error vector, with mean vector zero, also

has mean vector zero, and a linear transformed Gaussian distributed error vector, with covariance matrix $\sigma^2 R$, has covariance matrix $\sigma^2 R_l^{-1} R (R_l^{-1})^H$ (61:59). Because $R = R_l R_l^H$, we know

$$(135) \qquad \begin{aligned} R_l^{-1} R &= R_l^H \\ \Rightarrow \quad R_l^{-1} R (R_l^{-1})^H &= \text{I}. \end{aligned}$$

Therefore the covariance matrix of $R_l^{-1} \bar{e}$ is $\sigma^2 \text{I}$ and $R_l^{-1} \bar{e}$ is distributed

$$(136) \qquad R_l^{-1} \bar{e} \sim CN(\bar{0}, \sigma^2 \text{I}).$$

Since b_0 can be normalized to 1 without affecting the λ_n estimates, the equation

$$(137) \qquad Y \bar{b} = \bar{e}$$

can, once again, be rearrange to

$$(138) \qquad \bar{y}_0 + \tilde{Y} \tilde{b} = \bar{e}.$$

This implies

$$(139) \qquad R_l^{-1} \bar{y}_0 + R_l^{-1} \tilde{Y} \tilde{b} = R_l^{-1} \bar{e}.$$

Estimating $\tilde{b}$ in Equation 139 can be cast as a linear least squares problem. In this form, we know the unique minimum solution is obtained with the pseudo-inverse

estimator

$$
\begin{aligned}
\tilde{b} &= -\left((R_l^{-1}\tilde{Y})^H (R_l^{-1}\tilde{Y})\right)^{-1} (R_l^{-1}\tilde{Y})^H R_l^{-1}\bar{y}_0 \\
&= -\left(\tilde{Y}^H (R_l R_l^H)^{-1}\tilde{Y}\right)^{-1} \tilde{Y}^H (R_l R_l^H)^{-1}\bar{y}_0 \\
&= -(\tilde{Y}^H R^{-1}\tilde{Y})^{-1}\tilde{Y}^H R^{-1}\bar{y}_0.
\end{aligned}
$$

(140)

Recall from Chapter II, since $R_l^{-1}\bar{e}$ is Gaussian distributed with mean vector zero and covariance matrix $\sigma^2 I$, the pseudo-inverse estimator is also the MVU estimator. Linearly transforming a least squares problem to decorrelate the error vector is known as generalized least squares (GLS) (64:60–64).

In our application, R is unknown because $\bar{b}$ is unknown. To overcome this, we turn to an iterative generalized least squares (IGLS) approach which is similar to the IQML algorithm of Chapter II. In IGLS, previous estimates of $\bar{b}$ are used to create R for re-estimating $\bar{b}$ with the GLS estimator (140). The commonality of our IGLS algorithm and the IQML algorithm is in the cost function they both attempt to minimize. Since

(141)
$$
Y\bar{b} = \bar{e},
$$

we know

(142)
$$
R_l^{-1}Y\bar{b} = R_l^{-1}\bar{e}.
$$

In the least squares problem, we wish to minimize the cost function

$$
\begin{aligned}
J(\bar{b}) &= \left(R_l^{-1}\bar{e}\right)^H \left(R_l^{-1}\bar{e}\right) \\
&= \left(R_l^{-1}Y\bar{b}\right)^H \left(R_l^{-1}Y\bar{b}\right) \\
&= \bar{b}^H Y^H \left(R_l R_l^H\right)^{-1} Y\bar{b} \\
&= \bar{b}^H Y^H (R)^{-1} Y\bar{b} \\
(143) \qquad &= \bar{b}^H Y^H (BB^H)^{-1} Y\bar{b}.
\end{aligned}
$$

This expression for $J(\bar{b})$ is identical to the maximum likelihood cost function $L(\bar{b})$ for the superimposed exponential problem. Attempting to minimize $L(\bar{b})$ by using a prior estimate of $\bar{b}$ to fix $(BB^H)^{-1}$ and then solving the remaining quadratic problem, is similar to using a prior estimate of $\bar{b}$ to fix R_l^{-1} and then solving the generalized least squares problem. Like IQML, we typically initialize IGLS with a least squares estimate. We discuss, in detail, the intricacies of minimizing the cost function in Chapter IV. Figures 13, 14, 15, and 16 are phase inverse MSE and bias plots for comparing the performance of the IGLS algorithm with the OTLS estimator and the non SVD OLS estimator derived earlier. Again, the same Monte-Carlo experiment as that for OTLS, OLS, LS, and the original Prony method was accomplished. The number of IGLS iterations utilized in every execution was 10. From the figures, we see that the IGLS algorithm shares a common noise threshold with the OTLS estimator, but above the noise threshold, the IGLS algorithm attains the CRB. The performance of IGLS completely corroborates the development of this chapter and the maximum likelihood chapter.

Figure 13.   The λ_1 phase inverse MSE of the IGLS, OTLS, and OLS estimators over a range of SNRs. Also plotted is the CRB.

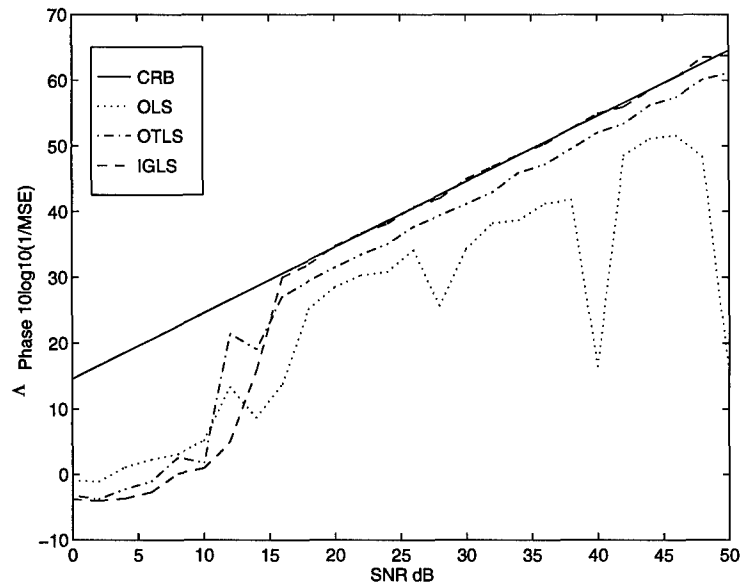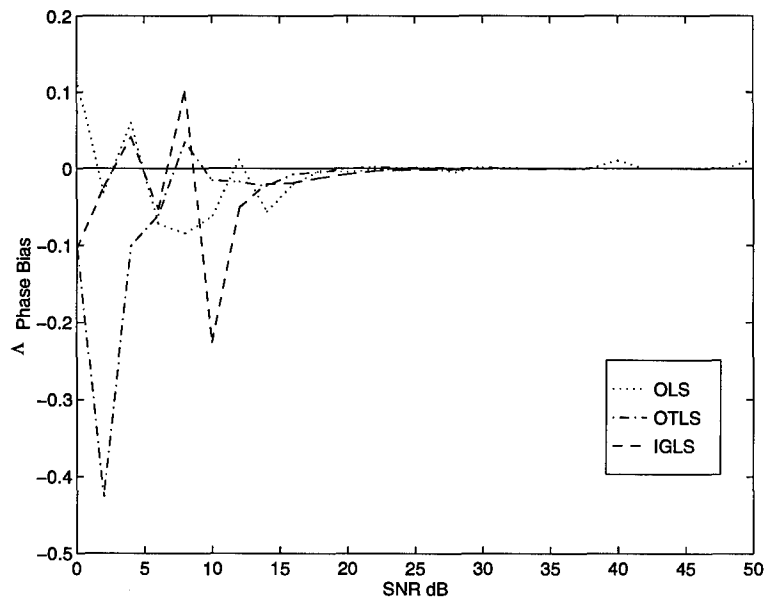Figure 14.   The λ_1 phase bias of the IGLS, OTLS, and OLS estimators over a range of SNRs.

61

Figure 15. The λ_2 phase inverse MSE of the IGLS, OTLS, and OLS estimators over a range of SNRs. Also plotted is the CRB.

Figure 16. The λ_2 phase bias of the IGLS, OTLS, and OLS estimators over a range of SNRs.
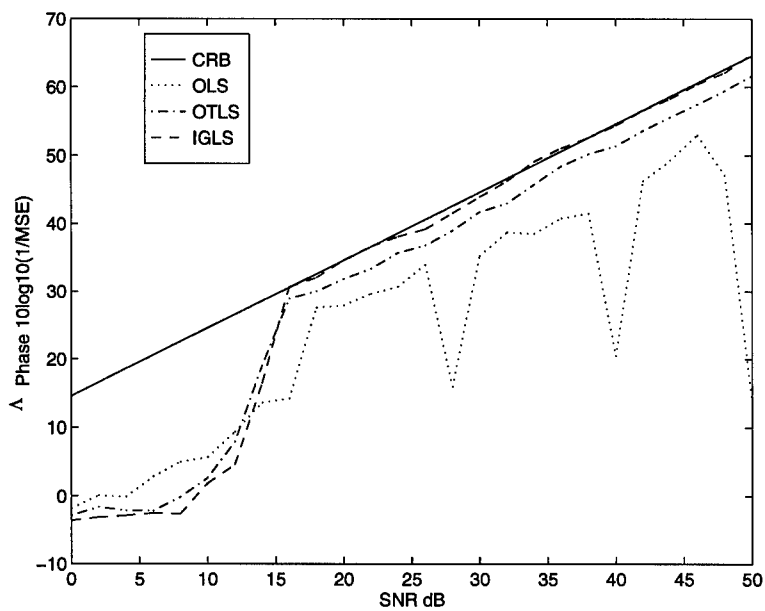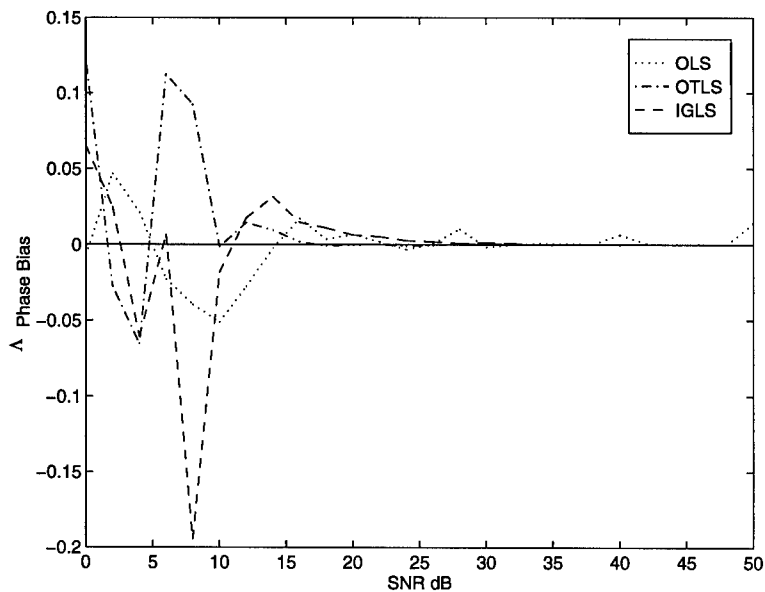
62

3.5 Chapter Conclusion

In this chapter, the methodology of linear prediction based parameter estimation for the superimposed exponential problem was developed from its earliest beginning. The introduction of least squares to linear prediction was explained, and then the heuristic path of overmodeling was examined. We concluded that although overmodeling initially improved linear prediction, it ultimately inhibited linear prediction from attaining truly maximum likelihood quality performance. To address this problem, we introduced a methodology that demonstrated how the noiseless LP equation can be modified to function in the noise corrupted scenario without heuristics. The result is a statistically sound, tractable, LP based algorithm for the parameter estimation problem with maximum likelihood performance. We call the algorithm iterative generalized least squares.

We concede IGLS differs little from the algorithm developed with maximum likelihood, but we insist our development, from linear prediction, is valuable. Maximum likelihood built its case at the expense of linear prediction (4:1081). Undoubtedly, this tarnished all previous linear prediction efforts. By giving linear prediction the capability to possibly obtain the MVU estimator, we allow the principles of the LP methodology to be maintained and carried to new applications. For some applications, it is more intuitive to construct an estimator with our methodology than with the maximum likelihood methodology. The application motivating our research—deep level transient spectroscopy—is in this category. In Chapter VI, we demonstrate how our linear prediction based methodology facilitates an estimator for DLTS and how deriving the same estimator from maximum likelihood is difficult. Before doing so, the next two chapters are used to analyze the details of implementing IGLS.

IV. Cost Function Minimization

4.1 Chapter Introduction

In Chapter II, we identified the exact maximum likelihood (ML) cost function for superimposed exponential parameter estimation. In the same chapter, we outlined an iterative quadratic minimization algorithm for minimizing the ML cost function. The algorithm is known as iterative quadratic maximum likelihood (IQML). In Chapter III, we developed our own linear prediction (LP) based estimation algorithm. The algorithm ultimately attempts to minimize the same ML cost function. We called it iterative generalized least squares (IGLS). In this chapter, we further analyze and compare the IQML and IGLS algorithms. We also introduce a similar total least squares (TLS) based algorithm for minimizing the ML cost function. The iterative total least squares (ITLS) algorithm is an original contribution of this dissertation. By introducing ITLS, we can address TLS without the heuristic of over-modeling and—by comparing IQML, IGLS, and ITLS—bring considerable insight to the minimizing task.

Before continuing, we need to re-emphasize an important issue. Despite the effectiveness of the IQML, IGLS, and ITLS algorithms, they do not directly minimize the ML cost function. Even in its simplified—LP coefficient based—form, the ML cost function,

$$(144) \qquad L(\bar{b}) = \bar{b}^H Y^H (BB^H)^{-1} Y \bar{b},$$

is nonlinear. When a previous estimate of $\bar{b}$ is used to calculate and fix $(BB^H)^{-1}$, the cost function has been "linearized". All three algorithms are variations of this assumption. They are not guaranteed to converge, and even if they converge, they

64

are not guaranteed to obtain the global minimum (38:1218). Excellent empirical performance is our best measurement for acceptance.

4.2 Iterative Quadratic Maximum Likelihood

When $(BB^H)^{-1}$ is calculated with a previous estimate of $\bar{b}$, the expression

$$(145) \qquad\qquad Y^H(BB^H)^{-1}Y$$

is Hermitian, and

$$(146) \qquad\qquad \bar{b}^H Y^H(BB^H)^{-1}Y\bar{b}$$

has quadratic form. To avoid the trivial solution while minimizing Equation 146, the vector $\bar{b}$ must be constrained. One way to constrain $\bar{b}$ is to insist $\bar{b}^H\bar{b} = 1$ or $\|\bar{b}\|_2^2 = 1$. Because we eventually root the LP polynomial formed from $\bar{b}$, we are not concerned with the actual values of b_n in $\bar{b}$, but rather the ratios between them. Therefore, the nontriviality constraint $\bar{b}^H\bar{b} = 1$ is reasonable.

Before explaining how to minimize the ML cost function with the $\bar{b}^H\bar{b} = 1$ constraint, we need to note that in the past, researchers have minimized the cost function with a different nontriviality constraint, *e.g.* (39, 4, 40, 8). Because recent implementations of IQML, (52, 26), utilize the $\bar{b}^H\bar{b} = 1$ nontriviality constraint, we are associating the $\bar{b}^H\bar{b} = 1$ constraint with IQML for this dissertation.

With this established, note $\bar{b}^H\bar{b}$ is a real scalar function of $\bar{b}$, and

$$(147) \qquad\qquad \bar{b}^H Y^H(BB^H)^{-1}Y\bar{b}$$

is also a real scalar function of $\bar{b}$, we can introduce a real Lagrange multiplier, λ, and recast this problem to one of minimizing the Lagrangian (70:691–694)

$$(148) \qquad \mathbb{L}(\bar{b}, \lambda) = \bar{b}^H Y^H (BB^H)^{-1} Y\bar{b} + \lambda \mathbb{C}(\bar{b})$$

where

$$(149) \qquad \mathbb{C}(\bar{b}) = 1 - \bar{b}^H \bar{b}.$$

At a stationary point, we expect

$$(150) \qquad \frac{\partial \mathbb{L}(\bar{b}, \lambda)}{\partial \bar{b}^*} = \bar{0}$$

and

$$(151) \qquad \frac{\partial \mathbb{L}(\bar{b}, \lambda)}{\partial \lambda} = 0.$$

From Appendix B, we know

$$(152) \qquad \frac{\partial \mathbb{L}(\bar{b}, \lambda)}{\partial \bar{b}^*} = Y^H (BB^H)^{-1} Y\bar{b} - \lambda\bar{b},$$

and by definition, we know

$$(153) \qquad \frac{\partial \mathbb{L}(\bar{b}, \lambda)}{\partial \lambda} = 1 - \bar{b}^H \bar{b}.$$

Therefore, at a stationary point,

$$(154) \qquad Y^H (BB^H)^{-1} Y\bar{b} = \lambda\bar{b}$$

and

(155)
$$\bar{b}^H \bar{b} = 1.$$

Equation 155 is the desired constraint, and Equation 154 is an eigenvalue-eigenvector expression. The Lagrange multiplier, λ, is an eigenvalue of $Y^H(BB^H)^{-1}Y$. Therefore, the eigenvector, $\bar{b}$, associated with the minimum eigenvalue, λ_n, from the eigenvalue decomposition of $Y^H(BB^H)^{-1}Y$ is the unique minimizing solution for $\bar{b}$.

The IQML algorithm is initialized with a previous estimate of $\bar{b}$ (typically from a least squares solution). With that estimate, $Y^H(BB^H)^{-1}Y$ is calculated and eigenvalue decomposed. From the decomposition, the minimizing eigenvector is selected as the next estimate of $\bar{b}$. The process is continued until an iteration limit or minimum estimate change limit is met. Maximum estimate improvement usually occurs in the first few iterations. When the signal consists of pure sinusoids in noise, as is the case in radar applications, the algorithm can be initialized with an estimate obtained from a spectral analysis of the signal.

4.3 Iterative Total Least Squares

At this time, we introduce our iterative total least squares approach for minimizing the ML cost function because it is analytically equivalent to IQML. As in Chapter III, let

(156)
$$R_l R_l^H = R = BB^H$$

where R_l is the nonsingular, lower triangular, $(M - N) \times (M - N)$, Cholesky decomposition of R. Now we can re-express the ML cost function as

$$
\begin{aligned}
L(\bar{b}) &= \bar{b}^H Y^H (BB^H)^{-1} Y\bar{b} \\
&= \bar{b}^H Y^H (R_l R_l^H)^{-1} Y\bar{b} \\
&= \bar{b}^H Y^H (R_l^{-1})^H R_l^{-1} Y\bar{b} \\
&= \|R_l^{-1} Y\bar{b}\|_2^2.
\end{aligned}
$$

(157)

Recall from the singular value decomposition (SVD), there exists unitary matrices U and W, and a diagonal matrix Σ, such that

$$
Y = U\Sigma W^H.
$$

(158)

Therefore,

$$
\begin{aligned}
Y^H Y &= (U\Sigma W^H)^H U\Sigma W^H \\
&= W\Sigma U^H U\Sigma W^H \\
&= W\Sigma\Sigma W^H.
\end{aligned}
$$

(159)

This implies

$$
Y^H Y \bar{w}_n = \sigma_n^2 \bar{w}_n
$$

(160)

Using the SVD to decompose $R_l^{-1} Y$ from Equation 157, let

$$
R_l^{-1} Y = U\Sigma W^H
$$

(161)

so that

$$
\begin{aligned}
(R_l^{-1}Y)^H(R_l^{-1}Y)\bar{w}_n &= \left(Y^H(R_l^{-1})^H R_l^{-1}Y\right)\bar{w}_n \\
&= Y^H(BB^H)^{-1}Y\bar{w}_n \\
&= \sigma_n^2 \bar{w}_n.
\end{aligned}
$$

(162)

This relationship is exactly the same as the eigenvalue-eigenvector relationship, Equation 154, developed in the previous IQML section. Therefore, through the Lagrangian, we know the singular vector $\bar{w}_n$ associated with the minimum singular value, σ_n, is the minimizing solution for $\bar{b}$. In the vein of TLS, we say that the minimum perturbation of $R_l^{-1}Y$—that reduces its full column rank by one and allows the right most singular vector of W into the null space—occurs when a low rank approximation of $R_l^{-1}Y$ is created by zeroing the smallest singular value of its SVD (51). Therefore, the singular vector $\bar{w}_{N+1}$ of $R_l^{-1}Y$ is equal to the eigenvector of $Y^H(BB^H)^{-1}Y$ associated with the smallest eigenvalue, and is also the minimizing solution of the linearized ML cost function. Consequently, the ITLS algorithm is identical to the IQML algorithm except that the eigenvalue decomposition of $Y^H(BB^H)^{-1}Y$ is replaced with the singular value decomposition of $R_l^{-1}Y$.

The ITLS and IQML equivalence was demonstrated to add insight to the $\bar{b}^H\bar{b} = 1$ constrained approach to minimization. For example, researchers in TLS have developed a closed form expression for $\bar{b}$ based on the assumption that b_0 can be normalized after solving for $\bar{b}$ (28:36–37). With the demonstrated equivalence, this insight applies to IQML and ITLS as well. Assume, after normalization, $\bar{b} = [\ 1\ \ \tilde{b}^T\ ]^T$. Recall that when $\bar{b}$ is the minimizing solution,

$$
(R_l^{-1}Y)^H(R_l^{-1}Y)\bar{b} = \sigma_{N+1}^2\bar{b}.
$$

(163)

Therefore,

$$
\left[R_l^{-1} \bar{y}_0; R_l^{-1} \tilde{Y} \right]^H \left[R_l^{-1} \bar{y}_0; R_l^{-1} \tilde{Y} \right] \begin{bmatrix} 1 \\ \tilde{b} \end{bmatrix} = \sigma_{N+1}^2 \begin{bmatrix} 1 \\ \tilde{b} \end{bmatrix}
$$

(164)

$$
\Rightarrow \begin{bmatrix} (R_l^{-1} \bar{y}_0)^H (R_l^{-1} \bar{y}_0) & (R_l^{-1} \bar{y}_0)^H (R_l^{-1} \tilde{Y}) \\ (R_l^{-1} \tilde{Y})^H (R_l^{-1} \bar{y}_0) & (R_l^{-1} \tilde{Y})^H (R_l^{-1} \tilde{Y}) \end{bmatrix} \begin{bmatrix} 1 \\ \tilde{b} \end{bmatrix} = \sigma_{N+1}^2 \begin{bmatrix} 1 \\ \tilde{b} \end{bmatrix}.
$$

From the bottom row of Equation 164,

$$
(R_l^{-1} \tilde{Y})^H (R_l^{-1} \bar{y}_0) + (R_l^{-1} \tilde{Y})^H (R_l^{-1} \tilde{Y}) \tilde{b} = \sigma_{N+1}^2 \tilde{b}
$$

(165)

$$
\Rightarrow \tilde{b} = - \left((R_l^{-1} \tilde{Y})^H (R_l^{-1} \tilde{Y}) - \sigma_{N+1}^2 I \right)^{-1} (R_l^{-1} \tilde{Y})^H R_l^{-1} \bar{y}_0
$$

$$
\Rightarrow \tilde{b} = - \left(\tilde{Y}^H R^{-1} \tilde{Y} - \sigma_{N+1}^2 I \right)^{-1} \tilde{Y}^H R^{-1} \bar{y}_0.
$$

Only the smallest singular value of $R_l^{-1} Y$ is necessary to express $\tilde{b}$ in a closed form solution.

Notice the closed form solution is not used in either the IQML or the ITLS algorithms. It was developed here to provide insight in the algorithm analysis section at the end of this chapter. Also note, $b_0 = 1$ is not constrained when determining the smallest singular value. It is only a result after assuming b_0 of the solution $\tilde{b}$ can be normalized. In the next section, $b_0 = 1$ is a constraint for minimization.

4.4 Iterative Generalized Least Squares

In Chapter III, we developed the IGLS algorithm from linear prediction. For continuity with this chapter, we recast the IGLS algorithm as a development from

70

the ML cost function. Recall from the previous section that

$$L(\bar{b}) = \bar{b}^H Y^H (BB^H)^{-1} Y \bar{b}$$

(166)
$$= \|R_l^{-1} Y \bar{b}\|_2^2$$

where $R_l R_l^H = BB^H$. For IGLS, a previous estimate of $\bar{b}$ is also utilized to calculate and fix R_l, but, instead of the $\bar{b}^H \bar{b} = 1$ nontriviality constraint, the b_0 of $\bar{b}$ is constrained to be one. With this constraint, we have the familiar relationship

$$(167) \qquad R_l^{-1} Y \bar{b} = R_l^{-1} \bar{y}_0 + R_l^{-1} \tilde{Y} \tilde{b} = R_l^{-1} \bar{e}$$

for minimizing the linearized ML cost function. In Chapter III, we showed that the pseudo-inverse estimator

$$\tilde{b} = -\left((R_l^{-1} \tilde{Y})^H (R_l^{-1} \tilde{Y})\right)^{-1} (R_l^{-1} \tilde{Y})^H R_l^{-1} \bar{y}_0$$

(168)
$$= -(\tilde{Y}^H R^{-1} \tilde{Y})^{-1} \tilde{Y}^H R^{-1} \bar{y}_0.$$

is the unique minimum solution for the sum of the squared error, $(R_l^{-1} \bar{e})^H (R_l^{-1} \bar{e})$, and that the R_l^{-1} pre-multiply has a decorrelating affect on the LP equation error vector, $\bar{e}$. In Chapter II, we showed that the pseudo-inverse estimator is the minimum variance unbiased (MVU) estimator for uncorrelated processes.

What naturally follows, is an iterative algorithm where the previous estimate of $\bar{b}$ is used to create R_l^{-1} for a new estimate of $\bar{b}$ with the pseudo-inverse estimator. Each iteration is known as a generalized least squares solution and, in theory, as each $\bar{b}$ estimate improves, a better decorrelating matrix, R_l^{-1}, is created. Like IQML

71

and ITLS, the algorithm is typically initialized with a least squares (LS) estimate for $\bar{b}$, and maximum estimate improvement usually occurs in the first few iterations.

We have now developed three similar algorithms for attempting to minimizing the ML cost function—IQML, ITLS, and IGLS. In the next section, we analyze the performance of each of the algorithms.

4.5 Algorithm Analysis

A Monte-Carlo experiment was conducted with the same parameters used throughout Chapter III. Two hundred, 25 point, noisy, data vectors were created—with underlying parameters $c_1 = c_2 = 1$, $\lambda_1 = e^{j2\pi(.20)}$, and $\lambda_1 = e^{j2\pi(.22)}$—for every 2 dB of SNR between 0 dB and 50 dB. At each SNR, 10 iterations of each algorithm were performed on each realization. In every case, the algorithms were initialized with a least squares estimate.

Figures 17, 18, 19, and 20 are phase inverse MSE and bias plots for comparing the performance of the three algorithms. As expected, every plot from the ITLS algorithm is identical to that of the IQML algorithm. From this, we conclude that numerical differences between the singular value approach of ITLS and the eigenvalue approach of IQML can not be visualized. On the other hand, a difference between the IGLS algorithm and the other two algorithms can be visualized. We note that the IGLS algorithm, with its $b_0 = 1$ nontriviality constraint, obtains slightly better performance than the ITLS and IQML algorithms, with their $\bar{b}^H \bar{b} = 1$ nontriviality constraint.

The potential for better performance with the $b_0 = 1$ constraint was identified in Golub and VanLoan's introductory TLS article (18:888-890), and complete studies of the subject have been accomplished by VanHuffel and Vandewalle (29) and

Figure 17. The λ_1 phase inverse MSE of the IGLS, ITLS, and IQML algorithms over a range of SNRs. Also plotted is the CRB.

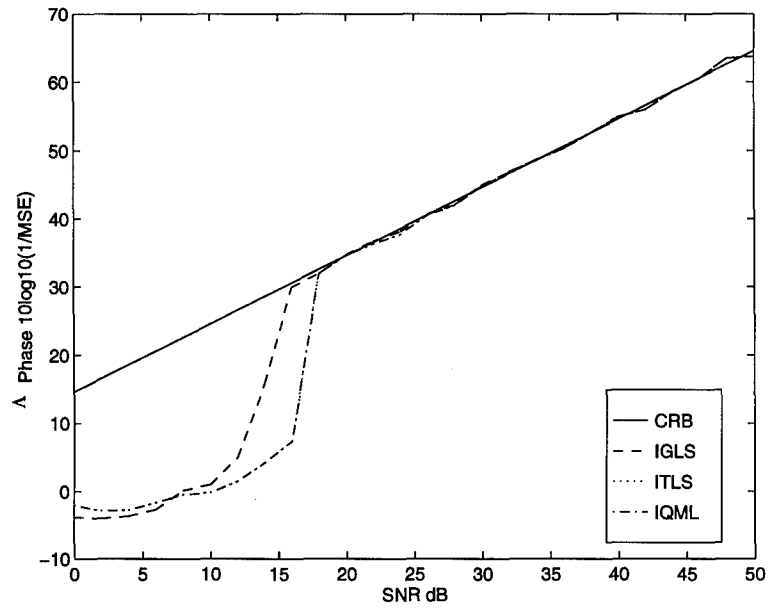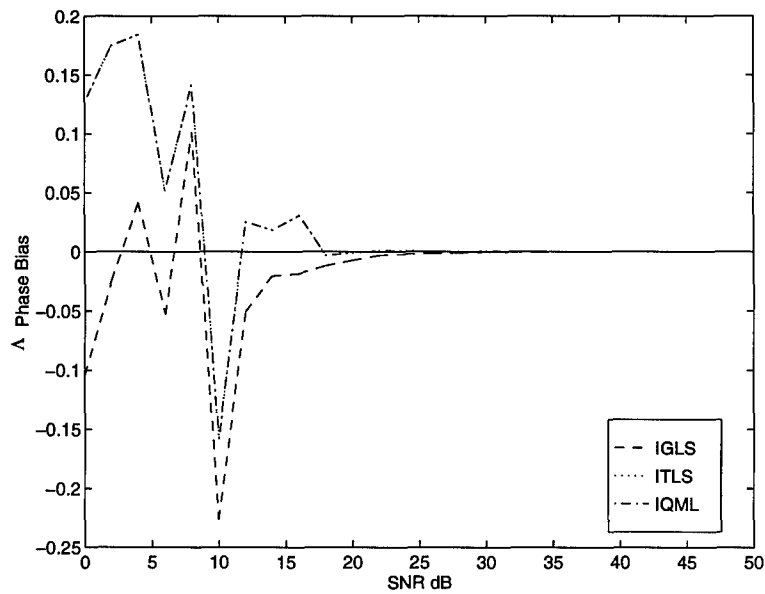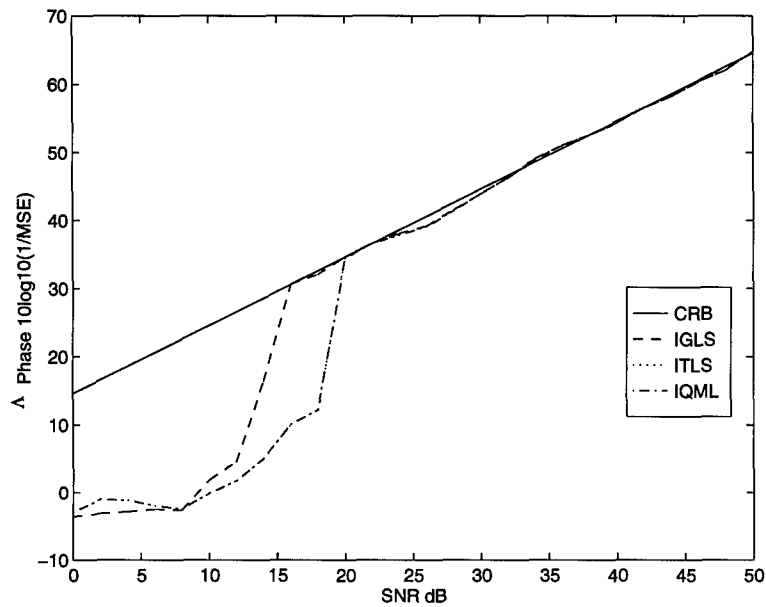Figure 18. The λ_1 phase bias of the IGLS, ITLS, and IQML algorithms over a range of SNRs.

73

Figure 19. The λ_2 phase inverse MSE of the IGLS, ITLS, and IQML algorithms over a range of SNRs. Also plotted is the CRB.
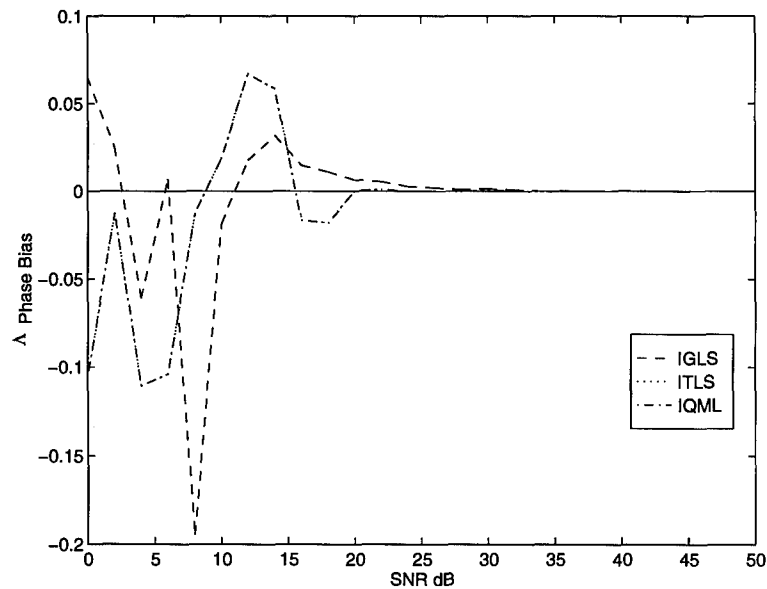

Figure 20. The λ_2 phase bias of the IGLS, ITLS, and IQML algorithms over a range of SNRs.

74

Stewart (68). Analytically, the difference between the IGLS algorithm and the other two algorithms can be seen in the IGLS pseudo-inverse estimator

$$(169) \qquad \tilde{b}_{IGLS} = -(\tilde{Y}^H R^{-1} \tilde{Y})^{-1} \tilde{Y}^H R^{-1} \tilde{y}_0,$$

and the closed form expression for $\tilde{b}$ developed in the ITLS section

$$(170) \qquad \tilde{b}_{ITLS/IQML} = - \left(\tilde{Y}^H R^{-1} \tilde{Y} - \sigma_{N+1}^2 I \right)^{-1} \tilde{Y}^H R^{-1} \tilde{y}_0.$$

The subtraction of $\sigma_{N+1}^2 I$ is the only difference between the three estimators. The ramification of the $\sigma_{N+1}^2 I$ subtraction is now considered. Recall from the noiseless LP equation,

$$(171) \qquad \bar{s}_0 + \tilde{S}\tilde{b} = \bar{0}$$

or

$$(172) \qquad S\bar{b} = \bar{0}.$$

In Appendix A we show that the rank of S is N. Therefore, the smallest singular value of S, σ_{N+1}, is 0. Equation 171 is said to be consistent because

$$(173) \qquad \tilde{S}\tilde{b} = -\bar{s}_0.$$

From the estimators in Equations 169 and 170, we see that if $\sigma_{N+1} = 0$, then $\tilde{b}_{IGLS} = \tilde{b}_{ITLS/IQML}$. When noise is considered,

$$(174) \qquad \bar{y}_0 + \tilde{Y}\tilde{b} \neq \bar{0}.$$

75

The smallest singular value of Y is not zero, and in fact, as SNR decreases, σ_{N+1} increases. The linear problem is said to be conflicting and when σ_{N+1} is large, the problem is highly conflicting. When the closed form expression for $\tilde{b}_{ITLS/IQML}$ is considered for a highly conflicting problem, the $\sigma_{N+1}^2 I$ subtraction on the diagonal of the matrix $\tilde{Y}^H R^{-1} \tilde{Y}$ is significant. This is an ill-conditioning effect on a matrix that is to be inverted. Golub and VanLoan call it deregularization and note that the condition of a TLS problem is always worse than the condition of the corresponding LS problem (18:889). Stewart in (68) shows that least squares estimates and total least squares estimates are equivalent within second order terms, and he asserts there is no reason to prefer using TLS because of the potential ill-conditioning. In our application, we know σ_{N+1} is large when SNR is low. Therefore, we actually discourage the use of TLS related algorithms like ITLS and IQML.

With regards to the overmodeled total least squares (OTLS) algorithm developed in Chapter III, we disagree with the claim that the slightly improved performance of OTLS, over the other overmodeled approaches, is due to an increase in perturbation capability (57, 27, 13, 1, 6). Given the insight we have gained in this chapter, the most likely reason for improved performance in OTLS is the readily available minimum norm solution for the near rank deficient structure artificially induced by overmodeling.

Returning to the task of minimizing the exact ML cost function, we must recognize a case where the $b_0 = 1$ nontriviality constraint is inappropriate. Researchers focused on estimating the parameters of exponential constrained to the unit circle (sinusoids) have found the performance of IQML can be significantly improved by placing an additional constraint, other than nontriviality, on the vector $\bar{b}$ (39, 4, 52, 66). It was recognized in (39) that for the roots of $\bar{b}$ to lie on the unit

circle, the LP coefficients of $\bar{b}$ must posses conjugate symmetry

$$(175) \qquad\qquad b_k = b^*_{N-k}, \qquad k = 0, 1, ..., N.$$

The symmetry constraint can be implemented in various ways, but it was noted by Nagesha and Kay that the $b_0 = 1$ nontriviality constraint can conflict with the symmetry constraint (52). The conflict arises because constraining b_0 to one also constrains the imaginary component of b_0 to zero. A symmetry constraint with the $b_0 = 1$ constraint can be too restrictive. However, when the symmetry constraint is applied with the $\bar{b}^H \bar{b} = 1$ constraint, no assumption about b_0 is made, and the IQML algorithm can accurately estimate the parameters of any two closely spaced sinusoids at SNRs below 10 dB (52). Accurate estimates at those SNRs are 5 to 10 dB below the noise threshold of the IGLS algorithm with only the $b_0 = 1$ constraint.

Unfortunately, our research is not limited to exponentials on the unit circle. Therefore, we utilize the $b_0 = 1$ nontriviality constraint implemented with the IGLS algorithm.

4.6   Chapter Conclusion

In this chapter, we considered the details of minimizing the exact ML cost function with the IQML and IGLS algorithms. We separated the two approaches as one that applies the $\bar{b}^H \bar{b} = 1$ nontriviality constraint (IQML) and one that applies the $b_0 = 1$ nontriviality constraint (IGLS). We also introduced our own ITLS algorithm and demonstrated its equivalence to IQML. Because of the closed form solution of TLS, we were able to bring considerable insight to the minimizing task with the ITLS algorithm.

77

We note, in signal processing, much has been written about the $\bar{b}^H \bar{b} = 1$ and the $b_0 = 1$ nontriviality constraint, *e.g.*, (39, 4, 40, 52). In the most recent correspondence, Nagesha and Kay conclude that the physical significance of the coefficients of $\bar{b}$ usually suggest an appropriate choice (52). Because of the potential ill-conditioning associated with the $\bar{b}^H \bar{b} = 1$ constraint, we strengthen the conclusion by saying, the $b_0 = 1$ constraint should always be used unless the physical significance of the coefficients dictates otherwise.

V. Real Exponentials

5.1 Chapter Introduction

From Chapter II we know the maximum likelihood solution is attained when the cost function

$$
\begin{aligned}
L(V) &= \bar{y}^H(\mathrm{I} - V(V^HV)^{-1}V^H)\bar{y} \\
&= \bar{y}^H\bar{y} - \bar{y}^H V(V^HV)^{-1}V^H\bar{y}
\end{aligned}
\tag{176}
$$

is minimized. For the particular scenario of parameter estimation with signals consisting of a sum of *real* exponentials, we now elaborate further. First, the Hermitian transpose, ^{H}, can be replaced with the traditional transpose, ^{T}. Also, because $\bar{y}^H\bar{y}$ is not a function of V, the Maximum Likelihood solution can be attained by finding the V that maximizes

$$
L(V) = \bar{y}^T V(V^TV)^{-1}V^T\bar{y}.
\tag{177}
$$

Additionally, we restrict ourselves to the study of decaying signals (transients). This implies that "stable" exponentials, λ_n, are considered only on the open interval $(0,1)$. Exponentials with $\lambda_n = 1$ (constants), are addressed in the deep level transient spectroscopy chapter of this dissertation. Finally, the amplitude coefficients, c_n, must also be real, but may be negative as well as positive.

With a building block approach in mind, we begin this chapter with an in depth analysis of maximum likelihood estimation for "simple" signals comprised of one real exponential. We then extend to the important case of two real exponentials, and conclude with the case of, N, real exponentials. For comparison with iterative

generalized least squares (IGLS), we develop maximum likelihood algorithms, unique to the specific cases, that realize a slightly more accurate ML solution. Ultimately though, we conclude that the IGLS algorithm is a very robust and efficient estimator for the superimposed real exponentials problem.

The analysis that follows is original work. Additionally, the maximum likelihood algorithms developed for the one and two mode cases can not be found in the literature. They contribute to signal processing as more appropriate maximum likelihood estimators when the real exponential problem can be scoped to a specific case, and, more importantly, they serve as an excellent benchmark for verification of the multi-mode IGLS algorithm.

5.2   One Real Exponential

For the one mode problem, $N = 1$, the signal model is

$$(178) \qquad\qquad \bar{y} = \bar{v}(\lambda_1)\bar{c}_1 + \bar{w}$$

where

$$(179) \qquad\qquad \bar{v}(\lambda_1) = \left[\begin{array}{ccccc} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{M-1} \end{array} \right]^T$$

is appropriate. Therefore, the cost function, L, is explicitly

$$
\begin{aligned}
L(\lambda) &= \bar{y}^T \bar{v}(\bar{v}^T \bar{v})^{-1}\bar{v}^T \bar{y} \\
(180) \qquad\qquad &= \frac{(\bar{y}^T \bar{v})^2}{\bar{v}^T \bar{v}}
\end{aligned}
$$

where the λ_1 reference is dropped from the $\bar{v}(\lambda_1)$ symbol.
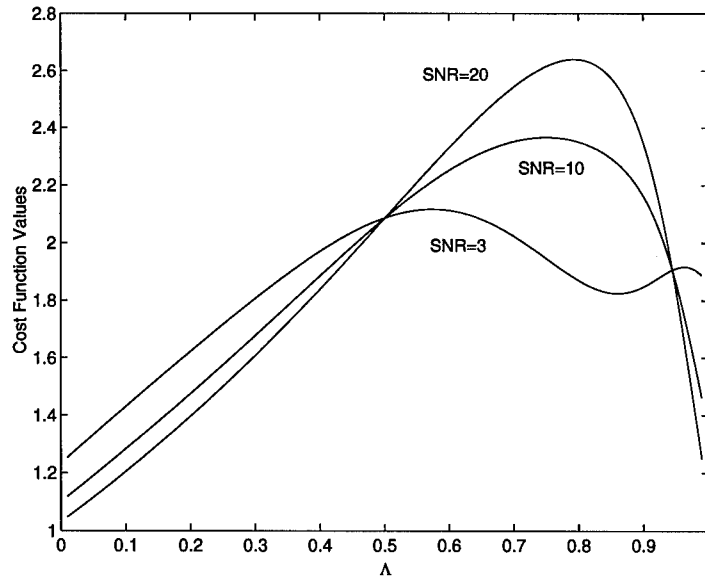
80

Figure 21. One mode real exponential cost function evaluations over the range of λ, $(0,1)$ at various SNRs. The underlying signal's true exponential is $\lambda = .8$

The behavior of the cost function L, over the range of λ, $(0,1)$, for various SNRs is displayed in Figure 21. The 25 data point observation vector, $\bar{y}$, was generated with $c = 1$, $\lambda = .8$, and the standard σ^2 definition

$$\text{(181)} \qquad\qquad \sigma^2 = \frac{|c|^2}{10^{\frac{\text{SNR}}{10}}}.$$

Recall that the noise vector $\bar{w}$ is composed of M Gaussian random variables, $w[m]$, uncorrelated across all m and identically distributed with mean zero and variance σ^2.

From the figure, it appears the cost function $L(\lambda)$ evaluated over the range of λ is unimodal until the SNR is very low. The exact SNR where $L(\lambda)$ becomes multimodal is dependent on the specific observed realization $\bar{y}$ and therefore is ultimately

random. Even though the SNR threshold for multi-modality is random, we have observed a rather small variance in its value over many realizations.

5.2.1   Unimodal Search.   Our first efforts for finding the λ that maximizes the cost function L were for the case of observations $\bar{y}$ above the multi-modal SNR threshold. Therefore, we used a unimodal search algorithm. Following the developments in Forsythe, Malcolm, and Moler (17:179–182) and Cheney and Kincaid (7:457–463), we settled on a golden section search algorithm for accomplishing this task.

By definition, any continuous unimodal function, f, over an interval $[a, b]$ has one maximum (or minimum). With two evaluations of f at a' and b', where $a < a' < b' < b$, we can determine if the maximum is in the interval $[a, b']$, if $f(a') > f(b')$, or in the interval $[a', b]$, if $f(a') < f(b')$. Iteratively, we rename the interval with the maximum, to $[a, b]$, evaluate f at two new points, a' and b', inside the new interval, and isolate the maximum to a new $[a, b']$ or $[a', b]$. At each iteration, the new search interval, called the interval of uncertainty, is reduced.

The goal of an efficient search plan is to reduce the interval of uncertainty with the minimum number of function evaluations. Intuitively, we can evaluate f at $a' = a + \frac{1}{3}(b - a)$ and $b' = b - \frac{1}{3}(b - a)$ and reduce the interval of uncertainty by $\frac{1}{3} \approx .3333$ at every iteration with two evaluations. Alternatively, we can evaluate f at $a' = a + (1 - r)(b - a)$ and $b' = b - (1 - r)(b - a)$ where

(182)
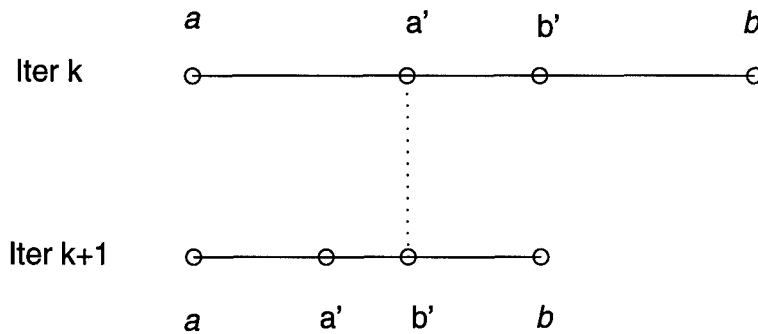$$r = \frac{-1 + \sqrt{5}}{2} \approx .6180$$

Figure 22. Iteration of the golden section search identifying the interval of uncertainty reduction and the common function evaluation between iterations. The maximum resides in the interval $[a, b']$ of iteration k.

and satisfies the quadratic equation

$$(183) \qquad\qquad r^2 + r - 1 = 0.$$

This golden section search reduces the interval of uncertainty by a factor of $(1 - r) \approx .3820$ at every iteration and eliminates one function evaluation in the next iteration. The eliminated function evaluation is inherent in the property that either $f(a')$ of the current iteration will equal $f(b')$ of the next iteration or $f(b')$ of the current iteration will equal $f(a')$ of the next iteration depending in which interval, $[a, b']$ or $[a', b]$, the maximum is determined to lie in.

Consider Figure 22, where the maximum is determined to be in the interval $[a, b']$. Let $l = b - a$ and $l' = b' - a$. In the current iteration,

$$(184) \qquad\qquad a' = a + (1 - r)l.$$

In the next iteration,

$$(185) \qquad\qquad b' = b - (1 - r)l' = a + rl'.$$

Since $l' = rl$,

$$(186) \qquad\qquad b' = a + r^2 l = a + (r - 1)l.$$

Therefore, a' of the first iteration equals b' of the next iteration.

The large reduction in the interval of uncertainty at every iteration and the eliminated function evaluation at the next iteration make this search plan close to optimal (17:182). The plan is known as the golden section search because of its association with the golden ratio, $\frac{1}{r} \approx 1.6180$. Its performance against the IGLS algorithm is illustrated in the inverse MSE and bias plots of Figures 23 and 24. For the figures, 400 Monte-Carlo realizations of $\bar{y}$ were created under the same one mode signal and noise parameters as before.

The performance of the golden section search and IGLS algorithms can be differentiated by the IGLS estimator's dependence on a good initial estimate, and its relatively complex update routine. At low SNR these factors contribute to an inferior performance when compared to the golden section search plan. However, the performance of IGLS is quite good when recalling that, unlike the one dimensional golden section search, IGLS has robust multi-exponential capability.

5.2.2 Multi-modal Search. To assess the ramifications of the unimodal assumption, an estimator, capable of determining the maximum of a multi-modal cost function in one variable, was derived. First, we began by considering the maximum
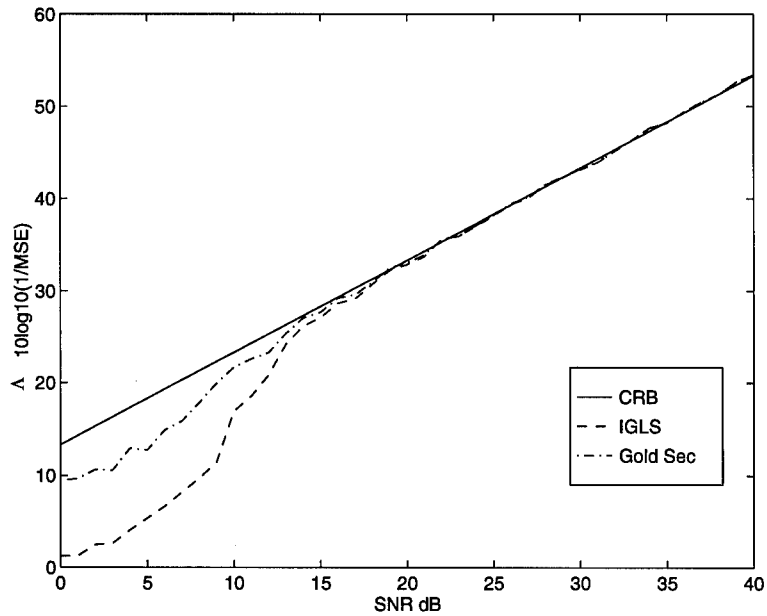
84

Figure 23. The inverse MSE of the golden section and IGLS algorithms versus the Cramer-Rao bound (CRB) over a range of SNRs. The underlying signal is of one real exponential with $\lambda = .8$
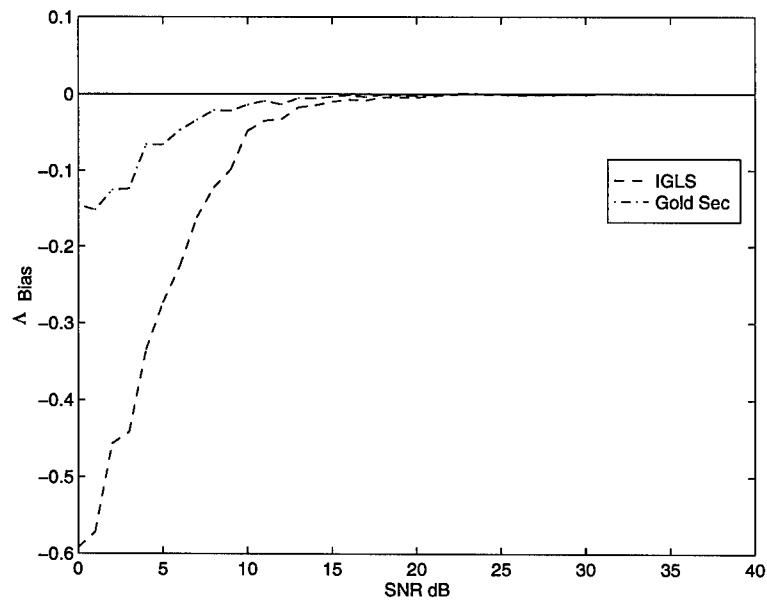

Figure 24. The bias of the golden section and IGLS algorithms over a range of SNRs. The underlying signal is of one real exponential with $\lambda = .8$

likelihood cost function in summation notation so that

$$(187) \qquad L(\lambda) = \frac{\left(\sum_{m=0}^{M-1} y[m]\lambda^m\right)^2}{\sum_{m=0}^{M-1} \lambda^{2m}}.$$

If we let c and λ represent the amplitude coefficient and exponential of the underlying signal of $y[m]$, respectively, and we let z represent the unknown exponential for estimation, we can say

$$(188) \qquad L(z; c, \lambda, M) = \frac{\left(\sum_{m=0}^{M-1}(c\lambda^m + w[m])z^m\right)^2}{\sum_{m=0}^{M-1} z^{2m}}.$$

Since

$$(189) \qquad \sum_{m=0}^{M-1} z^{2m} = \frac{1 - z^{2M}}{1 - z^2}$$

in closed form, we know

$$
\begin{aligned}
L(z; c, \lambda, M) &= \frac{(1 - z^2)\left(\sum_{m=0}^{M-1} c\lambda^m z^m + \sum_{m=0}^{M-1} w[m]z^m\right)^2}{1 - z^{2M}} \\
&= \frac{c^2(1 - z^2)\left(\sum_{m=0}^{M-1}(\lambda z)^m\right)^2}{1 - z^{2M}} + \\
&\qquad \frac{2c(1 - z^2)\sum_{m=0}^{M-1}(\lambda z)^m \sum_{m=0}^{M-1} w[m]z^m}{1 - z^{2M}} + \\
(190) &\qquad \frac{(1 - z^2)\left(\sum_{m=0}^{M-1} w[m]z^m\right)^2}{1 - z^{2M}}.
\end{aligned}
$$

Also, since

$$(191) \qquad \sum_{m=0}^{M-1}(\lambda z)^m = \frac{1 - (\lambda z)^M}{1 - \lambda z}$$

in closed form, we know

$$
L(z; c, \lambda, M) = \frac{c^2 \left(1 - (\lambda z)^M\right)^2 (1 - z^2)}{(1 - z^{2M})(1 - \lambda z)^2} +
$$

$$
\frac{2c(1 - (\lambda z)^M)(1 - z^2) \sum_{m=0}^{M-1} w[m] z^m}{(1 - z^{2M})(1 - \lambda z)} +
$$

(192)
$$
\frac{(1 - z^2) \left(\sum_{m=0}^{M-1} w[m] z^m\right)^2}{1 - z^{2M}}.
$$

Observe that the first rational function of Equation 192 exactly models the cost function of the noiseless signal, and therefore, the additional two rational functions exactly model the contribution of measurement noise in the cost function of the noisy signal, L. Our maximization plan is predicated on the fact that there exists coefficients for the numerator and denominator polynomials of the rational functions that model L. If known, these coefficients could be used with the derivative of L to find the z values that locally maximize and minimize L. The real value of z, in the interval (0,1), that yields the global maximum of L, gives the maximum likelihood solution. Unfortunately, the three rational functions that sum to L each have numerator and denominator polynomials of at least $2M$ degrees. With any significant length of signal, M, finding an analytical solution for the maximum of L is difficult. With some insight though, a rational low order function approximation of L can be created with a relatively small degree numerator polynomial and an even smaller degree denominator polynomial. To do so, we first observe the behavior of the functions $1 - (\lambda z)^M$ and $1 - z^{2M}$ as M increases. From Figures 25 and 26 we see that function evaluations, for z near $\lambda = .8$, quickly go to one as M increases. For approximating L, we assumed that by $M = 25$ both functions evaluate to one.
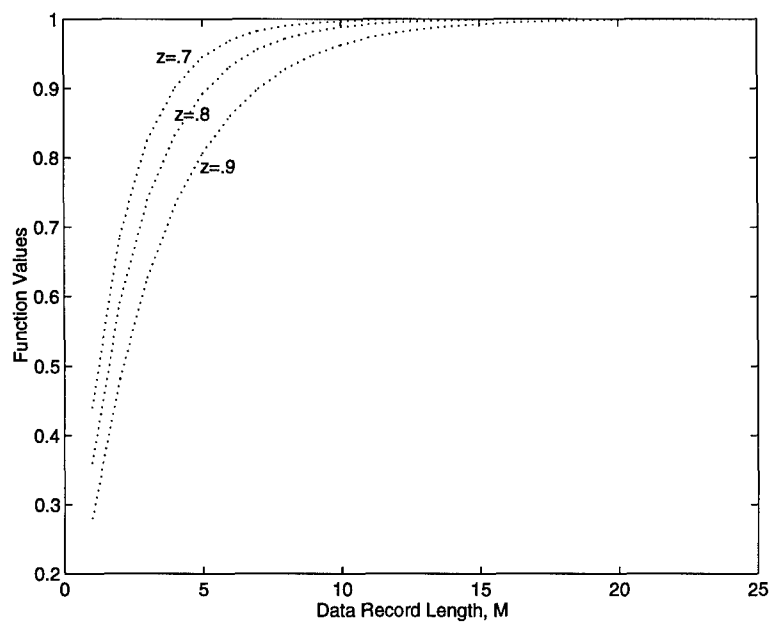
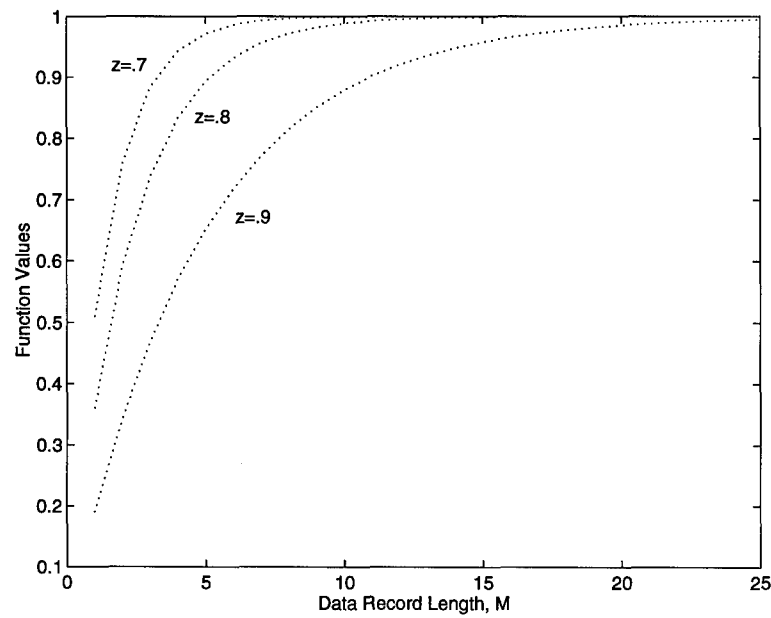Figure 25.   Evaluations of $1 - (\lambda z)^M$ for $\lambda = .8$ and $z = .7$, .8, and .9 with respect to M.

Figure 26. Evaluations of $1 - z^{2M}$ for $z = .7$, .8, and .9 with respect to M.

Therefore,

$$
L(z; c, \lambda, M) \approx \frac{c^2(1 - z^2)}{(1 - \lambda z)^2} +
$$

$$
\frac{2c(1 - z^2) \sum_{m=0}^{M-1} w[m]z^m}{(1 - \lambda z)} +
$$

$$
\frac{(1 - z^2) \left(\sum_{m=0}^{M-1} w[m]z^m \right)^2}{1}
$$

$$
\approx \frac{c^2(1 - z^2)}{1 - 2\lambda z + \lambda^2 z^2} +
$$

$$
\frac{2c(1 - \lambda z - z^2 + \lambda z^3) \sum_{m=0}^{M-1} w[m]z^m}{1 - 2\lambda z + \lambda^2 z^2} +
$$

(193)
$$
\frac{(1 - 2\lambda z + (\lambda^2 - 1)z^2 + 2\lambda z^3 - \lambda^2 z^4) \left(\sum_{m=0}^{M-1} w[m]z^m \right)^2}{1 - 2\lambda z + \lambda^2 z^2}.
$$

Again the first rational function of the model for L represents the noiseless signal. The noiseless function is a quadratic polynomial over a quadratic polynomial. To verify our assumptions and modeling thus far, we used an interpolation routine to estimate the coefficients of a quadratic over a quadratic rational function that fit the cost function of a noiseless, one real exponential signal. We then used the coefficients to determine the z that maximizes L with the analytical derivative approach just described.

The interpolation routine used stems from the Cauchy rational interpolation problem and utilizes the principle of "divided differences" as explained in Hildebrand (22:51-79) and Scharf (61:505-508). In 1990, Kumaresan applied divided differences—to estimate the coefficients of the transfer function associated with summed exponentials—with some success (37). It was his work that inspired us to use divided differences for estimating the coefficients of the rational maximum likelihood cost function. Unlike ourselves, Kumaresan's effort was not motivated by

the maximum likelihood approach to parameter estimation. His algorithm was a frequency domain analog of the extended Prony method with the same performance as the original extended Prony method (37).

We introduce approximation by divided differences for our application, by first illustrating its relationship to the linear interpolation problem and then extending it to the rational function interpolation problem. The properties of divided differences are numerous. We only identify those properties that lead to the rational function coefficient estimator.

Assume $U(z)$ is exactly defined by the linear model

$$(194) \qquad\qquad U(z) = u_0 + u_1 z$$

over a defined range of z. We can then say the ratio

$$(195) \qquad\qquad \frac{U(z_1) - U(z_0)}{z_1 - z_0}$$

is a constant, independent of z_0 and z_1 in the range of z. The ratio is called the first order divided difference of $U(z)$ and is denoted

$$(196) \qquad\qquad U[z_0, z_1] = \frac{U[z_1] - U[z_0]}{z_1 - z_0}$$

where $U[z_i] = U(z_i)$.

Furthermore, if $U(z)$ is exactly defined by a two degree polynomial, then $U[z_0, z_1]$ is a function of z, and the ratio

$$(197) \qquad\qquad \frac{U[z_1, z_2] - U[z_0, z_1]}{z_2 - z_0}$$

is also a constant, independent of z_0, z_1, and z_2 in the range of z. This ratio is called the second order divided difference of $U(z)$ and is denoted

$$(198) \qquad U[z_0, z_1, z_2] = \frac{U[z_1, z_2] - U[z_0, z_1]}{z_2 - z_0}.$$

Recursively, we can continue increasing the order, K, of the divided difference ratio with the equation

$$(199) \qquad U[z_0, z_1, \ldots, z_K] = \frac{U[z_1, z_2, \ldots, z_K] - U[z_0, z_1, \ldots, z_{K-1}]}{z_K - z_0}.$$

For example, assume $U(z)$ is exactly defined by a two degree polynomial such that

$$(200) \qquad U(z) = u_0 + u_1 z + u_2 z^2.$$

Then as defined,

$$
\begin{aligned}
U[z_0] &= u_0 + u_1 z_0 + u_2 z_0^2 \\
U[z_0, z_1] &= \frac{u_0 + u_1 z_1 + u_2 z_1^2 - (u_0 + u_1 z_0 + u_2 z_0^2)}{z_1 - z_0} \\
&= \frac{u_1(z_1 - z_0) + u_2(z_1^2 - z_0^2)}{z_1 - z_0} \\
&= u_1 + u_2(z_1 + z_0) \\
U[z_0, z_1, z_2] &= \frac{u_1 + u_2(z_2 + z_1) - (u_1 + u_2(z_1 + z_0))}{z_2 - z_0} \\
&= \frac{u_2(z_2 + z_1 - z_1 z_0)}{z_2 - z_0} \\
&= u_2 \equiv \text{constant} \\
U[z_0, z_1, z_2, z_3] &= \frac{u_2 - u_2}{z_3 - z_0} \\
(201) \qquad &= 0.
\end{aligned}
$$

This example illustrates an important property of divided differences: when the order of the divided difference exceeds the degree of the underlying polynomial, the divided difference ratio is zero. This property extends to all orders of divided differences that exceed any degree polynomial when the underlying function is exactly defined by the polynomial (22:57). If we assume a function is closely modeled by a Qth degree polynomial, we also assume all Kth order divided differences greater than Q are equal to 0.

Another feature of divided differences, needed to explain our rational function coefficient estimator, can be extracted from the recursive definition of the Kth order divided difference, Equation 199. We see that the desired order divided difference is a function of the previous order divided differences. When the definitions of the previous order divided differences are substituted back into Equation 199, a more generalized divided difference equation results (22:55). In the new equation

$$(202) \qquad U[z_0, z_1, \ldots, z_K] = \sum_{i=0}^{K} \frac{U(z_i)}{W(z_i)}$$

where

$$(203) \qquad W(z_i) = \prod_{\substack{j=0 \\ j \neq i}}^{K} (z_i - z_j)$$

Returning to the rational maximum likelihood cost function, let L take the form

$$(204) \qquad L(z) = \frac{\sum_{r=0}^{R} v_r z^r}{\sum_{q=0}^{Q} u_q z^q} = \frac{V(z)}{U(z)}$$

92

implying that the degrees of the numerator and of the denominator of L are R and Q, respectively. Equivalently, let

$$(205) \qquad\qquad V(z) = L(z)U(z).$$

From our previous divided difference example, we know that if we assume $V(z)$ is well modeled by an Rth order polynomial, we have

$$(206) \qquad\qquad V[z_0, z_1, \ldots, z_K] = \sum_{i=0}^{K} \frac{V(z_i)z_i^s}{W(z_i)} = 0$$

provided that $K > R$. Furthermore, if $K > R + s$,

$$(207) \qquad\qquad V[z_0, z_1, \ldots, z_K] = \sum_{i=0}^{K} \frac{V(z_i)z_i^s}{W(z_i)} = 0$$

holds for the polynomial $V(z)$ times the polynomial z^s. When we combine Equation 205 and Equation 207 we obtain, for $K > R + s$,

$$(208) \qquad
\begin{aligned}
\sum_{i=0}^{K} \frac{L(z_i)U(z_i)z_i^s}{W(z_i)} &= 0 \\
\Rightarrow \sum_{i=0}^{K} \frac{L(z_i)}{W(z_i)} \sum_{q=0}^{Q} u_q z_i^q z_i^s &= 0 \\
\Rightarrow \sum_{q=0}^{Q} u_q \sum_{i=0}^{K} \frac{L(z_i)}{W(z_i)} z_i^{q+s} &= 0.
\end{aligned}$$

Recall that Equation 208 holds for all z_i in the range of z associated with the polynomial assumption. By choosing at least Q values of s and designating $K + 1$ values of z_i, $(z_0, z_1, \ldots, z_K)$, where $K > R + s$, we can develop a system of linear equations for estimating the $Q + 1$ coefficients of $U(z)$, $(u_0, u_1, \ldots, u_Q)$. In the noiseless approximation of L for example, the numerator and denominator polynomial degrees, Q and R respectively, are equal to two. Therefore, after choosing

93

an s of 0 and 1, and letting u_0 normalize to 1, we can create the exactly determined system of two equations

$$
(209) \quad
\begin{bmatrix}
\sum_{i=0}^{K>2} \frac{L(z_i)}{W(z_i)} z_i^1 & \sum_{i=0}^{K>2} \frac{L(z_i)}{W(z_i)} z_i^2 \\
\sum_{i=0}^{K>3} \frac{L(z_i)}{W(z_i)} z_i^2 & \sum_{i=0}^{K>3} \frac{L(z_i)}{W(z_i)} z_i^3
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2
\end{bmatrix}
=
\begin{bmatrix}
-\sum_{i=0}^{K>2} \frac{L(z_i)}{W(z_i)} z_i^0 \\
-\sum_{i=0}^{K>3} \frac{L(z_i)}{W(z_i)} z_i^1
\end{bmatrix}
$$

where $L(z_i)$ is evaluated with the data, $y[m]$, under the maximum likelihood cost function

$$
(210) \quad L(z_i) = \frac{\sum_{m=0}^{M-1} (y[m] z_i^m)^2}{\sum_{m=0}^{M-1} z_i^{2m}}.
$$

After estimating the coefficients of $U(z)$ with Equation 209 we return to the relationship

$$
(211) \quad
\begin{aligned}
V(z_i) &= L(z_i)U(z_i) \\
\Rightarrow \sum_{r=0}^{R} v_r z_i^r &= L(z_i) \sum_{q=0}^{Q} u_q z_i^q.
\end{aligned}
$$

With $T+1$ values of z_i designated, where $T > R$, Equation 211 can be used in a system of equations for estimating the coefficients of $V(z)$. Under the noiseless, $Q = R = 2$, model assumption we can create the overdetermined system of linear equations in the unknowns $v_0, v_1, \ldots v_R$

$$
(212) \quad
\begin{bmatrix}
z_0^0 & z_0^1 & z_0^2 \\
z_1^0 & z_1^1 & z_1^2 \\
\vdots & \vdots & \vdots \\
z_T^0 & z_T^1 & z_T^2
\end{bmatrix}
\begin{bmatrix}
v_0 \\
v_1 \\
v_2
\end{bmatrix}
=
\begin{bmatrix}
L(z_0) \sum_{q=0}^{2} u_q z_0^q \\
L(z_1) \sum_{q=0}^{2} u_q z_1^q \\
\vdots \\
L(z_T) \sum_{q=0}^{2} u_q z_T^q
\end{bmatrix}.
$$

Continuing with the noiseless model of L analysis, the estimated numerator and denominator polynomial coefficients can be used with $\frac{\partial L}{\partial z}$ to find the maxima and minima of L. The derivative of L takes the form

$$
\begin{aligned}
\frac{\partial L}{\partial z} &= \frac{\partial}{\partial z}\left(\frac{v_0 + v_1 z + v_2 z^2}{1 + u_1 z + u_2 z^2}\right) \\
&= \frac{(v_1 + 2v_2 z)(1 + u_1 z + u_2 z^2) - (v_0 + v_1 z + v_2 z^2)(u_1 + 2u_2 z)}{(1 + u_1 z + u_2 z^2)^2} \\
(213) \qquad &= \frac{(v_1 - u_1 v_0) + 2(v_2 - u_2 v_0)z + (u_1 v_2 - u_2 v_1)z^2}{(1 + u_1 z + u_2 z^2)^2}.
\end{aligned}
$$

We know $\frac{\partial L}{\partial z} = 0$ when the numerator of $\frac{\partial L}{\partial z}$ is equal to zero. Therefore, we attain the maxima and minima by rooting the numerator polynomial of $\frac{\partial L}{\partial z}$. The real root in the range $(0, 1)$ that effects the global maximum L evaluation is considered the maximum likelihood solution.

Figures 27, 28, and 29 serve to validate our divided differences based estimation algorithm and model assumptions for the noiseless case. In the figures, actual L and modeled L evaluations are plotted over the range of z, for data record lengths of $M = 10$, 25 and 50 respectively. The noiseless signal was created with the same parameters used at the beginning of this chapter. A fourth order, $K = 4$, divided difference, with five equally spaced values of z_i between $(0, 1)$, was used in an exactly determined system of equations for estimating the coefficients of the two degree denominator polynomial, $U(z)$. Because two equations were necessary for the two unknown coefficients (u_0 was assumed normalized to 1), s values of 0 and 1 were sufficient. Therefore, $K = 4 > R + s$ satisfies the minimum divided difference order for both equations. One hundred values of z_i, equally spaced on the interval $(0, 1)$, were designated to estimate the coefficients of the two degree numerator polynomial, $V(z)$. This resulted in 100 equations for the three unknowns in the overdetermined
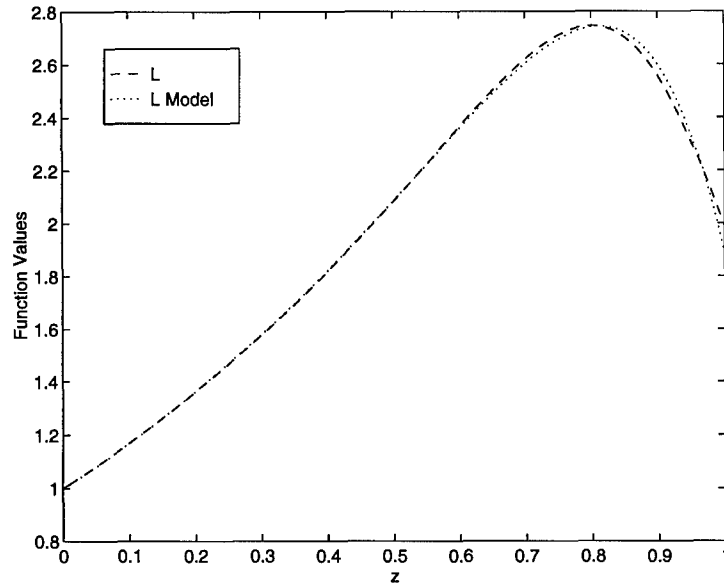
Figure 27. Evaluation of the actual L and the noiseless modeled L over the range of z for noiseless data with a record length of 10. The signal exponential is $\lambda = .8$ and the maximum L modeled occurs at $z = .8125$.

Figure 28. Evaluation of the actual L and the noiseless modeled L over the range of z for noiseless data with a record length of 25. The signal exponential is $\lambda = .8$ and the maximum L modeled occurs at $z = .8005$.

Figure 29. Evaluation of the actual L and the noiseless modeled L over the range of z for noiseless data with a record length of 50. The signal exponential is $\lambda = .8$ and the maximum L modeled occurs at $z = .7995$.
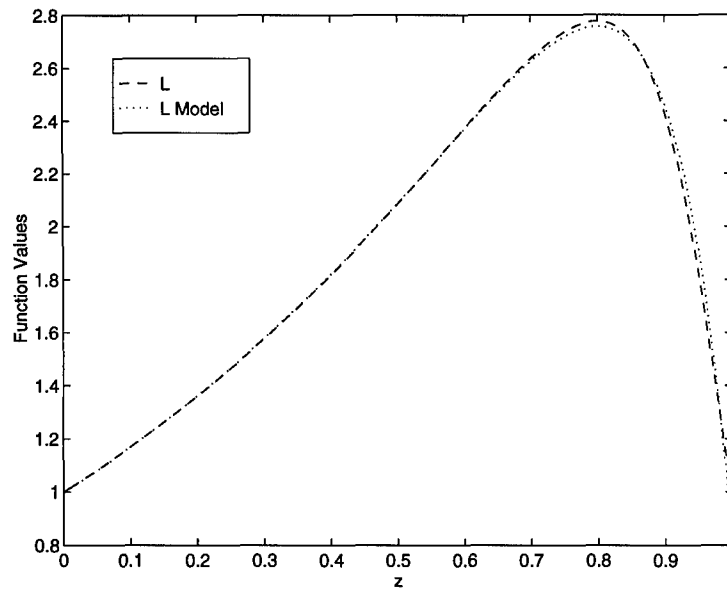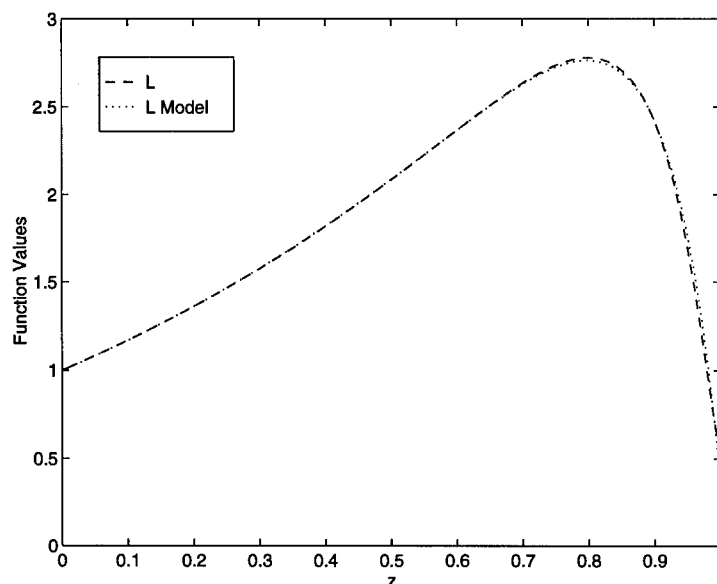
linear system. After the coefficients were estimated, maximizing z values of .8124, .8005, and .7995, for $M = 10$, 25, and 50 respectively, were calculated from the maximizing roots of the numerator polynomial of $\frac{\partial L}{\partial z}$.

From the figures and maximizing values of z, we see that the algorithm and model assumptions are accurate and valid for the noiseless case. As expected, increasing the data record length from 10 to 25 increased the estimator accuracy, but increasing the data record length from 25 to 50 had no appreciable affect, because at those values of M the low degree model assumption is appropriate. We also noted that increasing the divided difference order, K, for the $U(z)$ coefficient estimates was unnecessary and in fact reduced estimator accuracy. Additionally, varying the number of equations for the $V(z)$ coefficient estimates from 50 to 200 had little effect on the solution.
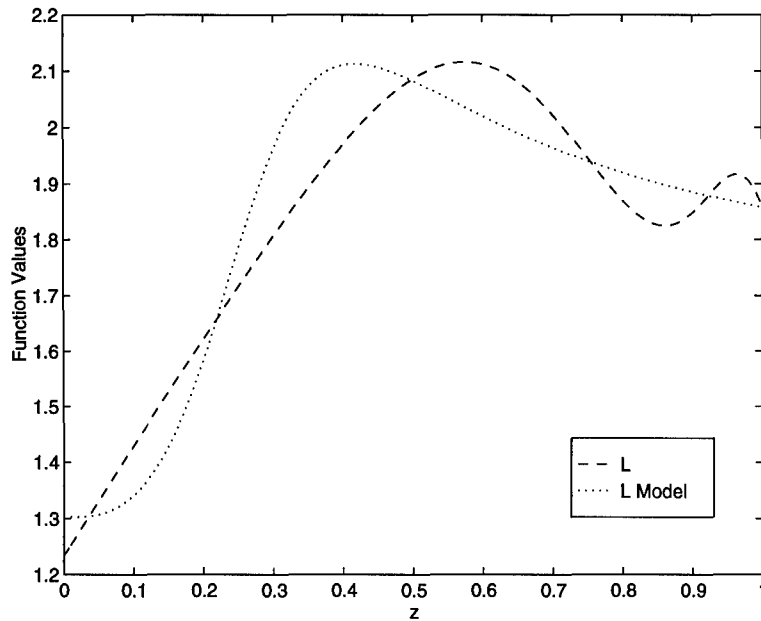
Figure 30. Evaluations of the actual L and the modeled L over the range of z for noisy data with a record length of 25. The signal exponential is $\lambda = .8$ and the SNR is 3 dB. The rational function model consisted of a two degree polynomial over a two degree polynomial.

The divided differences maximum likelihood estimator just developed is not predicated on a unimodal cost function. When the actual cost function is multi-modal over the range of z—as is the case in low SNR scenarios—a rational function will model multi-modality. Disregarding the inadequacies of a two degree polynomial over a two degree polynomial rational function for modeling the cost function of a noisy signal, the noiseless divided difference algorithm just created was utilized on the same signal with additive noise at 3 dB SNR. The data record length was 25. Figure 30 illustrates the results. The modeled L is indeed multi-modal, but its maximum is not in the vicinity of the actual maximum of L.

When we recall the noisy model of L,

$$L(z; c, \lambda, M) \approx \frac{c^2(1 - z^2)}{1 - 2\lambda z + \lambda^2 z^2} +$$

$$\frac{2c(1 - \lambda z - z^2 + \lambda z^3) \sum_{m=0}^{M-1} w[m]z^m}{1 - 2\lambda z + \lambda^2 z^2} +$$

(214)
$$\frac{(1 - 2\lambda z + (\lambda^2 - 1)z^2 + 2\lambda z^3 - \lambda^2 z^4) \left(\sum_{m=0}^{M-1} w[m]z^m\right)^2}{1 - 2\lambda z + \lambda^2 z^2},$$

we note a common two degree denominator polynomial exists, but the second and third rational function numerators are calling for $M+2$ and $2M+2$ degree polynomials respectively. Therefore, a divided difference algorithm for noisy signals requires a higher degree numerator polynomial in its noisy rational function model. Increasing the degree to $2M+2$ is not acceptable because of the increased numerical complexity.

We observed that the first few terms of

(215)
$$\sum_{m=0}^{M-1} w[m]z^m$$

are the most influential because of the decreasing nature of z^m as m increases, and used this rational to justify a smaller degree increase in the numerator polynomial. The coefficients for at least a four degree numerator polynomial are clearly supported in the third rational function of Equation 214. Thus, we assumed a four degree polynomial over a two degree polynomial rational function would be sufficient for modeling the noisy L.

The previous divided difference algorithm is easily generalized for accommodating a four degree numerator polynomial. A divided difference order, $K = 6 > R + s$, was utilized to estimate the two unknown coefficients of the denominator polyno-
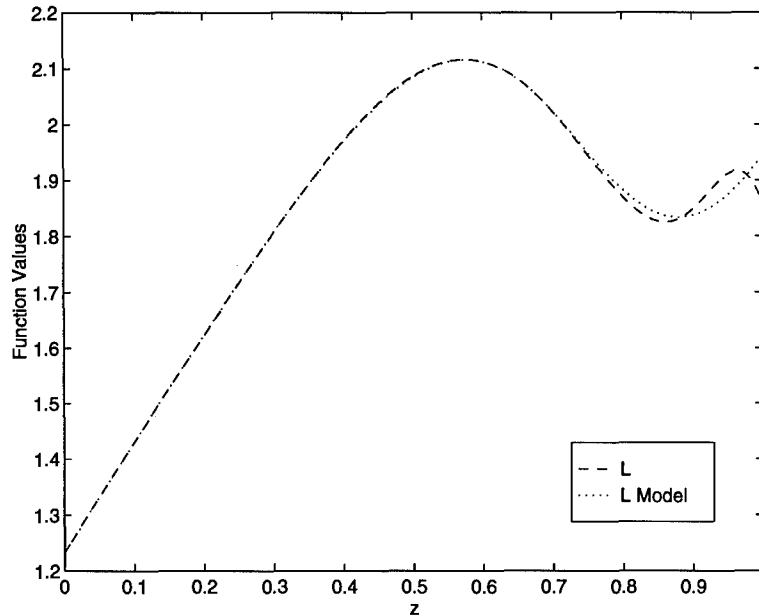
Figure 31. Realization 1 evaluations of the actual L and the modeled L over the range of z for noisy data with a record length of 25. The signal exponential is $\lambda = .8$ and the SNR is 3 dB. The rational function model consisted of a four degree polynomial over a two degree polynomial.

mial, $U(z)$, and 200 z_i values between $(0, 1)$ were designated for estimating the five unknown coefficients of the numerator polynomial, $V(z)$.

We see from Figures 31 and 32 that the divided difference algorithm, with a four degree numerator polynomial over a two degree denominator polynomial, provides a modeled L similar to the actual L of a noisy 25 data point signal at 3 dB SNR. Note that for two different realizations, neither the maximum of the modeled L or the actual L resides at the correct z estimate of .8. This variance can be anticipated when recalling from CRB theory that the minimum variance of all estimators must increases as SNR increases.

To accurately access the performance of the noisy divided difference algorithm, the same Monte-Carlo experiment, as performed on the golden section search and IGLS algorithms, was conducted. For a quick review of the experiment's parame-
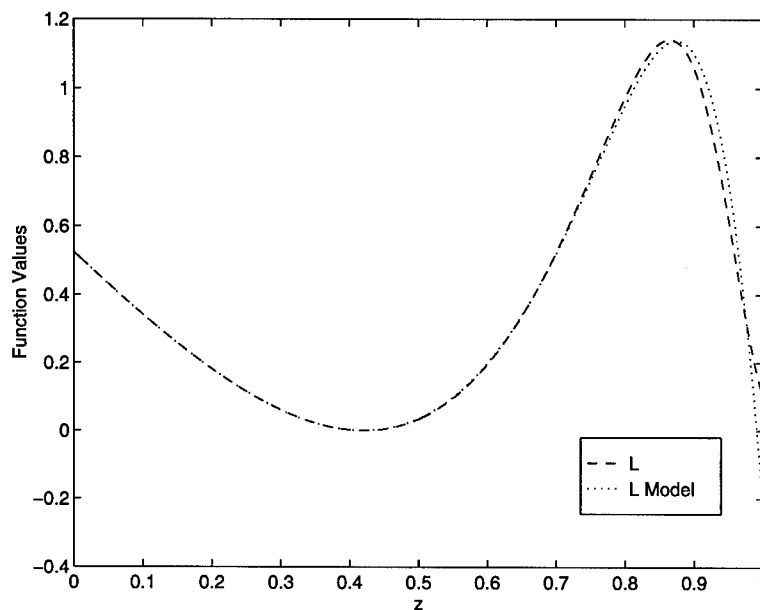
Figure 32. Realization 2 evaluations of the actual L and the modeled L over the range of z for noisy data with a record length of 25. The signal exponential is $\lambda = .8$ and the SNR is 3 dB. The rational function model consisted of a four degree polynomial over a two degree polynomial.
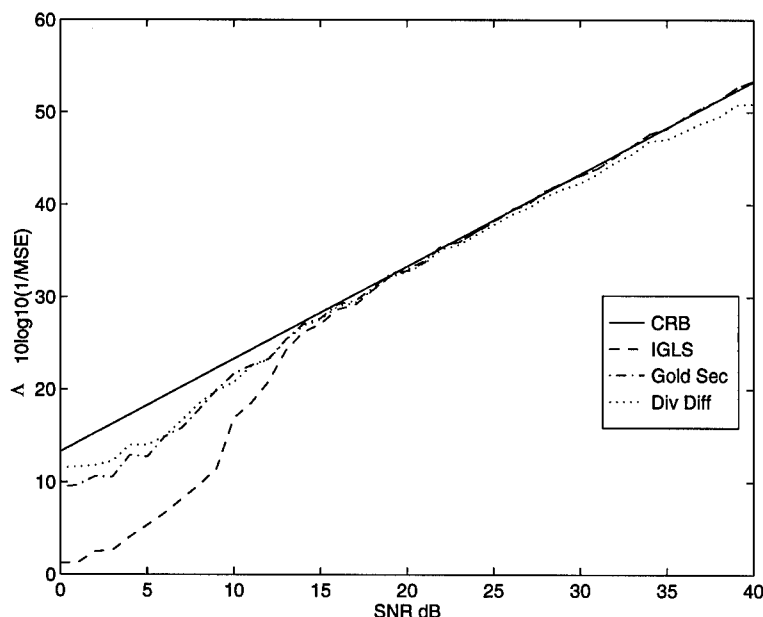
Figure 33. The inverse MSE of the divided difference, golden section, and IGLS algorithms versus the CRB over a range of SNRs. The underlying signal is of one real exponential with $\lambda = .8$

ters, 400, twenty-five data point realizations of $\bar{y}$ were created for each SNR. The underlying signal possessed a λ of .8 and a c of 1.

The performance of all three estimators is illustrated in Figures 33 and 34. The divided difference algorithm performed slightly better than the other two algorithms at low SNR. The improved performance is attributed to its multi-mode fitting capability. As the SNR increased, the divided difference algorithm performed slightly worse than the other two algorithms. This trend continues as the SNR increases to its maximum for machine precision.

The divided difference based estimation algorithm is best suited for noise corrupted signals with low SNRs. At moderate to high SNRs, the algorithm is too complex and the modeling is excessive. However, further experimentation with the algorithm itself is recommended. Investigation for appropriate divided difference
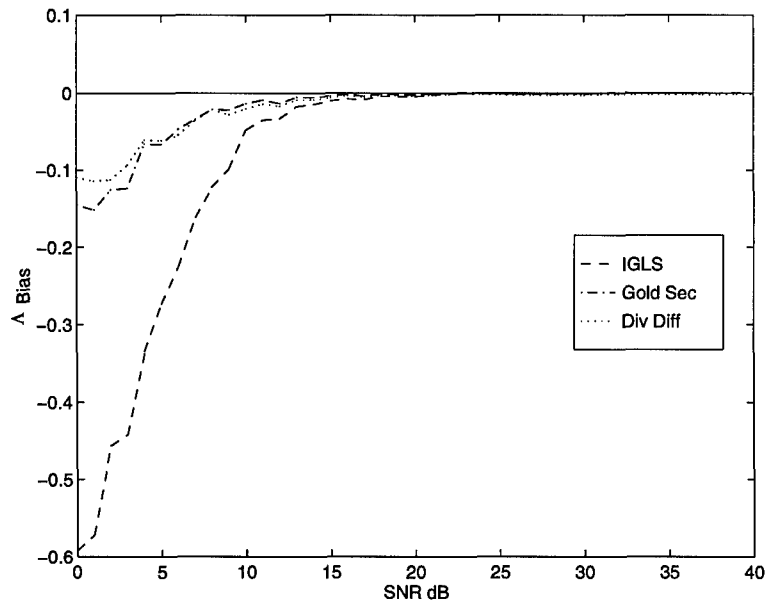
Figure 34.   The bias of the divided difference, golden section, and IGLS algorithms
over a range of SNRs. The underlying signal is of one real exponential
with $\lambda = .8$

orders, K, and system of equations sizes will provide additional insight into the
problem. Additionally, simplified cost functions for two and three mode, complex
and real applications may exist.

In conclusion of the one real exponential general analysis, the divided difference
algorithm's minimal improvement over the golden section search, at low noise, re-
moves our concerns for estimator degradation due to multi-modality. If the proposed
problem can be restricted to that of a single real exponential in noise, we recommend
the golden section search because of its consistently good performance over all SNRs.
However, if the SNR of the signal is not very low, the IGLS algorithm's performance
is definitely satisfactory.

5.3 Two Real Exponentials

The next phase of our real exponential parameter estimation analysis includes adding a second exponential for an in depth look at the summed exponential problem. The two real exponential problem is especially enlightening because minimization of the classical maximum likelihood cost function can be visualized on a two dimensional contour plot. Consider the 25 data point observation, $\bar{y}$, with underlying parameters $c_1 = .5$, $c_2 = 1$, $\lambda_1 = .8$, and $\lambda_2 = .5$ and the maximum likelihood cost function

$$(216) \qquad L(z_1, z_2) = \bar{y}^T(I - V(V^T V)^{-1} V^T)\bar{y}$$

where

$$
\begin{aligned}
V &= \begin{bmatrix} \bar{v}(z_1) & \bar{v}(z_2) \end{bmatrix} \\
\bar{v}(z_n) &= \begin{bmatrix} 1 & z_n & z_n^2 & \cdots & z_n^{M-1} \end{bmatrix}^T
\end{aligned}
$$

(217)

and z_1 and z_2 represent the unknown exponential values to be estimated. Figure 35 is a contour plot of the surface generated by the cost function evaluated over the range of z_1 and z_2 at an SNR of 30 dB. The two plus signs identify the location of two desired solutions and help illustrate the symmetry of the solution space. The intense relief on the diagonal is brought about by the singularity that occurs in $(V^T V)^{-1}$ when $z_1 \approx z_2$. Additionally, we observed that the general shape of the maximum likelihood contour surface is relatively invariant to changes in SNR. Like in the one mode analysis, additional multi-modal curves do not appear until SNR low.

A two mode nonlinear search of the cost function for the maximum likelihood solution appears to be achievable, but before embarking on that path, consider the
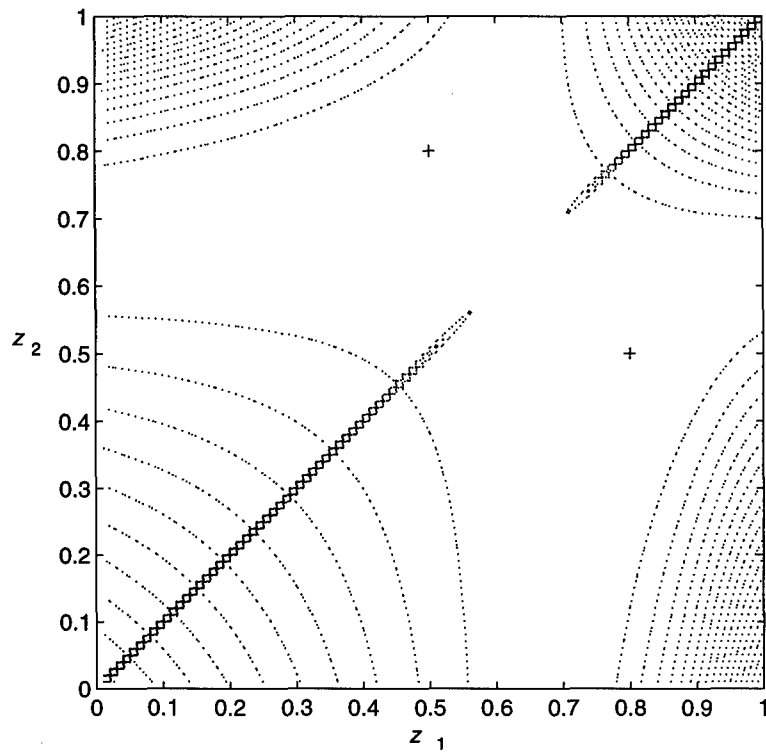
Figure 35. Contour plot of the maximum likelihood cost function evaluated over the range of z_1 and z_2 with underlying parameters $c_1 = .5$, $c_2 = 1$, $\lambda_1 = .8$, and $\lambda_2 = .5$ at an SNR of 30 dB.
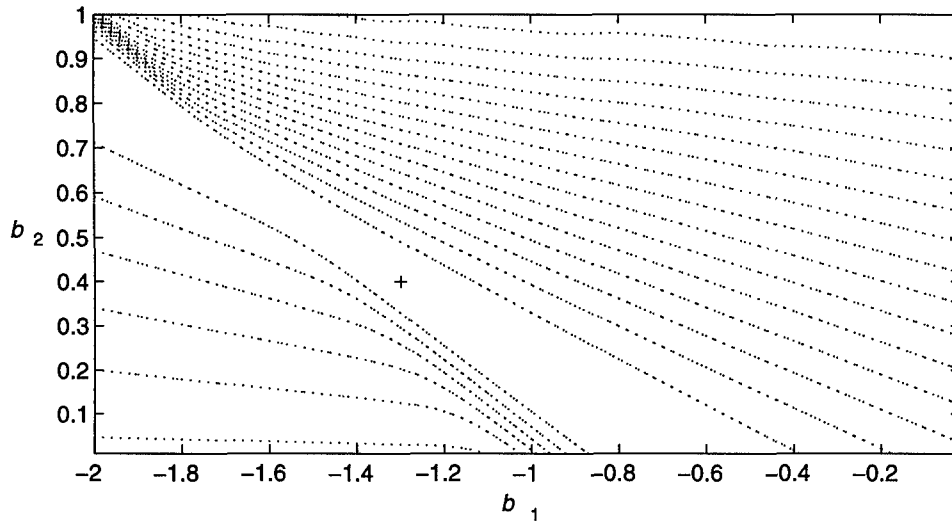
Figure 36. Contour plot of the maximum likelihood cost function evaluated over the range of b_1 and b_2 with underlying parameters $c_1 = .5$, $c_2 = 1$, $\lambda_1 = .8$, and $\lambda_2 = .5$ at an SNR of 30 dB.

cost function after being transformed by the relationship

$$(218) \qquad (I - V(V^T V)^{-1} V^T) = B^T (BB^T)^{-1} B.$$

Figure 36 is a contour plot of the 30 dB, two mode, linear prediction coefficient based, maximum likelihood surface. A plus sign identifies the desired solution. To facilitate a one-to-one mapping between z_n and b_n, b_0 was assumed to be one. Immediately, we see that there is no symmetry in the linear prediction coefficient based contour plot, and that the curved valley, enclosing the maximum likelihood solution in the previous plot, is now represented as a straight valley.

A closer review of the transformation on the parameters z_1 and z_2 limits the feasibility region of the linear prediction coefficient based contour plot as well. Because z_1 and z_2 are only considered in the interval $(0, 1)$, and because they are the quadratic roots of the polynomial formed by the linear prediction coefficients, we

106

know

$$(219) \qquad\qquad b_1 = -(z_1 + z_2)$$

and

$$(220) \qquad\qquad b_2 = z_1 z_2.$$

Therefore,

$$(221) \qquad\qquad -2 < b_1 < 0$$

and

$$(222) \qquad\qquad 0 < b_2 < 1.$$

Additionally, from the quadratic equation

$$(223) \qquad\qquad z_n = \frac{-b_1 \pm \sqrt{b_1^2 - 4b_2}}{2},$$

we know that a real z_n mandates

$$(224) \qquad\qquad \sqrt{b_1^2 - 4b_2} > 0.$$

Therefore,

$$(225) \qquad\qquad b_2 < \frac{1}{4}b_1^2.$$

Finally, since

(226)
$$\sqrt{b_1^2 - 4b_2} > 0,$$

and z is confined to the interval $(0, 1)$, we know

$$
\begin{aligned}
\frac{-b_1 + \sqrt{b_1^2 - 4b_2}}{2} &< && 1 \\
\Rightarrow \quad \sqrt{b_1^2 - 4b_2} &< && 2 + b_1 \\
\Rightarrow \quad b_1^2 - 4b_2 &< && 4 + 4b_1 + b_1^2 \\
\Rightarrow \quad -b_2 &< && 1 + b_1 \\
\Rightarrow \quad b_2 &> && -b_1 - 1.
\end{aligned}
$$

The boundaries of the inequalities just identified are illustrated in Figure 37. The label A is assigned to the $b_2 > 0$ boundary, the label B is assigned to the $b_2 > -b_1 - 1$ boundary, and the label C is assigned to the $b_2 < \frac{1}{4}b_1^2$ boundary. The boundaries are not a function of SNR, and their relationship to the 30 dB SNR maximum likelihood contour plot is identified in Figure 38. Interestingly, the curved boundary C equates to the case of $z_1 = z_2$, but in the linear prediction coefficient based representation, the cost function is not subject to the same singularities as previously noted in the z_1 versus z_2 plot.

When considering a search plan for the two mode maximum likelihood solution, the straight valley—prevalent in all two mode, linear prediction coefficient based, contour plots—attracts particular attention. The valley consistently runs through the feasibility region, and if the two mode problem could be reduced to a one dimensional line search along the valley, considerable efficiency would be attained.
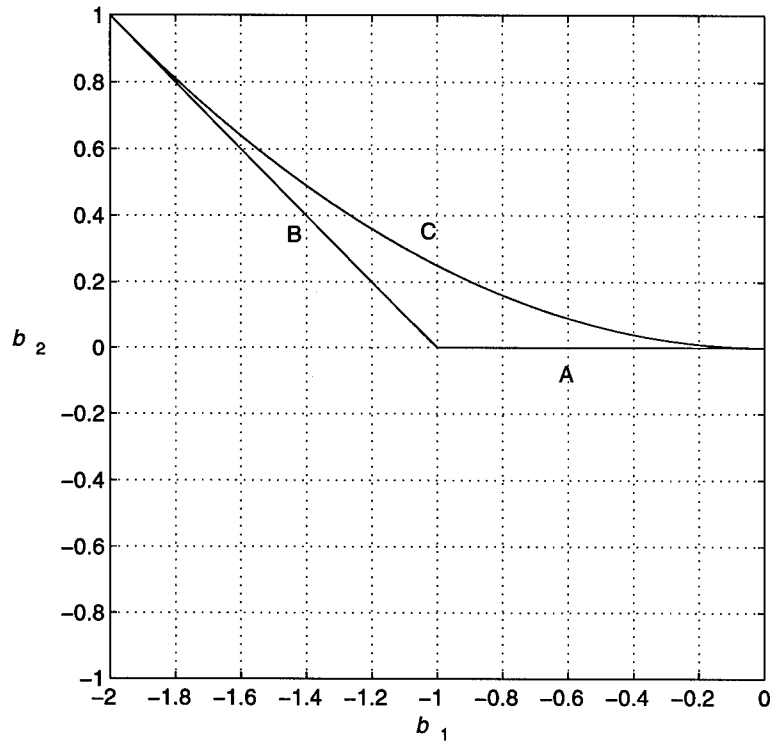
Figure 37. Boundaries of the feasibility region for the sum of two decaying real exponentials.
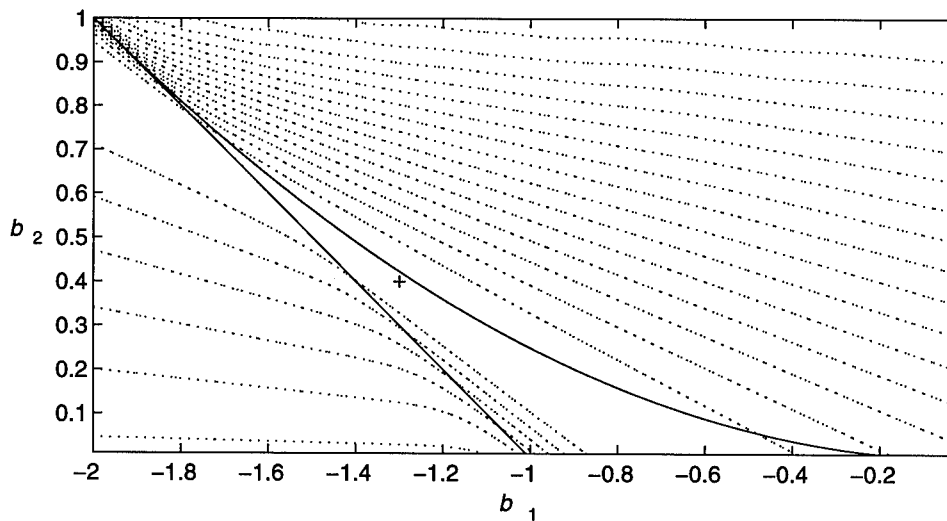

Figure 38. Contour plot of the maximum likelihood cost function evaluated over the range of b_1 and b_2 at an SNR of 30 dB. The boundaries of the feasibility region are also identified with solid lines.

109

Initially, a simple algorithm was devised to search the line, drawn between the minimums of the A and B borders, for a minimal solution. The behavior of the maximum likelihood cost function along the A and B, and even C, borders was similar to that of the previous one mode section: L evaluated unimodally, over the range of the border, under all but very low SNR scenarios. A golden section search was utilized to determine the minimums of the A and B borders and thus estimate the endpoints of a line in the valley. A third golden section search was utilized to identify the minimum along the line of the valley. The algorithm always provided reasonable estimates, but as SNR decreased, the line of the valley appeared to only graze the true minimum. This resulted in estimates slightly, but consistently inferior to the IGLS method of searching the maximum likelihood surface.

To identify the reason for this inferior performance, the linear valley assumption was reconsidered. The behavior of the valley in general is explained by the summed exponential's relationship to the linear difference, or linear prediction, equation. Recall that the underlying, two mode, noiseless signal is a solution to

$$(227) \qquad\qquad s[m] + b_1 s[m-1] + b_2 s[m-2] = 0$$

Again b_0 is assumed to be normalized to 1. Equation 227 can be rearranged to take the form of a line expressed in the unknowns b_1 and b_2,

$$(228) \qquad\qquad b_2 = \frac{-s[m-1]}{s[m-2]} b_1 - \frac{s[m]}{s[m-2]}.$$

The slope and intercept of the line varies as $s[m]$ varies from $m = 0, 1, \ldots, M-1$, but all lines defined by Equation 228 intersect at the desired solution for b_1 and b_2. In figure 39 the lines defined by Equation 228 are incorporated with the feasibility
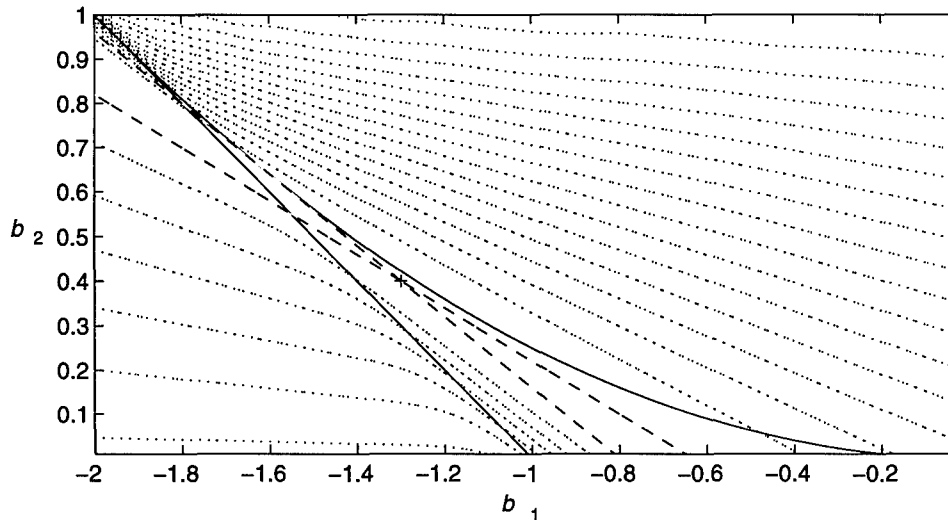
Figure 39. Contour plot of the maximum likelihood cost function evaluated over the range of b_1 and b_2 at an SNR of 30 dB. The boundaries of the feasibility region and the linear prediction solution lines are also identified with solid lines.

region and the contour plots for the 30 dB, two mode, problem. Only the upper and lower lines are displayed for clarity. From the figure, the relationship between the intersecting lines and the valley of the contour plot is illustrated.

Although this relationship has been identified, the lines defined by Equation 228 consist of entirely unknown information in an estimation context. In theory, a line in the valley could be used for a one dimensional search of the minimum. In practice, accurately estimating the parameters of such a line requires the same, if not more, effort as the original plan of searching for the two dimensional minimum.

With the one dimensional search option dismissed, attention was given to improving the two dimensional search. The linear prediction coefficient based graphical representation just created for the real, two mode, problem provides considerable insight into the mechanics of the IGLS algorithm. As discussed in Chapter IV, initializing the IGLS algorithm with a least squares estimate of $\bar{b}$ is equivalent to building
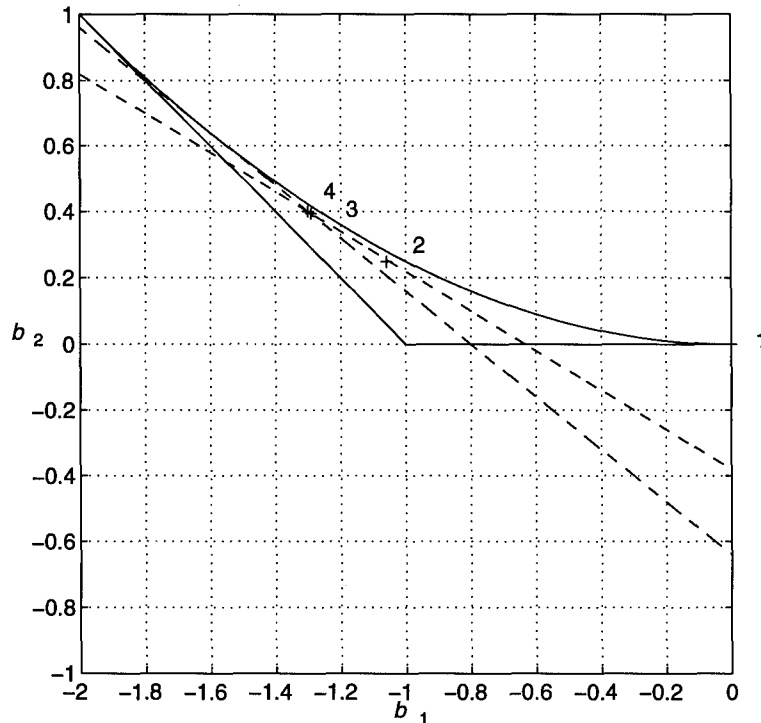
Figure 40. Location of b_1 and b_2 for each iteration of the IGLS algorithm plotted with respect to the feasibility region for the 40 dB SNR, two mode, real exponential problem.

the B matrix, for an IGLS iteration, with $b_0 = 1$ and $b_1, b_2, \ldots, b_N = 0$. In the two mode real exponential problem, the point $b_1 = b_2 = 0$ resides at the A and C border intersection of the feasibility region. In Figure 40, b_1 and b_2 estimates from each iteration of the IGLS algorithm for the 40 dB SNR problem are identified. We see that after initializing at $b_1 = b_2 = 0$, the first iteration estimates b_1 and b_2 in the valley and near the center of the feasibility region. In subsequent estimates, b_1 and b_2 track up the valley to the desired solution.

When each iteration of the IGLS algorithm is observed at lower SNRs, subsequent estimates of b_1 and b_2 occasionally occur outside of the feasibility region. In Figure 41 we see the first update of b_1 and b_2 is in the valley but well below the feasibility region. For this realization and SNR, subsequent estimates of b_1 and b_2
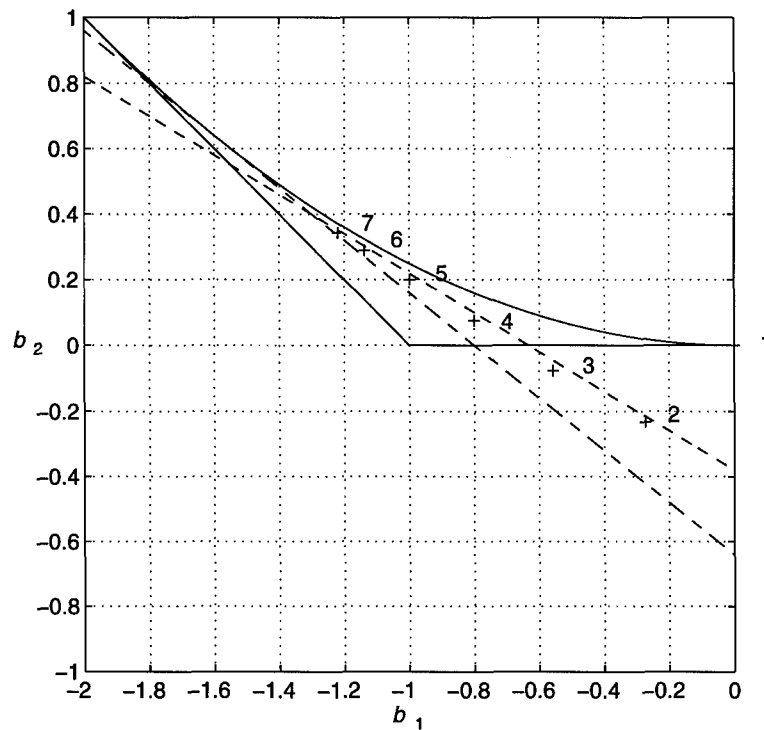
112

Figure 41. Locations of b_1 and b_2 for each iteration of the IGLS algorithm plotted with respect to the feasibility region for the 25 dB SNR, two mode, real exponential problem.
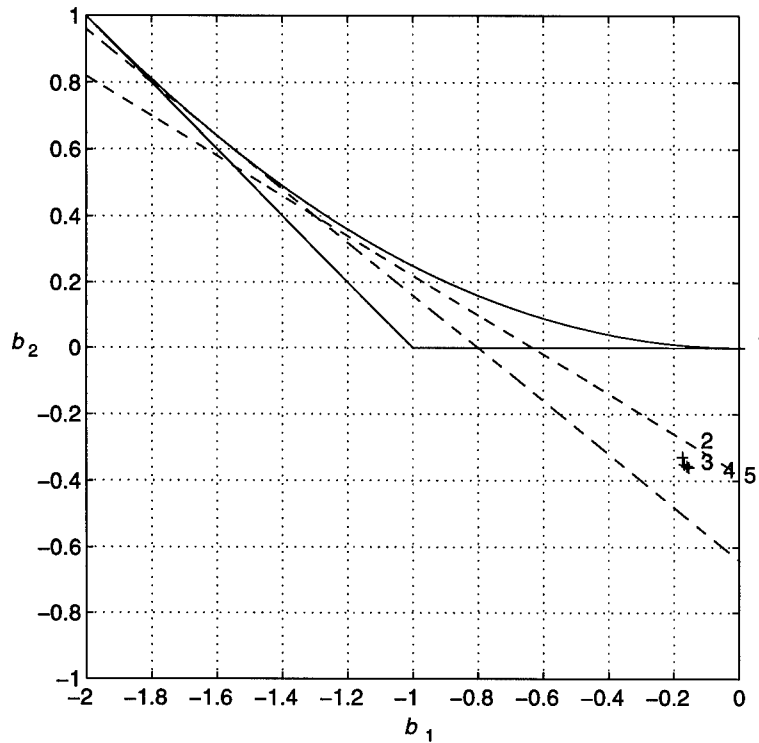
Figure 42. Locations of b_1 and b_2 for each iteration of the IGLS algorithm plotted with respect to the feasibility region for the 20 dB SNR, two mode, real exponential problem.

track up the valley to the desired solution. However, returning to the feasibility region in subsequent updates is not always the case. In Figure 42 we see, for a realization at 20 dB, the first update occurs outside of the feasibility region, and the subsequent updates track further away from the region. To avoid this malady, alternative initialization points for the IGLS algorithm were considered.

In the previous work, we efficiently found the location of the minimums on the borders of the feasibility region with a golden section search. The A and B border minimums typically occur in the valley of the linear prediction coefficient based contour plots, even at low SNR. Similarly, for the same realizations, the C border minimum typically occurs very near the desired solution. In Figure 43, the

114
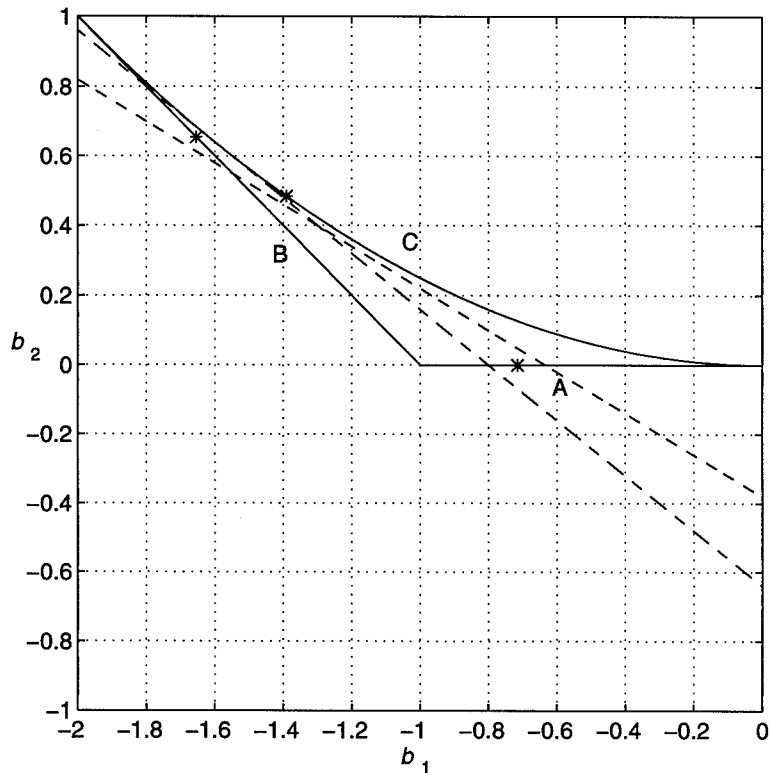
Figure 43. Minimums identified on the A, B, and C borders of the feasibility region
for the real, two mode, problem at 5 dB SNR.

border minimums are identified for a realization of the standard two mode problem

at 5 dB SNR.

With this information, a Monte-Carlo experiment was conducted to asses the

affect of initializing the IGLS algorithm from the A, B, and C border minimums.

The results of the experiment are displayed, in terms of inverse MSE and bias, in

Figures 44, 45, 46, and 47.   The performance of the IGLS algorithm, initialized

from the least squares estimate ($b_1 = b_2 = 0$), versus the IGLS algorithm, initialized

from each border minimum, is illustrated in each plot. In every case, 15 iterations

of the IGLS algorithm were implemented. From the figures, we conclude that very

Figure 44. The λ_1 inverse MSE of the standard IGLS algorithm, initialized from $b_1 = b_2 = 0$, and the IGLS algorithm, initialized from the minimums of the A, B, and C borders. Also plotted is the CRB.

Figure 45. The λ_1 bias of the standard IGLS algorithm, initialized from $b_1 = b_2 = 0$, and the IGLS algorithm, initialized from the minimums of the A, B, and C borders, over a range of SNRs.

116
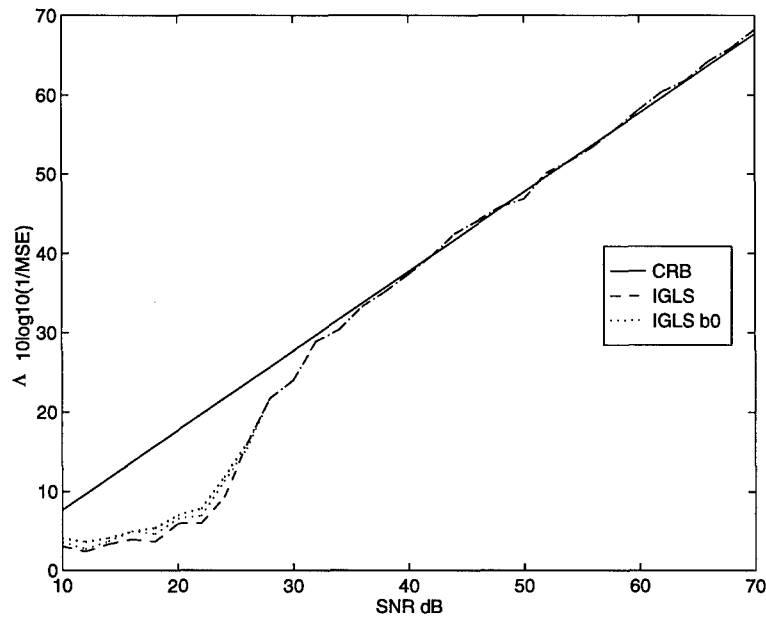
Figure 46. The λ_2 inverse MSE of the standard IGLS algorithm, initialized from $b_1 = b_2 = 0$, and the IGLS algorithm, initialized from the minimums of the A, B, and C borders. Also plotted is the CRB.
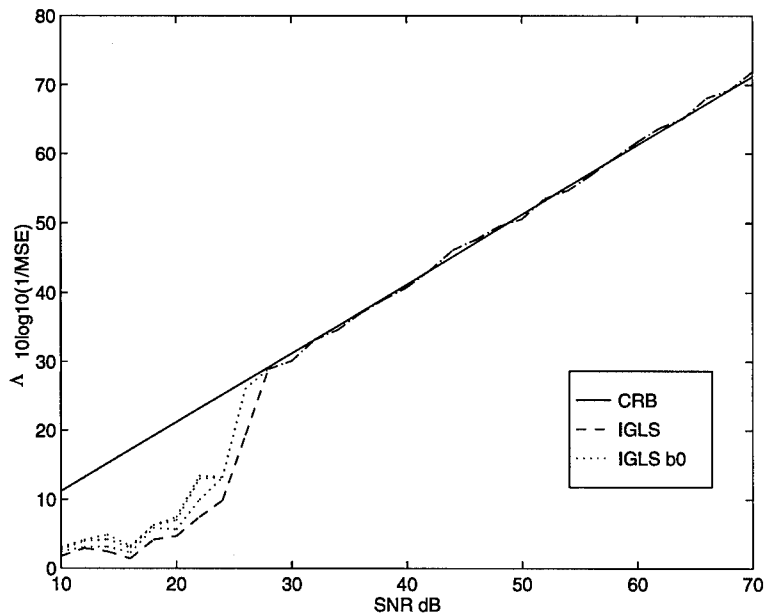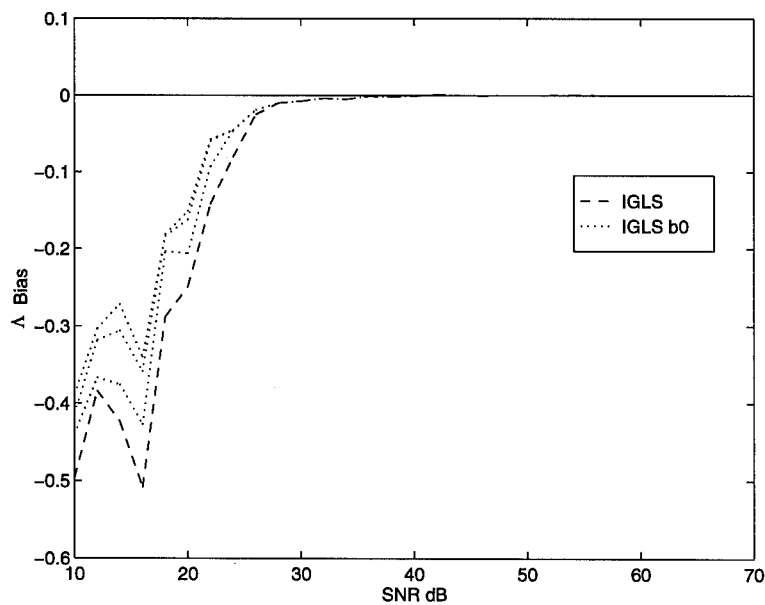

Figure 47. The λ_2 bias of the standard IGLS algorithm, initialized from $b_1 = b_2 = 0$, and the IGLS algorithm, initialized from the minimums of the A, B, and C borders, over a range of SNRs.

117
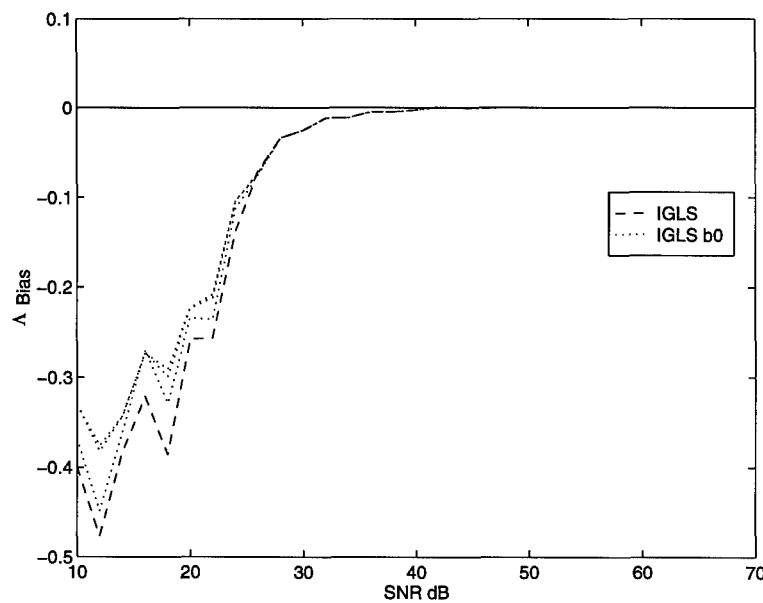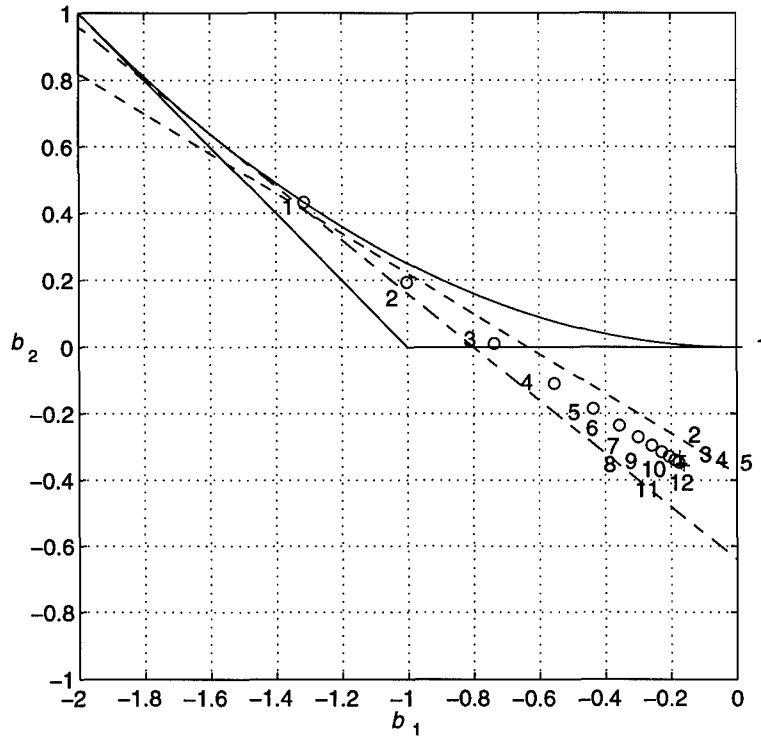
Figure 48. Realization 1 locations of b_1 and b_2 for each iteration of the IGLS algorithm, initialized at two different points. One starting point was at the least squares solution and the other starting point was at the minimum of the C border. The SNR was 20 dB.

little is gained by initializing the IGLS algorithm at a point nearer to the desired solution.

We postulate that given enough iterations, the IGLS algorithm will terminate at the same estimate regardless of being initialized at the least squares solution or a border minimum. This statement is corroborated by observing the b_1 and b_2 estimates from each IGLS iteration on a realization that terminates outside of the feasibility region. In Figure 48 the b_1 and b_2 estimates from each iteration of the IGLS algorithm initialized at the C border minimum are overlayed on estimates from the IGLS algorithm initialized at the least squares solution. The same realization,

118
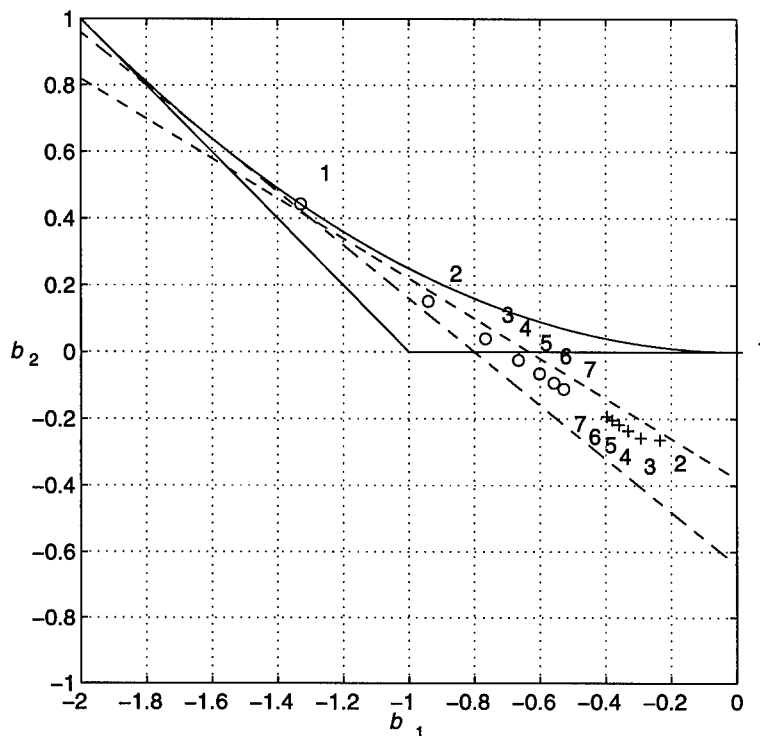
Figure 49.   Realization 2 locations of b_1 and b_2 for each iteration of the IGLS algorithm, initialized at two different points. One starting point was at the least squares solution and the other starting point was at the minimum of the C border. The SNR was 20 dB.

at 20 dB SNR, used in Figure 42 was used for Figure 48. Both applications of the IGLS algorithm terminated at the same location outside of the feasibility region.

The same phenomena for a different realization at 20 dB SNR is demonstrated in Figure 49. In this case, the successive estimates of b_1 and b_2 from the IGLS algorithm, initialized at $b_1 = b_2 = 0$, track towards the feasibility region but fall short of the A border. The successive estimates of b_1 and b_2 from the IGLS algorithm, initialized at the C border, track below the A border towards the same point outside of the feasibility region as well.

From the Monte-Carlo experiment and these sample realizations, we conclude that the IGLS algorithm is a robust estimator for the global minimum of the max-

imum likelihood surface. The algorithm is essentially unaffected by alternative initializations on the feasibility region border and by departures from the feasibility region for subsequent iterations.

A final Monte-Carlo experiment was conducted to investigate the relationship between the global minimum of the maximum likelihood surface and the feasibility region. With a well defined feasibility region and a readily available maximum likelihood cost function—as is the case for the real, two mode problem—implementation of a constrained optimization routine is straight forward. For this reason, the MATLAB **constr** function was utilized on the same realizations of $\bar{y}$ created for the previous Monte-Carlo experiment. The MATLAB **constr** function employs a sequential quadratic programming (SQP) method which can be assisted with closed form expressions for the first derivatives of the cost function and region constraints (20:3—9,3—12). The derivative of the cost function and region constraints are developed in Appendix D. Similar to the alternative initializations of the IGLS algorithm, the SQP algorithm was initialized from the minimums of the A, B, and C borders.

The performance of the SQP algorithm, in terms of inverse MSE and bias is compared to the IGLS algorithm, initialized at the least squares solution, in Figures 50, 51, 52, and 53. From the inverse MSE plots, it appears that the SQP algorithm performs slightly better than the IGLS algorithm at lower SNR, but the behavior of the bias at the same SNR makes the estimates suspect. Also note, the alternative initializations have minimal effect on the SQP algorithm. Regardless, any improved performance is minor when realizing that the IGLS algorithm is not restricted to the real, two mode, problem. Additionally, the IGLS algorithm is computationally more efficient than the SQP algorithm.
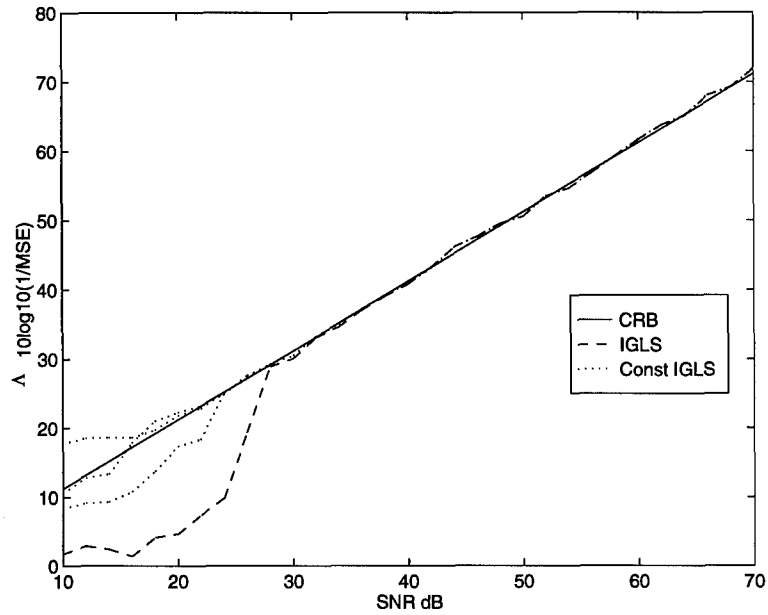
120

Figure 50. The λ_1 inverse MSE of the standard IGLS algorithm, initialized from $b_1 = b_2 = 0$, and the SQP algorithm, initialized from the minimums of the A, B, and C borders. Also plotted is the CRB.

Figure 51. The λ_1 bias of the standard IGLS algorithm, initialized from $b_1 = b_2 = 0$, and the SQP algorithm, initialized from the minimums of the A, B, and C borders.

121
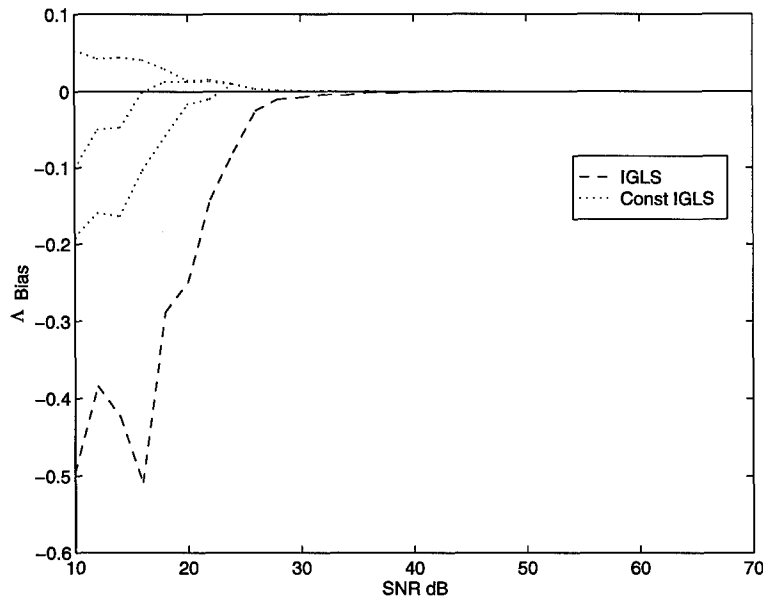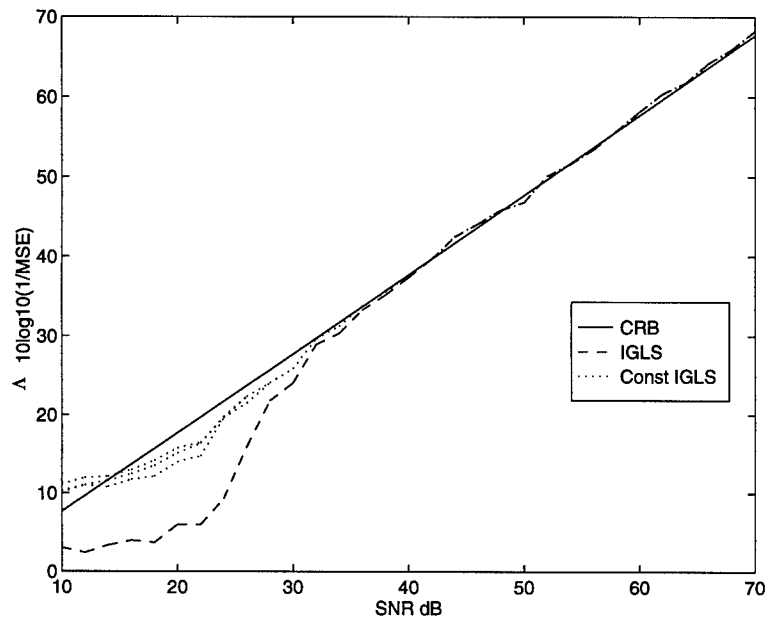
Figure 52.   The λ_2 inverse MSE of the standard IGLS algorithm, initialized from $b_1 = b_2 = 0$, and the SQP algorithm, initialized from the minimums of the A, B, and C borders. Also plotted is the CRB.
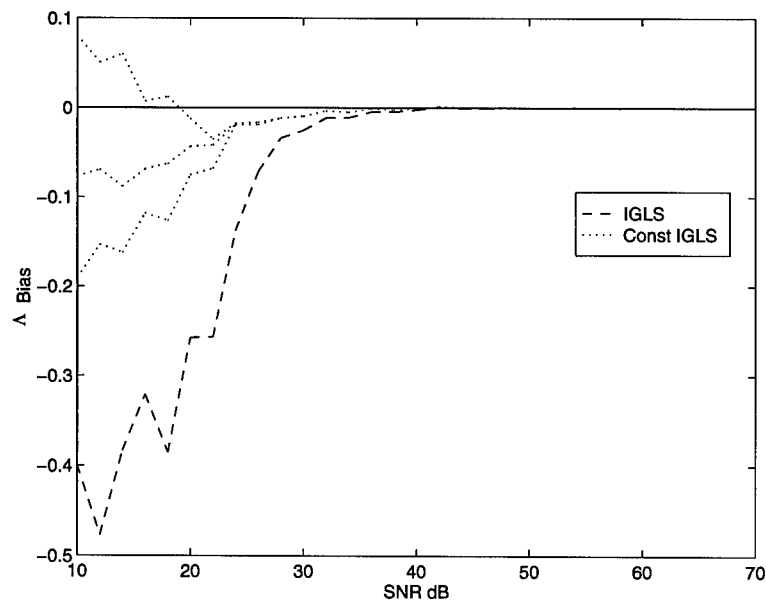
Figure 53.   The λ_2 bias of the standard IGLS algorithm, initialized from $b_1 = b_2 = 0$, and the SQP algorithm, initialized from the minimums of the A, B, and C borders.

5.4 Chapter Conclusion

Beyond the real, two mode, problem, visualization of the mechanics of the IGLS algorithm is difficult. As dimensionality increases, estimation accuracy decreases, but given the performance of the IGLS algorithm for the one and two mode problem, we find no reason to search for a better multi-dimensional maximum likelihood algorithm. We have shown, in isolated scenarios at low SNR, specialized algorithms can be developed to estimate parameters slightly better than the IGLS algorithm, but ultimately, these efforts serve better to exalt the robust capabilities of the IGLS estimator. We move on to the application portion of the dissertation with confidence in the IGLS algorithm.

VI. Deep Level Transient Spectroscopy Application

6.1 Chapter Introduction

As alluded to throughout this dissertation, a particular physics experiment that motivates our real exponential analysis is deep level transient spectroscopy (DLTS). DLTS is a capacitance transient thermal scanning technique used to characterize defects present in semiconductors (45:3023). We begin this chapter by developing the fundamentals of DLTS. Ultimately, we show that the signals from a DLTS experiment are well modeled by superimposed, real, decaying exponentials. Each exponential, and corresponding amplitude coefficient, characterize a defect. Typically the number of detectable defects, N, is small.

Previous approaches to estimating the parameters of DLTS signals include boxcar integrator analysis (45), modulating function waveform analysis (58), and linear prediction (LP) analysis (65). The boxcar integrator approach is the original technique proposed by Lang—the founder of DLTS—in 1974. It is an analog approach and is still popular today. In the DLTS fundamentals section, we show why it is limited. The modulation function approach, like linear prediction, is digital, but unlike linear prediction, is heuristic. Recently, Doolittle and Rohatgi published a series of papers highlighting the benefits of linear prediction analysis over modulating function analysis (10, 11, 12). The LP approach they use for DLTS analysis is taken from the work of Shapiro *et al.* published in 1984. In our research, we have developed concepts, unavailable to Shapiro *et al.*, that significantly improve on the LP approach to DLTS analysis. This work has been accepted for publication by the DLTS community (31). The increased fidelity in our parameter estimates has lead

to defect characterizations that were previously unobserved. These too have been submitted for publication (62).

After developing the fundamentals of DLTS, we review the LP approach of Shapiro *et al.* As we review the previous LP approach to DLTS analysis, we introduce our improvements. We then verify our contribution on actual DLTS signals. At the end of the Chapter, we include additional algorithms, developed by us, for DLTS.

6.2 DLTS Fundamentals

In this section, we identify the DLTS signal and its relationship to the superimposed exponential model. We begin by developing the semiconductor and *pn* junction theory necessary for explaining DLTS. Volumes 1 and 2 of the series by Pierret and Neudeck (55, 53), and the book by Sze (69) are our primary references. The actual DLTS signal development follows the methodology of Lang (45, 46) and Elsaesser (14). Elsaesser clearly extends DLTS to the digital signal processing (DSP) environment.

Consider the *n*-type and *p*-type semiconductor materials of Figure 54a prior to contact. Both *n*-type and *p*-type materials are crystal lattices primarily composed of atoms with four valence electrons such as silicon, Si. The lattice in the *n*-type material is doped with a small quantity of atoms with five valence electrons such as arsenic, As. The lattice in the *p*-type material is doped with a small quantity of atoms with three valence electrons such as boron, B. The doped atoms of both *n*-type and *p*-type materials are called impurities even though their introduction into the crystal lattice is intentional. The four valence electrons of the primary atoms in the lattice tend to share electrons with the other atoms to form covalent bonds. At room temperature, the impurities described are more likely to form covalent bonds

125
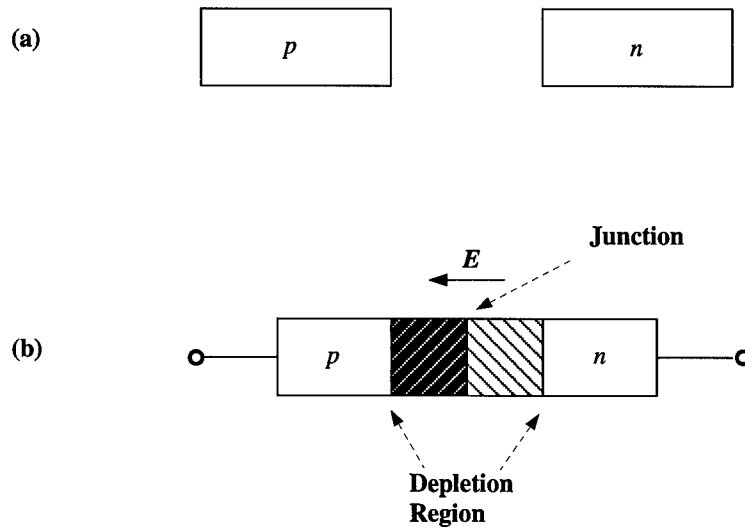
Figure 54. Doped n-type and p-type semiconductor materials (a) before and (b) after forming a pn junction.

with the primary atoms of the lattice than keep their neutral states. Therefore, the four valence electrons of the impurity atoms in n-type material form covalent bonds with the primary atoms of the lattice, and the fifth electrons are "donated" to the conduction band. The impurity atoms in this case are called donors, and the material is called n-type because of the addition of negatively charged carriers. With p-type material, the exact opposite occurs. The impurity atoms, with three valence electrons, "accept" electrons from the primary atoms of the lattice to form covalent bonds. Consequently, positively charged "holes" are created in the valence band. The impurity atoms in this case are called acceptors, and the material is called p-type because of the addition of positively charged carriers. The covalent bonding and resulting charged carriers are illustrated for n-type and p-type silicon in Figure 55.

The impurities described thus far ionize, or emit electrons or holes, at room temperatures. Their ionization energies are just inside the energy band gap boundaries. The ionization energy of donors in n-type material is just below the conduction
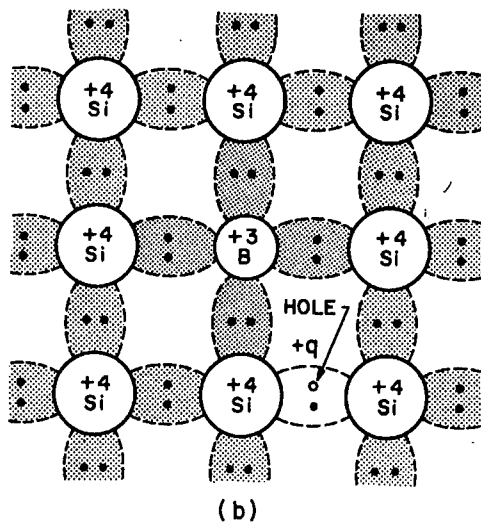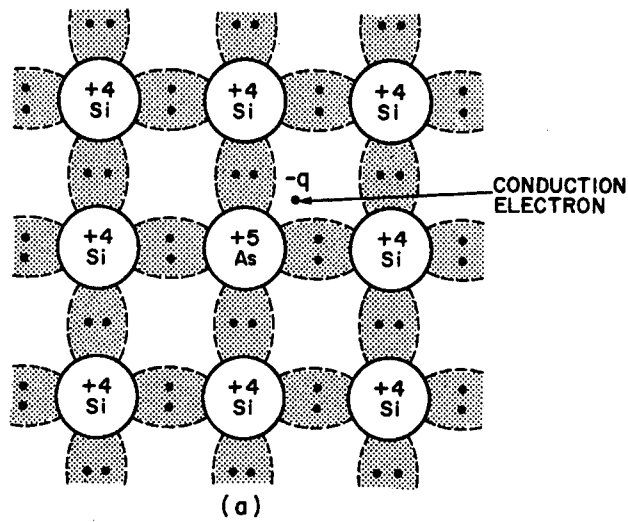
Figure 55. Covalent bonding and resulting charged carriers for (a) n-type silicon and (b) p-type silicon. The figure was obtained from Sze (69).

127

band edge, and the ionization energy of acceptors in p-type material is just above the valence band edge. Consequently, we refer to these intentional impurities as shallow donors and shallow acceptors, respectively. Impurities with more than five or less than three valence electrons can be unintentionally introduced into n-type or p-type material as well. Their ionization energies are typically deep into the energy band gap, and their emission of electrons or holes is significantly more complicated than that of shallow donors and acceptors. Without knowing more about the deep level impurity, it must be assumed that each impurity has the potential to capture, as well as emit, both electrons and holes. Defects in the crystal structure, alone or in combination with deep level impurities, also are assumed to have the potential to capture and emit both electrons and holes. In general, we refer to deep level impurities and crystal structure defects as deep level defects. DLTS is a technique for identifying and characterizing theses deep level defects.

Before explaining DLTS, we need to develop some key concepts for pn junctions. Consider the n-type and p-type materials joined in Figure 54b, and assume there are no defects in either materials. Because of the larger concentration of electrons on the n-type side of the junction versus the p-type side of the junction, electrons diffuse into the p-type material. Likewise, holes from the p-type material diffuse into the n-type material. Recall, the impurity atoms are covalently bonded into the lattice. With the negatively charged carriers leaving the n-type material, a positive space charge forms on the n-type side of the junction. likewise, a negative space charge forms on the p-type side of the junction. This space charge differential creates an electric field, E, in the direction from positive space charge to negative space charge. See Figure 54b. The electric field is opposite the direction of diffusion for both the negatively and positively charged carriers. Therefore, an equilibrium between diffusion and space charge differential is achieved. At equilibrium, the electric

field is significant. It sweeps all the positively and negatively free charged carriers out of a region near the junction. Consequently, this region is dubbed the depletion region. Again, see Figure 54b.

The width, w, of the depletion region is a function of the concentration of shallow donors, N_D, and acceptors, N_A, doped into the n-type and p-type materials. It is also a function of the built in voltage, V_{BI}, due to the space charge differential. The equation for the width is (69:78)

$$(229) \qquad w = \left(\frac{2\epsilon}{q} \left[\frac{N_A + N_D}{N_A N_D} \right] V_{BI} \right)^{\frac{1}{2}}$$

where ϵ is the dielectric constant of the semiconductor and q is the electronic unit charge. If $N_A > N_D$, the depletion region is shifted over to the n-type side of the pn junction. If $N_A >> N_D$ the portion of the depletion region on the p-type side of the junction is so small that the depletion region can be considered to exist only in the n-type material (69:78). The junction is denoted p^+n, and Equation 229 can be simplified to

$$(230) \qquad w = \left(\frac{2\epsilon V_{BI}}{q N_D} \right)^{\frac{1}{2}}.$$

See Figure 56a. In our development, we only consider p^+n junctions. The development for n^+p junctions is complementary and straight forward. We will see that one sided junctions assist DLTS by isolating the location of deep level defects.

By observing Equation 230 we can see a relationship between the width of the depletion region and the voltage. If a reverse bias voltage, V_R, is added to the built in voltage, V_{BI}, the width of the depletion region will increase. See Figure 56b. The
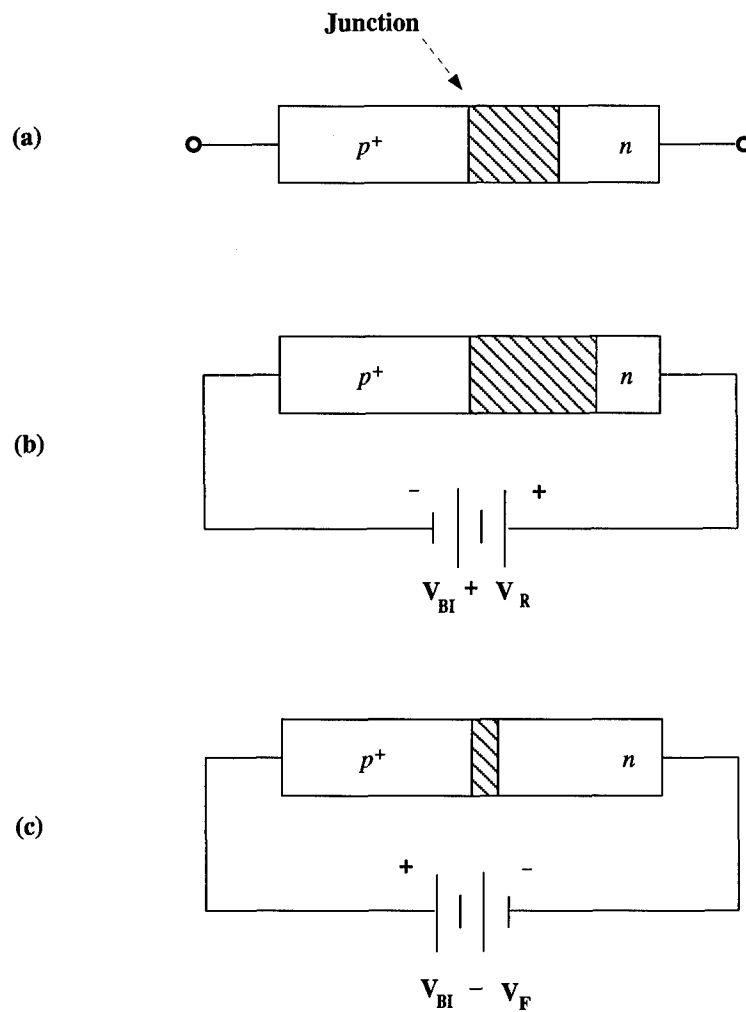
Figure 56. Doped p^+n junction with a variable depletion region width dependent on (a) built in voltage, (b) built in voltage plus reversed biased voltage, and (c) built in voltage minus forward biased voltage.

130

governing equation is (69:80)

$$(231) \qquad w = \left(\frac{2\epsilon(V_{BI} + V_R)}{qN_D} \right)^{\frac{1}{2}}.$$

Similarly, if a forward bias voltage, V_F, is subtracted from the built in voltage, the width of the depletion region will decrease. See Figure 56c. In this case, the governing equation is complicated by the fact that the forward bias voltage forces charged carriers into the depletion region. The actual equation is unnecessary for our development and, for clarity, is omitted.

We have now established that under biased voltage conditions, the width of the depletion region can be varied. Because the shallow donors are easily ionized at room temperature, and because the charged carriers are quickly swept out of the depletion region by the electric field, changes in the bias voltage result in changes in the width of the depletion region with very little temporal lag. This is not the case when deep level defects are present. In DLTS, deep level defects can be characterized by their temporal influence on the width of the depletion region after changes in the bias voltage are made.

As mentioned earlier, we assume each deep level defect has the potential to emit and capture electrons and holes. The propensity of these four events is governed by the conservation of energy. Possible energy interactions with a deep level defect are the absorption or emission of phonons (thermal processes), the absorption or emission of photons (optical processes), and interactions with free conduction band electrons or valence band holes (Auger processes). If bias voltages on the p^+n semiconductor are applied in the dark—which is always the case in DLTS—optical processes can be ignored. Also, under forward biased conditions, Auger-capture processes dominate, over emission processes. If the width of the depletion region is made small enough

131

by forward biasing, the previously depleted region is flooded with free electrons and holes. If the duration of this even is sufficient, we can assume any deep level defect in the n-type material has captured electrons or holes. If the defect primarily captures electrons, we call it an electron trap, and if the defect primarily captures holes, we call it a hole trap. If after forward biasing, the width of the depletion region is opened and held at its built in voltage width, or even further at a reversed biased width, thermal processes will prompt the traps in the depletion region to emit their captured electrons or holes with an emission rate that can be used to characterize the particular defect. If temperature is held constant (isothermal) during emission, the decrease in concentration of traps with captured electron or holes, N_T^c, is modeled with (14:26)

$$(232) \qquad\qquad N_T^c(\mathbb{T}, t) = N_T e^{-rt}$$

where r is the electron or hole emission rate, and N_T is the total concentration of the trap—with or without captured electrons or holes. The temperature dependence, $\mathbb{T}$, is brought in to Equation 232 through the emission rate and is discussed in detail shortly. Finally, because the electric field sweeps the free charged carriers out of the depletion region, re-trapping under reversed bias conditions does not occur.

Consider a p^+n junction with a single deep level electron or hole trap in the depletion region of n-type material. If, under isothermal conditions, we forward bias the semiconductor to fill the trap with electrons or holes, and then reverse bias the semiconductor to insure the trap is well inside the depletion region, the width of the depletion region will change as the capture concentration of the trap changes with time due to the change in depletion charge. By incorporating the capture concentration model, Equation 232, into the reverse biased depletion region width

equation, we can model this change with (14:27)

$$(233) \qquad w(\mathbb{T}, t) = \left(\frac{2\epsilon(V_{BI} + V_R)}{q(N_D + N_T e^{-rt})} \right)^{\frac{1}{2}}.$$

The process just described is the procedure employed in DLTS. The width of the depletion region can be monitored by measuring the capacitance across the region. The measured capacitance is governed by the classic relationship (46:95)

$$(234) \qquad \mathbb{C} = \frac{\epsilon A}{w}$$

where A is the area of the junction. Therefore, the capacitance per unit area, C, for the depletion region is

$$(235) \qquad C(\mathbb{T}, t) = \left(\frac{q\epsilon(N_D + N_T e^{-rt})}{2(V_{BI} + V_R)} \right)^{\frac{1}{2}}.$$

This implies

$$
\begin{aligned}
(236) \qquad C^2(\mathbb{T}, t) &= \frac{q\epsilon N_D}{2(V_{BI} + V_R)} + \frac{q\epsilon N_T e^{-rt}}{2(V_{BI} + V_R)} \\
&= \frac{q\epsilon N_D}{2(V_{BI} + V_R)} + \frac{q\epsilon N_D}{2(V_{BI} + V_R)} \frac{N_T}{N_D} e^{-rt}.
\end{aligned}
$$

In both terms on the right side of Equation 236, we know

$$(237) \qquad \frac{q\epsilon N_D}{2(V_{BI} + V_R)} = \left(\frac{\epsilon}{w} \right)^2.$$

133

This is the square of the capacitance per unit area of the reversed biased depletion region after the trap has emitted all of its captured electrons or holes. Let

$$(238) \qquad C_{ss}^2 = \frac{q\epsilon N_D}{2(V_{BI} + V_R)}$$

where C_{ss} is the steady state capacitance per unit area. Then, we can say

$$(239) \qquad C^2(\mathbb{T}, t) = C_{ss}^2 + C_{ss}^2 \frac{N_T}{N_D} e^{-rt}.$$

In a DLTS experiment, this is the square of the measured capacitance of the depletion region, at a constant temperature, for the time period immediately after the forward bias is removed and the reverse bias is applied. If the trap is primarily an electron trap, the charge distribution is such that $\frac{N_T}{N_D}$ is negative and the capacitance transient will increase to C_{ss}. If the trap is primarily a hole trap, the charge distribution is such that $\frac{N_T}{N_D}$ is positive and the capacitance transient will decrease to C_{ss}.

Multiple hole or electron traps are linearly modeled by (14:34)

$$(240) \qquad C^2(\mathbb{T}, t) = C_{ss}^2 + \sum_{n=1}^{N} C_{ss}^2 \left(\frac{N_T}{N_D}\right)_n e^{-r_n t}.$$

If discretized, this expression can fit the form of the superimposed exponential signal model used throughout this dissertation. Let

$$(241) \qquad s[m] = c_0 + \sum_{n=1}^{N} c_n \lambda_n^m$$

where

$$c_0 = C_{ss}^2$$

$$c_n = C_{ss}^2 \left(\frac{N_T}{N_D}\right)_n$$

(242)
$$\lambda_n = e^{-r\delta}$$

and δ is the sample interval. Also, for now, assume the baseline constant c_0 is the amplitude coefficient of an added exponential with $\lambda_0 = 1$. The estimation algorithms developed thus far can provide estimates of c_0, c_n, and λ_n. In the latter sections of this chapter, we will address more appropriate approaches to estimating the intercept c_0. Nevertheless, with estimates of c_n and λ_n, we can solve for the trap concentration N_T and the emission rate, r, of each defect. With trap emission rate estimates obtained over a range of temperatures, we can further characterize the trap. To do so, we first need to develop more background on electron and hole trap emission rates.

In the DLTS experiment, we can limit our discussion to thermal emission rates which can be modeled as a function of temperature (14:13-17). Because the thermal emission rate has dependence on the thermal capture rate, we can show

(243)
$$r_{n,p} = Q_{n,p} \mathrm{T}^2 \sigma_{n,p} e^{-\Delta E_{n,p} \frac{1}{k\mathrm{T}}}$$

where the n subscript indicates electron trap, the p subscript indicates hole trap, $Q_{n,p}$ are assumed constants, $\sigma_{n,p}$ are thermal capture cross sections, k is the Boltzmann constant, and ΔE_n is the separation energy between the lower edge of the conduction band and the energy level of an electron trap ($\Delta E_n = E_c - E_T$), and ΔE_p is the

135

separation energy between the energy level of a hole trap and the upper edge of the valence band ($\Delta E_p = E_T - E_v$).

Lang's analog DLTS analysis technique (45), which is also explained shortly, estimates the inverse of the trap emission rate, $\tau = \frac{1}{r}$. Therefore, estimating τ, rather than r, has become the convention. From the thermal emission rate equation, consider

(244)
$$\tau_{n,p} = \left(Q_{n,p}\mathbb{T}^2\sigma_{n,p}\right)^{-1} e^{\Delta E_{n,p}\frac{1}{kT}}$$
$$\Rightarrow \quad \mathbb{T}^2\tau_{n,p} = \left(Q_{n,p}\sigma_{n,p}\right)^{-1} e^{\Delta E_{n,p}\frac{1}{kT}}$$
$$\Rightarrow \quad \ln\left(\mathbb{T}^2\tau_{n,p}\right) = -\ln\left(Q_{n,p}\sigma_{n,p}\right) + \Delta E_{n,p}\frac{1}{kT}.$$

The last expression has linear form. If isothermal capacitance transient signals are created and measured at various temperatures, and the emission rates are estimated for each signal, $\ln\left(\mathbb{T}^2\tau_{n,p}\right)$ can be plotted versus $\frac{1}{kT}$. The slope of the plot corresponds to ΔE, which can be solved for the energy level of the trap, and the intercept of the plot corresponds to $\ln\left(Q_{n,p}\sigma_{n,p}\right)$, which can be solved for the thermal capture cross section of the trap. Plots of this nature are called Arrhenius plots.

When Lang introduced DLTS, high speed digitizers were not available for digital analysis. Despite this, he proposed an easy to implement analog approach for estimating the τ of various isothermal transients and subsequent Arrhenius plotting. At this time, Lang's approach is developed to identify its limitations and to give physical insight into the DLTS problem. At the end of this chapter, Lang's approach is used to highlight the difficulty of obtaining adequate estimates on actual DLTS data.

First, Lang assumes there is only one trap in the depletion region and that its concentration, N_T, is much less than the concentration of the shallow donors, N_D.
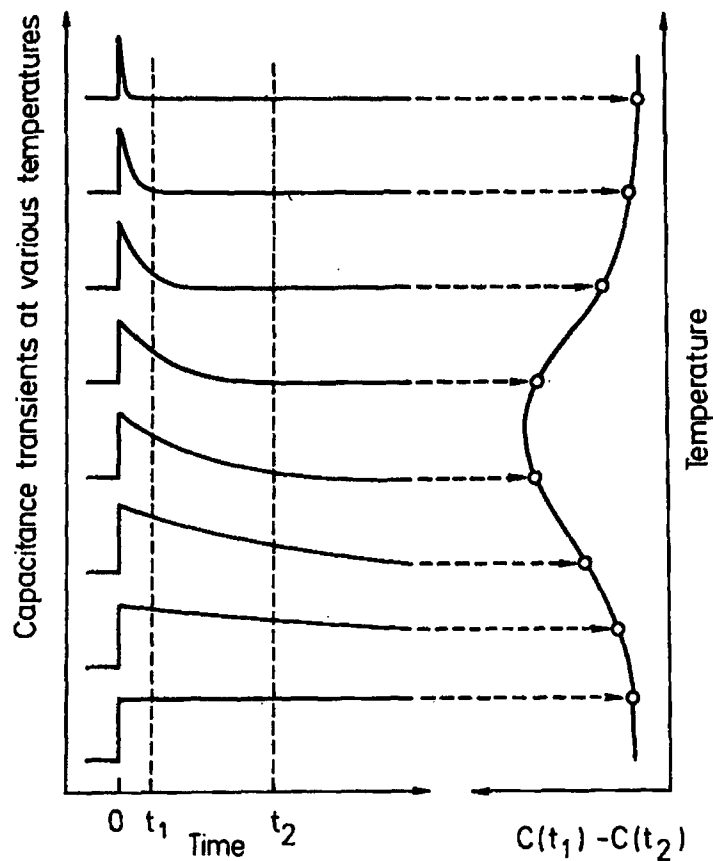
Figure 57. Capacitance transients for a single hole trap in the depletion region of a p^+n junction at various temperatures. Also shown is the $\Delta C(\mathbb{T})$ for t_1 and t_2 in each transient. The figure was obtained from Lang (46).

Then, a binomial expansion of Equation 239 leaves

$$(245) \qquad C(\mathbb{T}, t) = C_{ss} + C_{ss}\frac{N_T}{2N_D}e^{-\frac{t}{\tau}}.$$

The left side of Figure 57 is a plot of the capacitance transients for a hole trap at various temperatures. As expected with the thermal emission rate equation, τ decreases as temperature increases. Notice, from the right side of Figure 57, the

137

change in capacitance between times t_1 and t_2,

$$(246) \qquad \Delta C(\mathbb{T}) = C(\mathbb{T}, t_1) - C(\mathbb{T}, t_2),$$

is minimal at low temperatures, minimal at high temperatures, and maximal at one temperature in the middle. With the capacitance transient model for a single trap, Equation 245, we can obtain an analytical expression for $\Delta C(\mathbb{T})$ where

$$
\begin{aligned}
\Delta C(\mathbb{T}) &= C_{ss} + C_{ss}\frac{N_T}{2N_D}e^{-\frac{t_1}{\tau}} - C_{ss} - C_{ss}\frac{N_T}{2N_D}e^{-\frac{t_2}{\tau}} \\
(247) \qquad &= C_{ss}\frac{N_T}{2N_D}e^{(\frac{t_2}{\tau} - \frac{t_1}{\tau})}.
\end{aligned}
$$

From the previous development, we know $\Delta C(\mathbb{T})$ is a function of temperature through the inverse of the thermal emission rate, τ. By differentiating $\Delta C(\mathbb{T})$ with respect to τ, and equating to zero, we see

$$(248) \qquad \tau_{max} = \frac{t_2 - t_1}{\ln(t_2/t_1)}.$$

Therefore, referring to Figure 57, the temperature associated with $\Delta C(\mathbb{T})_{max}$ corresponds to a singe trap capacitance transient with $\tau_{max} = \frac{t_2 - t_1}{\ln(t_2/t_1)}$.

By changing t_1 and/or t_2, the peak for $\Delta C(\mathbb{T})_{max}$ changes and a new temperature associated with $\Delta C(\mathbb{T})_{max}$ corresponds to a new $\tau_{max} = \frac{t_2 - t_1}{\ln(t_2/t_1)}$. The process can be repeated as many times as necessary to plot $\ln(\mathbb{T}^2\tau)$ versus $\frac{1}{k\mathbb{T}}$ in an Arrhenius plot. This approach is known as the rate window approach. By anticipating the t_1 and t_2 samples necessary to produce an adequate Arrhenius plot, double boxcar integrators can be placed in line to calculate $\Delta C(\mathbb{T})$ for every transient. The rate

window approach is quick and simple, and yields single trap parameter estimates without a digitizer.

However, the rate window approach does not have multi-trap capability, and as will be shown at the end of this chapter, degrades with noise. The advent of high speed digitizers and improvements in digital signal processing, namely linear prediction, have provided some relief to these problems. In the next section, we review an early LP approach to DLTS analysis and present our improved iterative generalized least squares (IGLS) approach.

6.3 Linear Prediction for DLTS Analysis

As mentioned in the beginning of this chapter, Shapiro *et al.* (65) introduced an LP based approach for DLTS analysis in 1984. Their proposed algorithm is essentially the overmodeled least square algorithm (OLS) developed in Chapter III. In DLTS, Shapiro *et al.* recognized that because the concentrations of any deep level traps, N_T, are usually much smaller than the concentration of the donors, N_D, the baseline constant, c_0, is usually much larger than the amplitude coefficients, c_n. Treating c_0 as the amplitude coefficient of an additional exponential, λ_0, without properly constraining λ_0 to equal one, seriously degraded the performance of their algorithm. They improved their estimates by differencing the discrete transient at successive points prior to implementing OLS (65:3457-3458). In a sense, they obtained estimates from the derivative of the data, which eliminated the constant. At the end of this section, we take a similar tack, but first we must recognize another approach.

In 1992, Doolittle and Rohatgi (10) introduced hardware to estimate and re-move (null) the baseline constant from the capacitance transients. They then applied

the OLS algorithm of Shapiro *et al.* and obtained good results. For comparison, we simulated capacitance transient data without a baseline constant and tested the performance of the IGLS algorithm against the OLS algorithm.

It should be noted, that—from the work of Kumaresan and Tufts (72)—Shapiro *et al.* (65) and Doolittle and Rohatgi (11) both assumed any extraneous λ_n estimates would be complex and used this as a criterion for actual λ_n acceptance. We find this criterion to be insufficient and have added the energy sort described in Chapter III to all OLS estimates.

For our simulation, one hundred data point realizations of $\bar{y}$ were created with randomly generated Gaussian noise and the underlying signal parameters: $c_1 = -1$, $c_2 = -.5$, $\lambda_1 = .8$, $\lambda_2 = .9$. The noise contributions were governed by SNR, in dB, under the standard formula. For each realization, 10 iterations of the IGLS algorithm and three separate runs of the OLS algorithm, with prediction orders $P = 20, 30$ and 40, were performed.

Figures 58, 59, 60, and 61 are inverse MSE and bias plots for comparing the performance of the IGLS algorithm with the OLS algorithm. The MSE and bias were calculated after 200 Monte-Carlo simulations for each SNR. From the figures, we note IGLS consistently provides lower variance estimates than OLS, regardless of prediction order, over a usable range of SNRs. Also, above an SNR threshold— approximately 45 dB—IGLS essentially attains the CRB. The performance of OLS decreases as SNR increases because the overmodeling becomes inappropriate.

Given the experiments we performed in Chapter III on the sum of two complex exponentials, these results for the sum of two real exponentials are expected. The IGLS algorithm should replace the OLS algorithm for linear prediction based DLTS when the baseline constant, c_0, is nulled with hardware. However, we claim, the

Figure 58.   The λ_1 inverse MSE of the IGLS and OLS estimators over a range of SNRs. Also plotted is the CRB.

Figure 59. The λ_1 bias of the IGLS and OLS estimators over a range of SNRs.
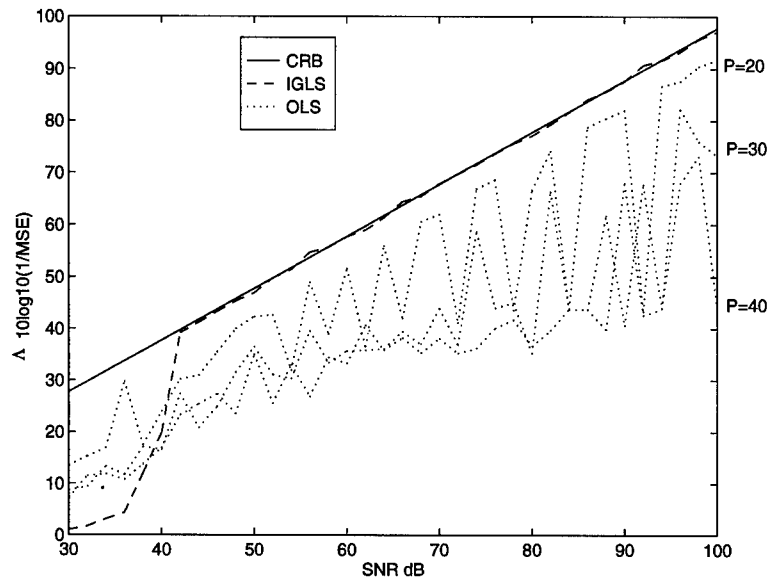
141

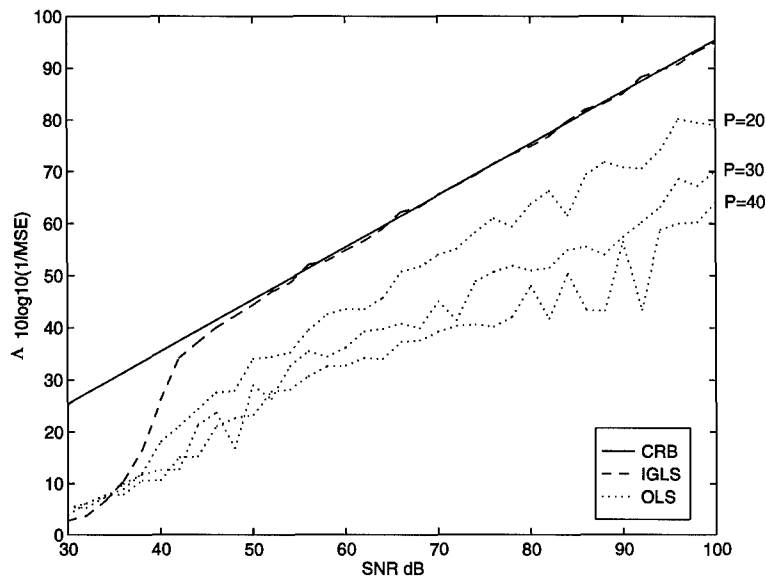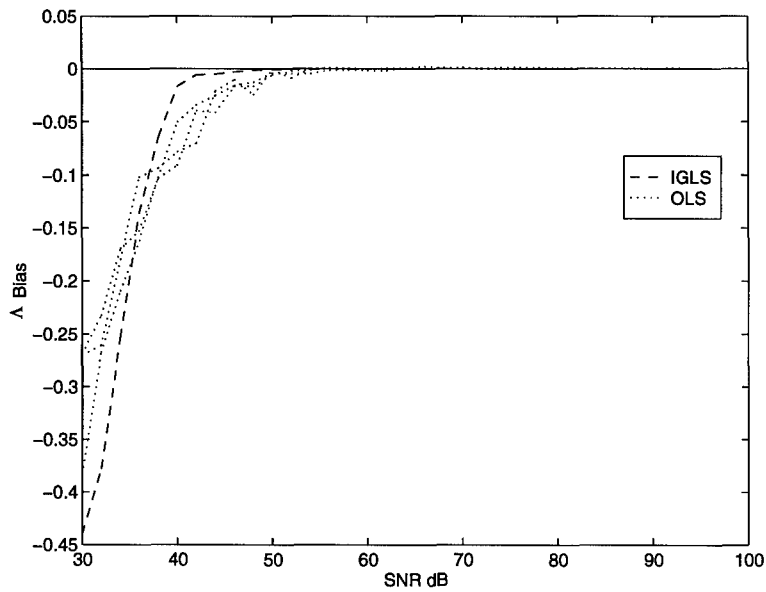Figure 60. The λ_2 inverse MSE of the IGLS and OLS estimators over a range of SNRs. Also plotted is the CRB.
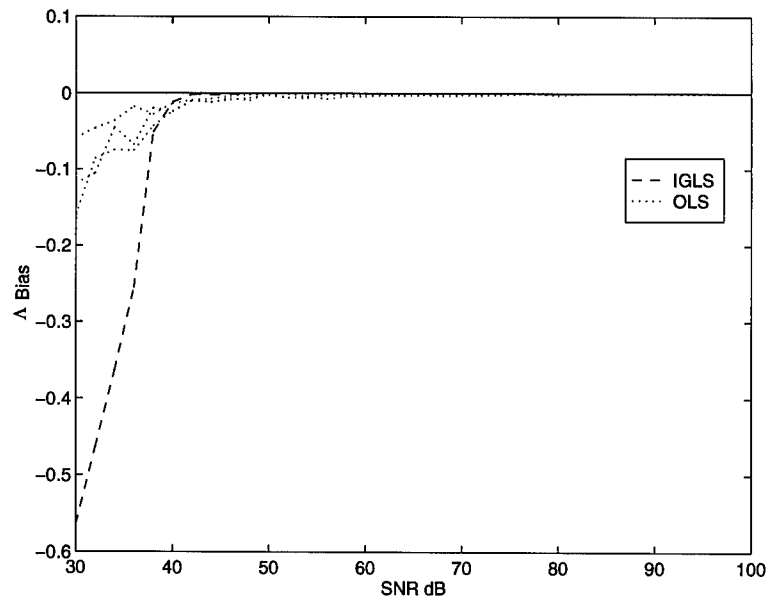
Figure 61. The λ_2 bias of the IGLS and OLS estimators over a range of SNRs.

need for the nulling hardware is unnecessary. Unlike the OLS algorithm, the IGLS algorithm can be accurately extended to incorporate baseline constant estimation. The extension is not obvious when the algorithm is derived form the maximum likelihood (ML) methodology, but with the introduction of our linear prediction methodology, the ML cost function for superimposed exponentials with a baseline constant follows nicely.

In our development, we keep the differencing approach suggested by Shapiro *et al.* (65). Incidentally, differencing was also recommended by Hildebrand (22:458–462), in 1956, for the general superimposed exponential problem with a baseline constant. Unfortunately, both Hildebrand and Shapiro *et al.* did not account for the effects of noise in their linear prediction methodology. In this dissertation, we have established the inferior performance of LP based estimators without accurate noise analysis. The performance of differenced LP based estimators without accurate noise analysis is even more degraded. As alluded to before, with our methodology, properly accounting for the effects of noise in differenced linear prediction is straight forward.

Let

$$
\begin{aligned}
y_d[m] &= y[m] - y[m+1] \\
&= s_d[m] + w[m] - w[m+1], \qquad m = 0, 1, ..., M-2
\end{aligned}
$$

(249)

where

(250)
$$
s_d[m] = \sum_{n=1}^{N} (1 - \lambda_n) c_n \lambda_n^m.
$$

The noiseless signal, $s_d[m]$ is also a solution to a difference equation with constant coefficients. Therefore, if we let

$$(251) \qquad\qquad s_d[m] = y_d[m] - w[m] + w[m+1]$$

we know

$$(252) \qquad b_0 y_d[m] + b_1 y_d[m-1] + \ldots + b_N y_d[m-N] = e_d[m]$$

where

$$
\begin{aligned}
e_d[m] &= b_0 w[m] + b_1 w[m-1] + \ldots + b_N w[m-N] - \\
(253) &\qquad -b_0 w[m+1] - b_1 w[m] - \ldots - b_N w[m-N+1].
\end{aligned}
$$

The LP error $e_d[m]$ has more elements than the previous $e[m]$ of Chapter III, but it is still composed of a tractable sum of weighted Gaussian random variables. Under an overdetermined set of linear equations, let

$$(254) \qquad\qquad Y_d \bar{b} = \bar{e}_d$$

where

$$
\begin{aligned}
Y_d &= \begin{bmatrix} \bar{y}_{d0} & \bar{y}_{d1} & \cdots & \bar{y}_{dN} \end{bmatrix} \\
\bar{y}_{dn} &= \begin{bmatrix} y_d[N-n] & y_d[N+1-n] & \cdots & y_d[M-2-n] \end{bmatrix}^T \\
\bar{b} &= \begin{bmatrix} b_0 & b_1 & \cdots & b_N \end{bmatrix}^T \\
(255) \qquad \bar{e} &= \begin{bmatrix} e_d[N] & e_d[N+1] & \cdots & e_d[M-2] \end{bmatrix}^T.
\end{aligned}
$$

144

The data matrix, Y_d, has dimensions $(M - N - 1) \times (N + 1)$ and Toeplitz structure. The LP error vector $\bar{e}_d$ is distributed with mean vector zero and covariance matrix $\sigma^2 R_d$. If we define the $(M - N - 1) \times (M - 1)$ Toeplitz matrices

$$(256) \qquad B_1 = \begin{bmatrix} b_0 & b_1 & \cdots & b_N & 0 & \cdots & 0 \\ 0 & b_0 & b_1 & \cdots & b_N & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & b_0 & b_1 & \cdots & b_N \end{bmatrix}$$

and

$$(257) \qquad B_2 = \begin{bmatrix} 0 & b_0 & b_1 & \cdots & b_{N+1} & b_N & 0 & \cdots & 0 \\ 0 & 0 & b_0 & b_1 & \cdots & b_{N+1} & b_N & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \vdots & & \ddots & \ddots & \ddots & \ddots & \ddots & b_N \\ 0 & 0 & \cdots & \cdots & 0 & b_0 & b_1 & \cdots & b_{N+1} \end{bmatrix},$$

then the differenced LP error covariance matrix is obtained with

$$(258) \qquad R_d = 2B_1 B_1^T - B_1 B_2^T - B_2 B_1^T$$

and the generalizes least squares (GLS) estimator becomes

$$(259) \qquad \bar{b} = \begin{bmatrix} 1 \\ -(Y_d^T R_d^{-1} Y_d)^{-1} Y_d^T R_d^{-1} \bar{y}_{d0} \end{bmatrix}.$$

As before, an iterative algorithm naturally follows. After rooting, the amplitude coefficients, c_0 and c_n, can be estimated from an expanded pseudo-inverse

145

estimator

$$(260) \qquad \begin{bmatrix} c_0 \\ \bar{c} \end{bmatrix} = (V_{c_o}^T V_{c_o})^{-1} V_{c_o}^T \bar{y}$$

where the matrix V_{c_o} consists of an M dimensional column of ones appended in front of the standard Vandermonde matrix, V.

In simulations, realizations of $\bar{y}$ were created with the same parameters as earlier in this section, only a baseline constant, c_0, was added at an elevated value of 50. The differenced iterative generalized least squares (DIGLS) estimator was tested against a differenced overmodeled least squares (DOLS) estimator with an energy sort. The experiment was conducted identical to that of the IGLS and OLS estimators in this chapter.

Figures 62, 63, 64, and 65 are inverse MSE and bias plots for comparing the performance of the DIGLS algorithm with the DOLS algorithm. Similar to the experiment without c_0, DIGLS consistently performs better than DOLS over a usable range of SNRs and attains the CRB. Figure 66 illustrates the difference between DIGLS and IGLS for the λ_1 estimates. We see that adding the c_0 constant does not significantly effect the accuracy of the iterative estimators. Note, the error associated with nulling is not reflected in the IGLS inverse MSE plot. Obtaining all estimates with software may indeed be more accurate than introducing hardware for assisting in the task. In the next section, the capabilities of the DIGLS algorithm are verified on actual DLTS signals. Like many DLTS facilities, we do not utilize hardware for c_0 nulling. Therefore, we use the differenced approach.
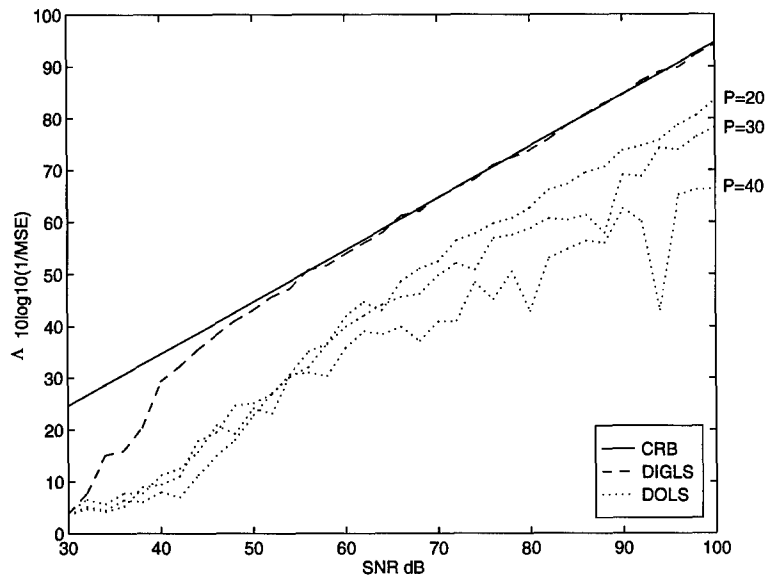
Figure 62. The λ_1 inverse MSE of the DIGLS and DOLS estimators over a range of SNRs. Also plotted is the CRB.

Figure 63. The λ_1 bias of the DIGLS and DOLS estimators over a range of SNRs.
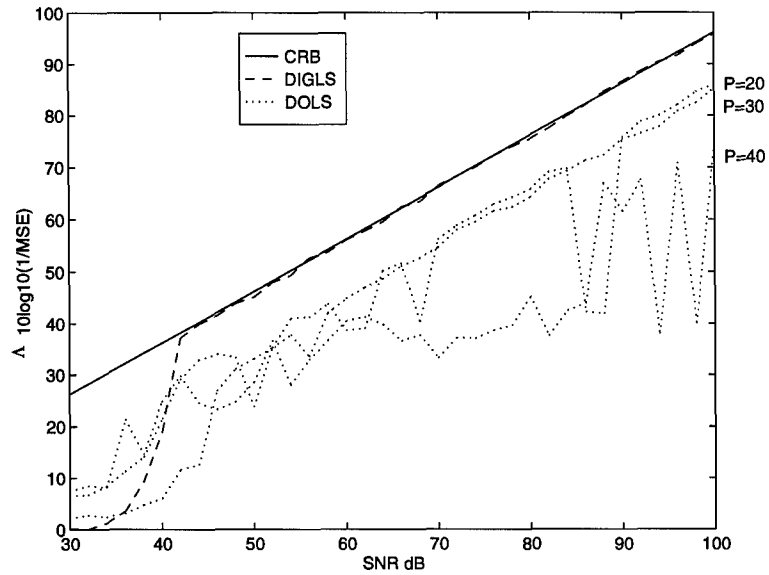
147

Figure 64.   The λ_2 inverse MSE of the DIGLS and DOLS estimators over a range of SNRs. Also plotted is the CRB.
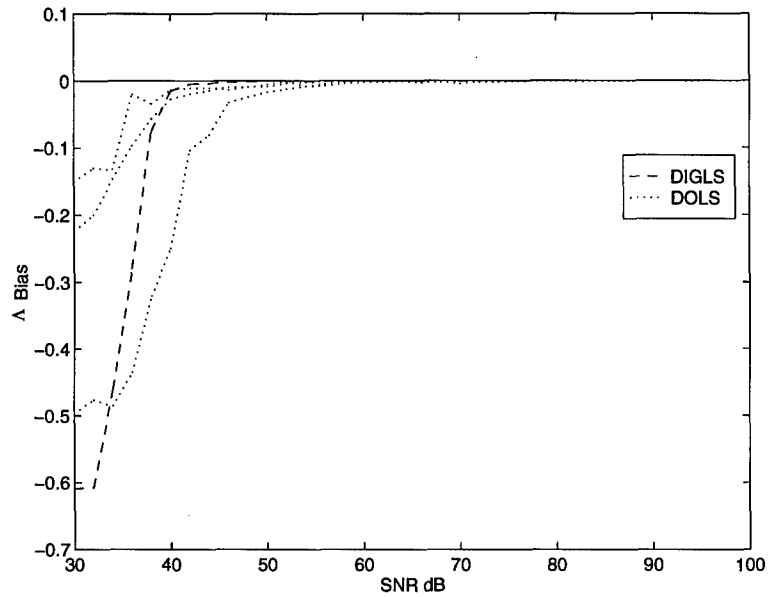
Figure 65. The λ_2 bias of the DIGLS and DOLS estimators over a range of SNRs.
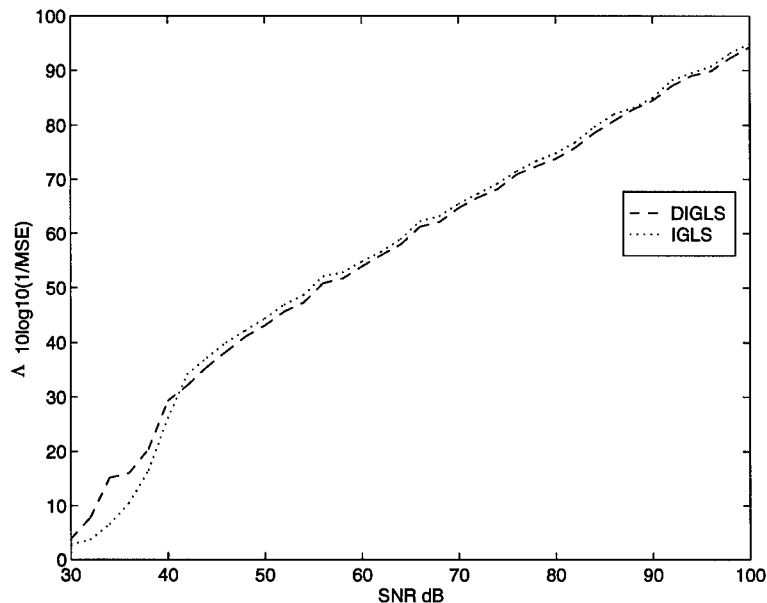
Figure 66. Inverse MSE versus SNR plot for λ_1. DIGLS and IGLS compared on simulated data with and without a baseline constant, respectively.

6.4 DLTS Experiments

The DIGLS and DOLS estimators were compared on the capacitance transients of two different semiconductors. First, the arsenic anti-site defect, As_{Ga}, in GaAs— namely the EL2 level—was considered. The EL2 deep level was selected for initial verification of the DIGLS estimator because of its well known trapping kinetics and deep level parameters (43, 73). The EL2 capacitance transient signals were recorded in both maximal and artificially reduced SNR scenarios. They were digitized at isothermal increments of 4 K using conventional constant-voltage biasing (CVDLTS) of Ti/Pt/Au Schottky diodes formed on low-temperature grown, Sb-doped, GaAs substrate material. A large CVDLTS reverse bias voltage of -5.0 volts was utilized to maximize the width of the depletion region and minimize the non-exponential effects due to a free carrier tail at the edge of the depletion region. It has been shown, that the large reverse bias voltage does not effect the expected separation

149

energy, ΔE_n, in EL2 (5). Forward bias filling pulse heights and pulse durations, used to completely saturate the traps, were 1.5 *volts* and 10 *ms* respectively. The resulting capacitance transients were sampled at 100 μs intervals with 500 sample points recorded to yield a 50 *ms* digitized transient period. Noise averaging at each isothermal step was nominally high (100 transients) to facilitate the zero mean Gaussian assumption. Artificial SNR reduction was accomplished using a double correlated mode of DLTS operation (DDLTS) to control and reduce the volume of space charge from which emission occurs. Specifically, the difference signal, resulting from the capacitive decay for differing forward bias filling pulse heights under the same measurement bias, was analyzed. Variations in the forward bias difference, $|VF1 - VF2|$, resulted in controllable variations of the SNR.

The Arrhenius plot shown in Figure 67 is for comparing the performance of the DIGLS and DOLS algorithms on the EL2 deep defect. In all cases, the DIGLS estimator was implemented with 10 iterations, and the DOLS estimator was implemented with a prediction order $P = 30$ and an energy sort. The data shown in the figure correspond to the solutions for a single mode ($N = 1$) analysis using the DIGLS and DOLS estimators under the maximal SNR scenario. Immediately apparent from the Arrhenius plots is the very significant extension of the range of useful emission rate data in the case of the DIGLS estimator. In other words, the DIGLS estimator is seen from the figure to be much more robust at the lower temperatures where the capacitance transient is significantly truncated, or equivalently, is characterized by slower emission rates. A more important point to make regarding the data shown in the figure is that the separation energy, obtained from the resulting linear slope of the DIGLS emission rate estimates, equals 791 *meV*. This is in much better agreement with accepted transient spectroscopy values than that of the DOLS estimates. The best fit to the linear region data obtained from the DOLS

150
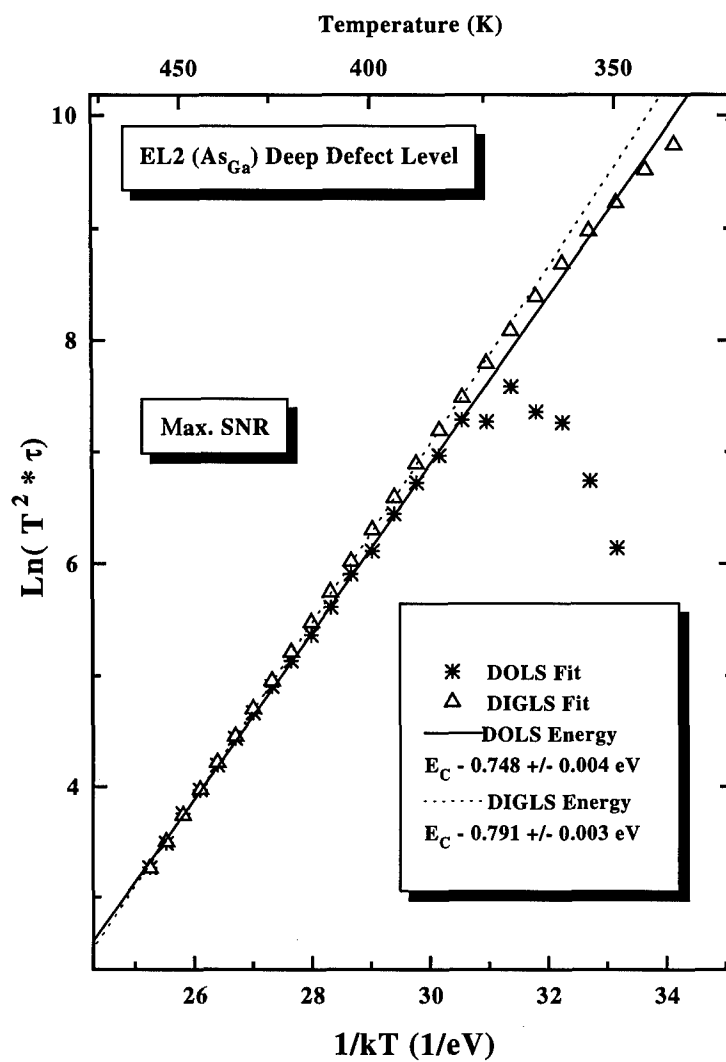
Figure 67. Arrhenius plot for EL2 with maximal SNR.

algorithm yields a separation energy of 748 meV. This underestimates even the low end separation energies obtained using alternative Hall effect measurements (48).

Figure 68 illustrates more concisely the superiority of the improved DIGLS linear prediction algorithm. The data represents the DIGLS and DOLS estimates on EL2 transients with artificially reduced SNR. The DIGLS algorithm slightly underestimates the EL2 energy level (-10%) but the DOLS estimator is completely ineffective. For confidence in our experiment, we have shown in our article (31) that the OLS algorithm, implemented by Doolittle and Rohatgi (12), obtains inferior estimates to our DOLS estimator under equivalent EL2 conditions.

Finally, the extent of signal degradation and the significance of the improvements afforded by the DIGLS estimator can be fully appreciated by observing the rate window plots of the maximal and degraded SNR scenarios. They are illustrated in Figures 69 and 70 respectively. By convention, six rate windows were applied from each end of the transients with t_2/t_1 ratios of 2, 5, and 10, respectively. The degraded SNR rate window plot is unintelligible.

After verification under simulated and controlled DLTS experimental conditions, the DIGLS and DOLS estimators were compared on a more general problem potentially involving multi-mode transient decay. The 6H polytypic modification of the SiC material system was analyzed. In n-6H-SiC bulk substrate material, we anticipate a DLTS signal with two resolvable separation energies. To date, multi-mode DLTS signals have not been resolved. Only heuristic arguments and postulation of their existence, based on experimental data yielding closely spaced energy levels, have been suggested (16).

Figure 71 is a rate window plot for n-6H-SiC material. Typically, broad overlapping peaks and shoulders are observed in the rate widow plots of multi-mode
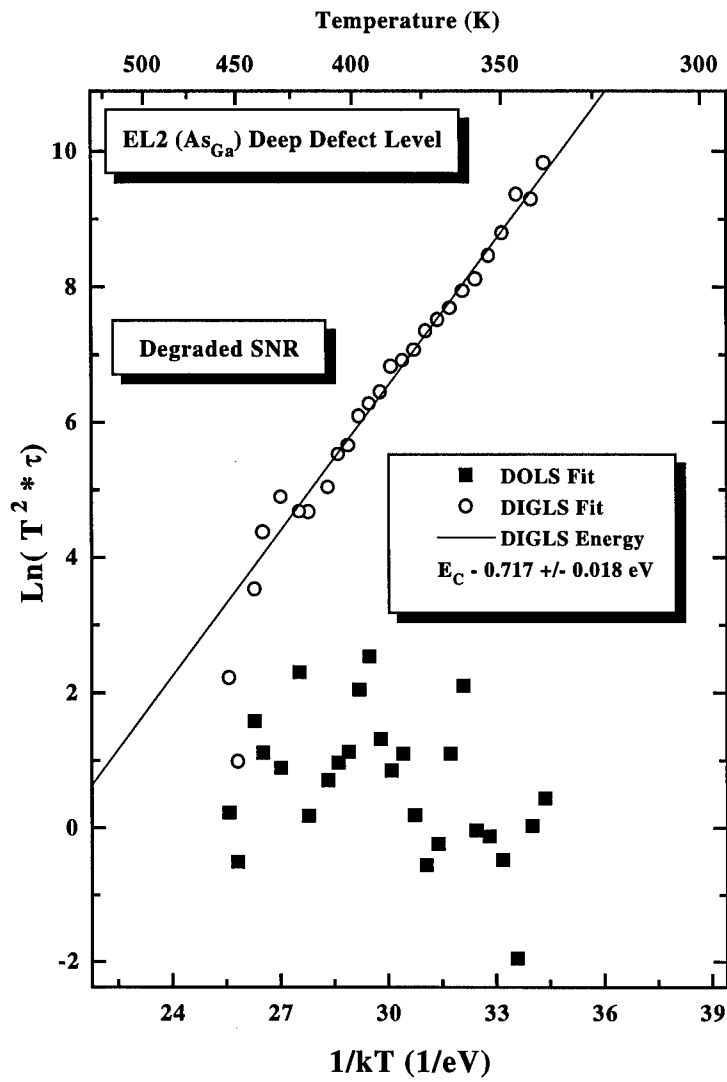
Figure 68. Arrhenius plot for EL2 with artificially degraded SNR.
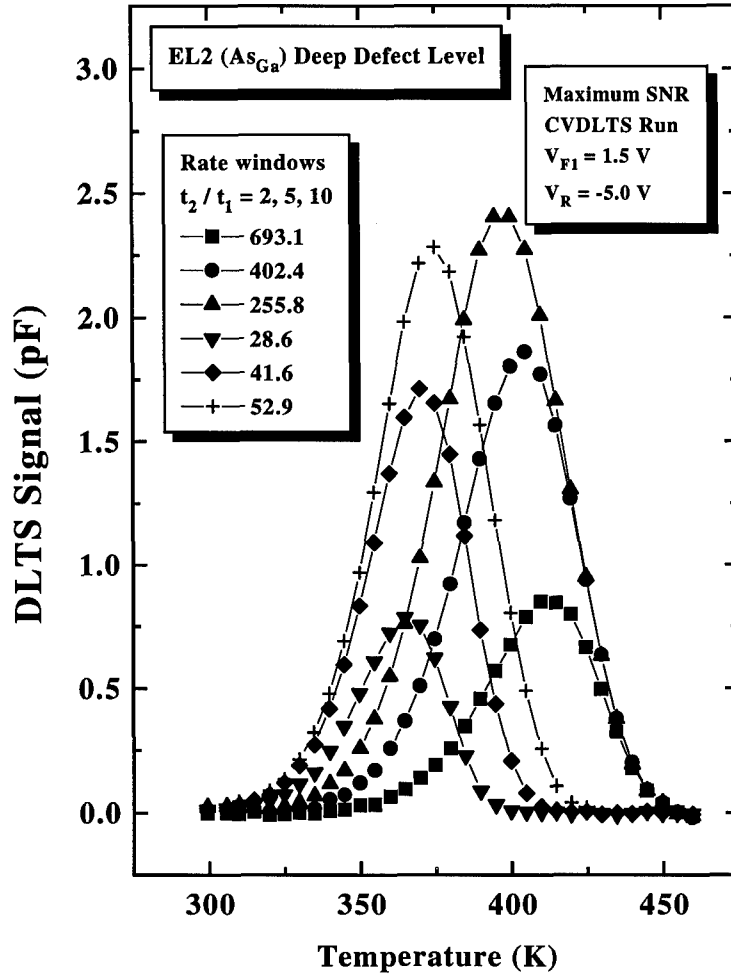
Figure 69. Rate window plot for EL2 with maximal SNR.
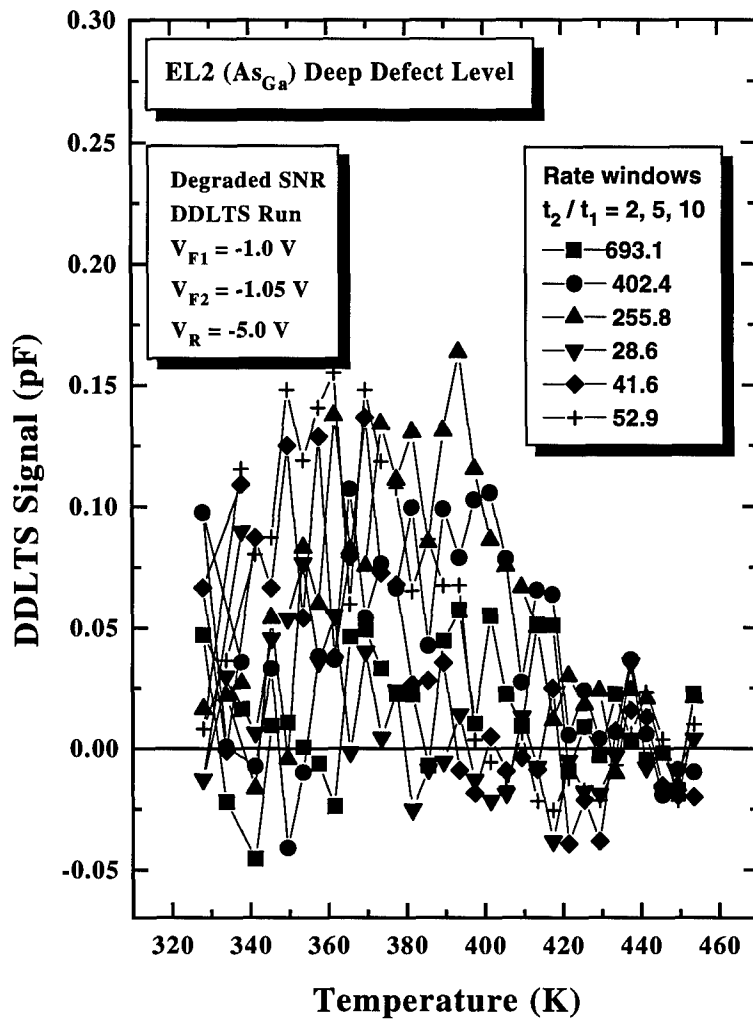
Figure 70. Rate window plot for EL2 with artificially degraded SNR.

155
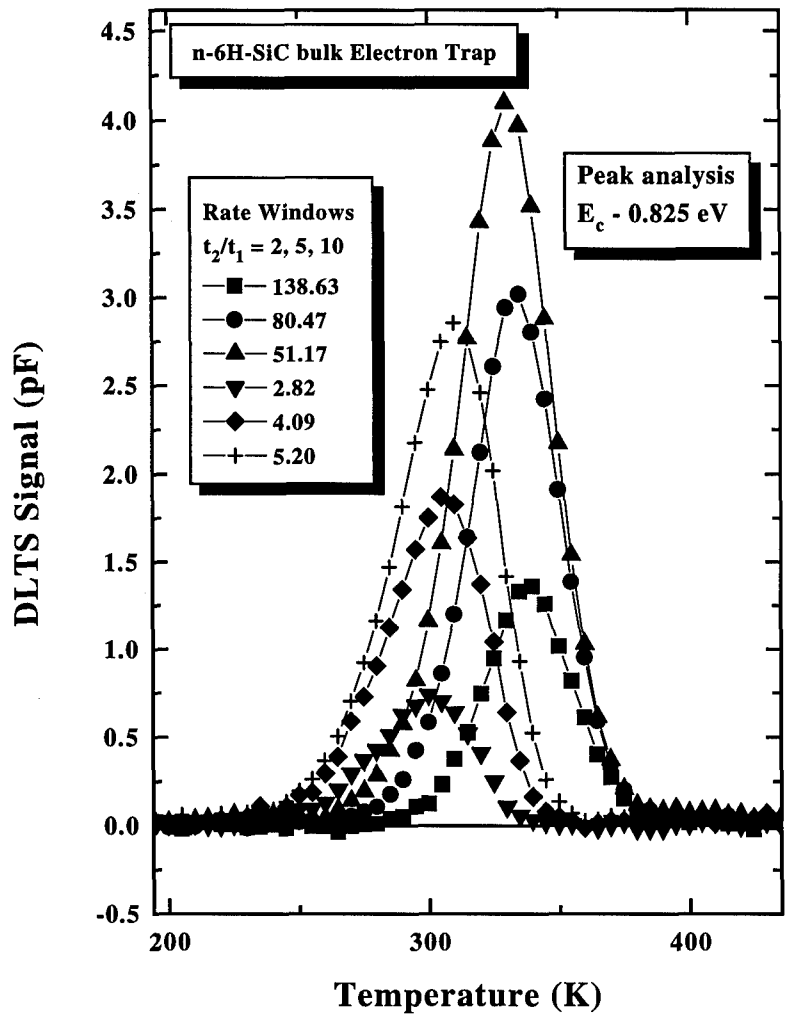
Figure 71. Rate window plot for n-6H-SiC.

156

materials. This is not the case in Figure 71. Similarly, the DOLS estimator only detected one decay mode. However, our DIGLS estimator revealed a definite indication of the presence of two exponential decay modes. Figure 72 illustrates the resulting Arrhenius analysis of the DIGLS fitted decay time constants showing convincing evidence for a two mode transient decay. The separation energies of 509 and 543 meV obtained from the linear slope are seen to deviate significantly from the rate window peak analysis estimate of 760 meV. This is not surprising if we recall that the classical peak analysis method is invalid if more than a single decay mode exists (45). The results presented, may be the first explicit DLTS data supporting the existence of SiC deep level energetic pairs for an apparent spectral feature indicating single mode decay.

The apparent success obtained in resolving closely spaced levels of the electron traps in n-6H-SiC led us to apply the DIGLS estimator to a commonly observed hole trap in p-6H-SiC. This defect has been observed by us to be present in most substrate wafer material and is readily formed upon ion-implantation into epitaxial material (63). Figure 73 is the rate window plot for the CVDLTS transients. In the figure, the shoulder (marked by the large arrow) is indicative of multi-mode decay. However, both the IGLS and DOLS estimates for the transients indicate the presence of only a single exponential component with a separation energy of 861 meV. Figure 74 illustrates the fitted data for this defect. The striking feature of the DIGLS estimator is the extension of useful transient data by almost 50 K, or 10 transient emission rate data points, on the low temperature side. Consequently, characterization of the defect level in p-6H-SiC can be made with increased confidence.
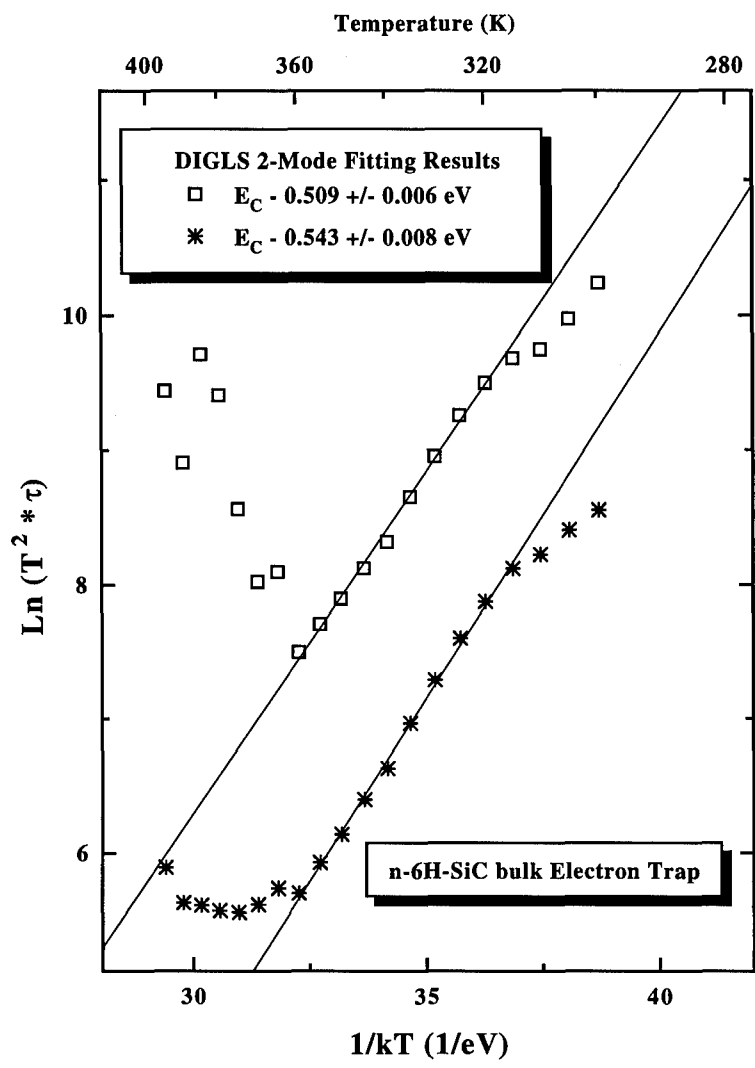
Figure 72. Arrhenius plot for n-6H-SiC.
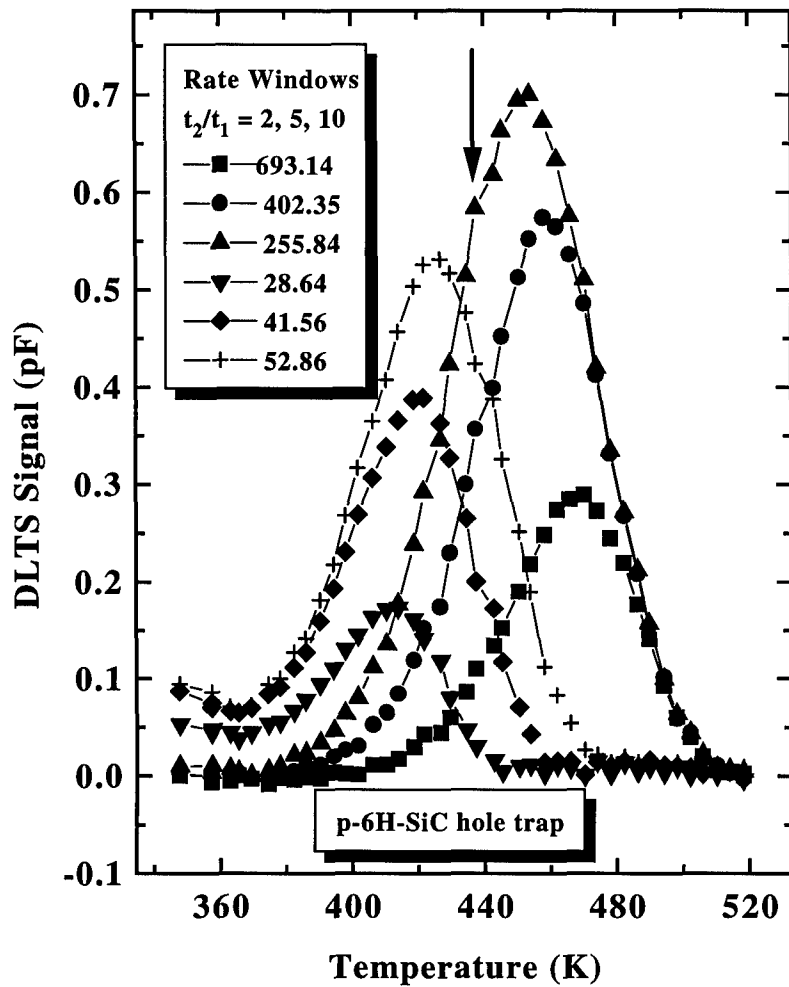
Figure 73. Rate window plot for p-6H-SiC. The large arrow points at a shoulder that typically represents multi-mode decay.
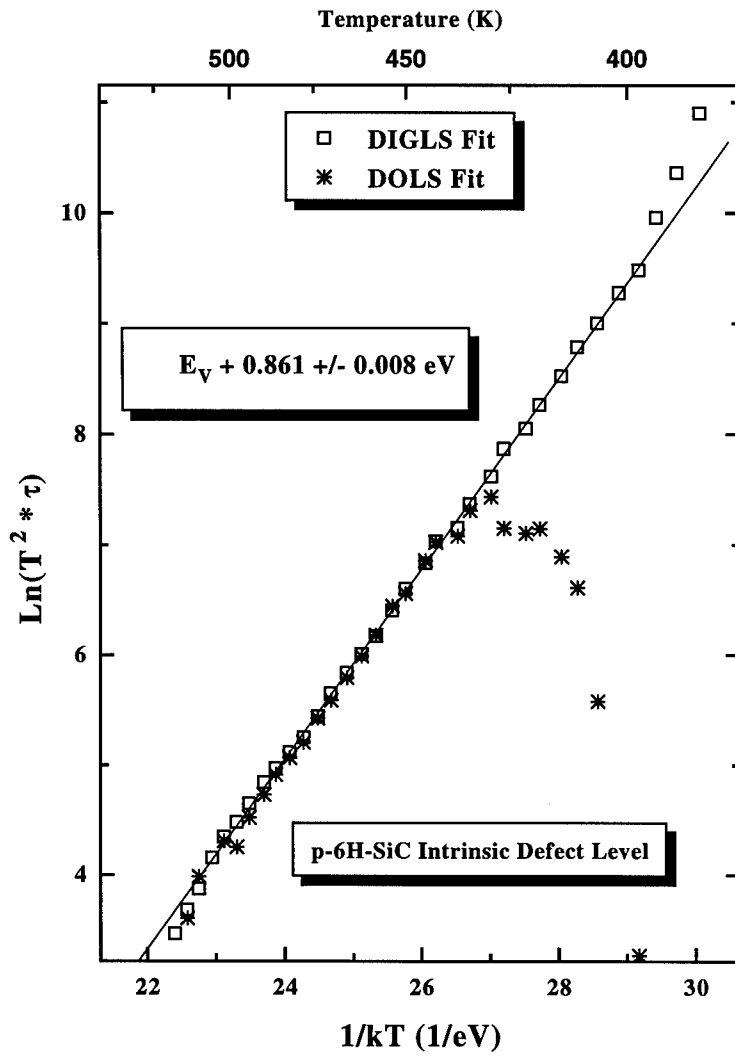
Figure 74. Arrhenius plot for p-6H-SiC.

6.5 Additional Estimation Algorithms for DLTS

In the course of our research, we realized the exact maximum likelihood cost function for the one mode DLTS signal can be minimized directly without differencing and without linear prediction coefficient representation. Let

$$(261) \qquad V_{c_0} = \begin{bmatrix} \bar{1} & \bar{v}_1 \end{bmatrix}$$

where $\bar{1}$ is a $M \times 1$ vector of ones and

$$(262) \qquad \bar{v}_1 = \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{M-1} \end{bmatrix}^T,$$

so that

$$(263) \qquad \bar{y} = V_{c_0} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} + \bar{w}.$$

From Chapters II and III we know the ML solution is attained by maximizing

$$(264) \qquad L(V_{c_0}) = \bar{y}^T V_{c_0} (V_{c_0}^T V_{c_0})^{-1} V_{c_0}^T \bar{y}.$$

We find the behavior of the one mode DLTS ML cost function is similar to the behavior of the one mode real exponential ML cost function analyzed in Chapter V. Over a large range of SNRs, the one mode DLTS ML cost function is unimodal. Consequently, with only one variable, λ_1, a golden section search can be implemented. Details of the golden section search are given in Chapter V.

Figures 75 and 76 are inverse MSE and bias plots for comparing the performance of the DIGLS algorithm with the golden section search (Gold DLTS) algo-
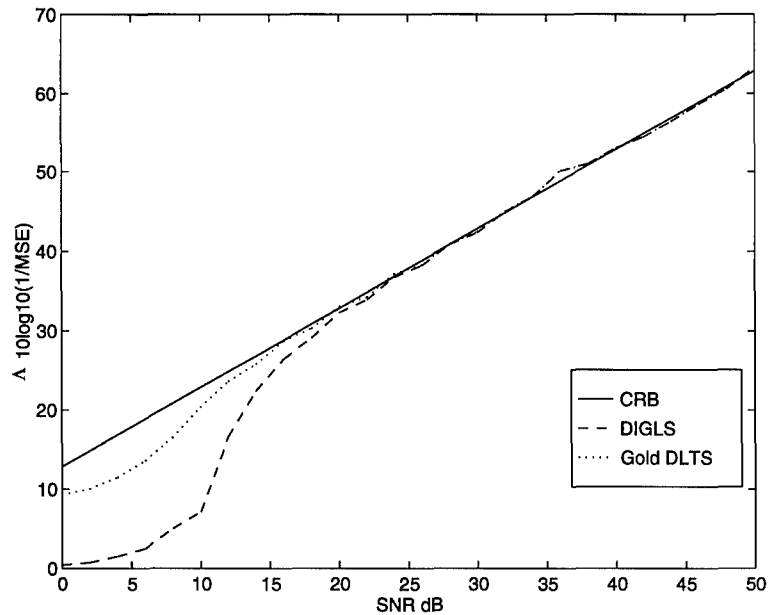
Figure 75. The inverse MSE of the DIGLS and Gold DLTS estimators over a range
of SNRs. Also plotted is the CRB.

rithm for the one mode DLTS problem. The Monte-Carlo experiment used to create
the plots is identical to the experiments accomplished earlier in the chapter except
the underlying signal consists of only one real mode with parameters $c_1 = -1$ and
$\lambda_1 = .8$. As expected, the Gold DLTS algorithm is superior at lower SNRs. For
the one mode problem, the golden section search is more effective at maximizing
the exact ML cost function. However, the Gold DLTS algorithm is restricted to one
mode DLTS problems.

In addition to the one mode DLTS algorithm just delivered, we also developed a
multi-mode DLTS algorithm that performs differencing implicitly. In this algorithm,
we treat the baseline constant as an additional exponential, λ_0, with an amplitude
coefficient, c_0, but we constrain λ_0 to equal one in an elegant and insightful way.
The algorithm essentially follows the development of IGLS.
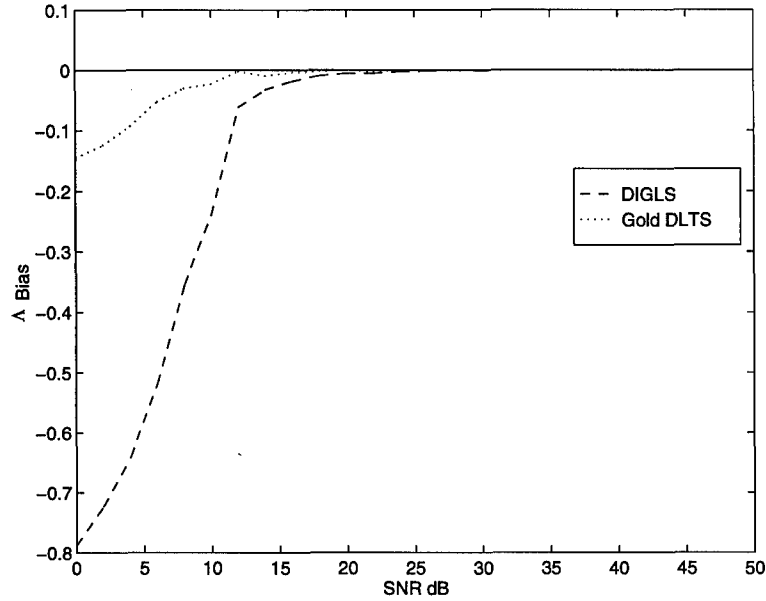
Figure 76.   The bias of the DIGLS and Gold DLTS estimators over a range of SNRs.

With a model order of $N + 1$, consider the noiseless LP equation

$$(265) \quad b_0 s[m] + b_1 s[m-1] + \ldots + b_N s[m-N] + b_{N+1} s[m-(N+1)] = 0.$$

Recall, the estimates of λ_n are the roots of the polynomial formed from the LP coefficients. Therefore, we assume.

$$(266) \qquad b_0 z^{N+1} + b_1 z^N + b_2 z^{N-1} + \ldots + b_N + b_{N+1} = 0.$$

Since c_0 is a constant, we know $\lambda_0 = 1$ is a root of the LP polynomial, Equation 266. Therefore,

$$(267) \qquad \begin{aligned} b_0 + b_1 + b_2 + \ldots + b_N + b_{N+1} &= 0 \\ \Rightarrow \quad b_{N+1} &= -(b_0 + b_1 + b_2 + \ldots + b_N). \end{aligned}$$

163

This is a constraint we can apply in the standard IGLS development. Since

$$(268) \qquad s[m] = y[m] - w[m],$$

we know

$$b_0 y[m] + b_1 y[m-1] + \ldots + b_N y[m-N] + b_{N+1} y[m-(N+1)]$$

$$(269) \quad = \quad b_0 w[m] + b_1 w[m-1] + \ldots + b_N w[m-N] + b_{N+1} w[m-(N+1)].$$

When we introduce the constraint of Equation 267, we can say

$$b_0(y[m] - y[m-(N+1)]) + b_1(y[m-1] - y[m-(N+1)]) + \ldots +$$

$$b_N y([m-N] - y[m-(N+1)])$$

$$= \quad b_0 w[m] + b_1 w[m-1] + \ldots + b_N w[m-N] -$$

$$(b_0 + b_1 + b_2 + \ldots + b_N) w[m-(N+1)]$$

$$(270) \qquad = \quad e_c[m].$$

The LP error $e_c[m]$ is still composed of a tractable sum of weighted Gaussian random variables. Under an overdetermined set of linear equations, let

$$(271) \qquad Y_c \bar{b} = \bar{e}_c$$

where

$$Y_c = \begin{bmatrix} \bar{y}_{c0} & \bar{y}_{c1} & \cdots & \bar{y}_{cN} \end{bmatrix}$$

$$\bar{y}_{cn} = \begin{bmatrix} (y[N+1-n] - y[0]) & (y[N+2-n] - y[1]) & \cdots \\ y([M-1-n] - y[M-N-2]) \end{bmatrix}^T$$

$$\bar{b} = \begin{bmatrix} b_0 & b_1 & \cdots & b_N \end{bmatrix}^T$$

(272)
$$\bar{e} = \begin{bmatrix} e_c[N+1] & e_c[N+2] & \cdots & e_c[M-1] \end{bmatrix}^T.$$

The data matrix, Y_c, has dimensions $(M - N - 1) \times (N + 1)$ and Toeplitz structure. The LP error vector $\bar{e}_c$ is distributed with mean vector zero and covariance matrix $\sigma^2 R_c$. If we define the $(M - N - 1) \times M$ Toeplitz matrix

(273)

$$B_c = \begin{bmatrix} b_0 & b_1 & \cdots & b_N & -(b_0+b_1+...+b_N) & 0 & \cdots & 0 \\ 0 & b_0 & b_1 & \cdots & b_N & -(b_0+b_1+...+b_N) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & b_0 & b_1 & \cdots & b_N & -(b_0+b_1+...+b_N) \end{bmatrix},$$

then the differenced LP error covariance matrix is obtained with

(274)
$$R_d = B_c B_c^T$$

and the GLS estimator becomes

(275)
$$\bar{b} = -(Y_c^T R_c^{-1} Y_c)^{-1} Y_c^T R_c^{-1} \bar{y}_{d0}.$$
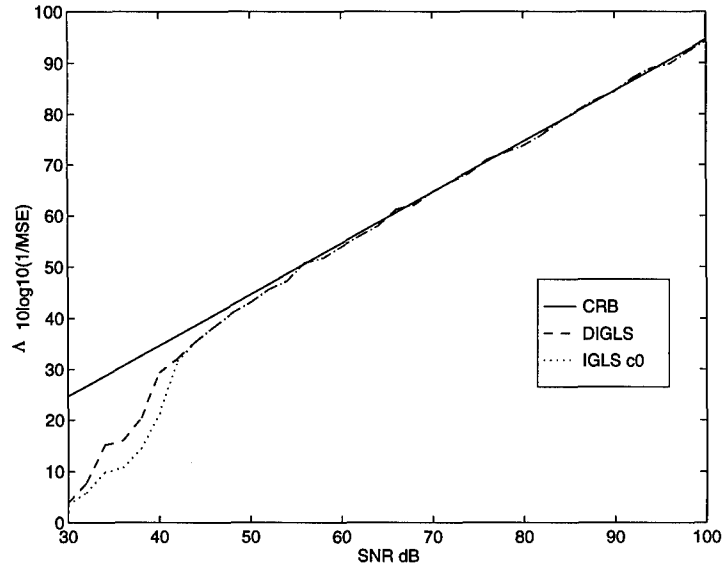
Figure 77. The λ_1 inverse MSE of the DIGLS and IGLS $c0$ estimators over a range of SNRs. Also plotted is the CRB.

As before, an iterative algorithm naturally follows, and after rooting, the amplitude coefficients, c_0 and c_n, can be estimated from the expanded pseudo-inverse estimator

$$(276) \qquad \begin{bmatrix} c_0 \\ \bar{c} \end{bmatrix} = (V_{c_o}^T V_{c_o})^{-1} V_{c_o}^T \bar{y}.$$

Figures 77, 78, 79, and 80 are inverse MSE and bias plots for comparing the performance of the IGLS algorithm with the c_0 constraint (IGLS c_0) and the DIGLS algorithm derived earlier for the two mode DLTS problem.   The performance of the two algorithms is essentially identical until below the SNR threshold. This is expected because both algorithms—although differing in appearance—attempt to minimize the ML cost function. This bolsters our claim that introducing our

166
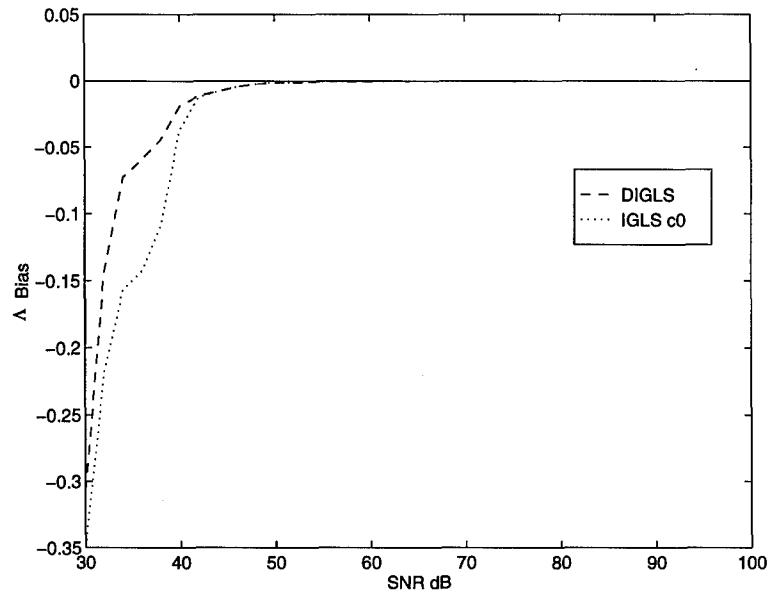
Figure 78. The λ_1 bias of the DIGLS and IGLS c_0 estimators over a range of SNRs.
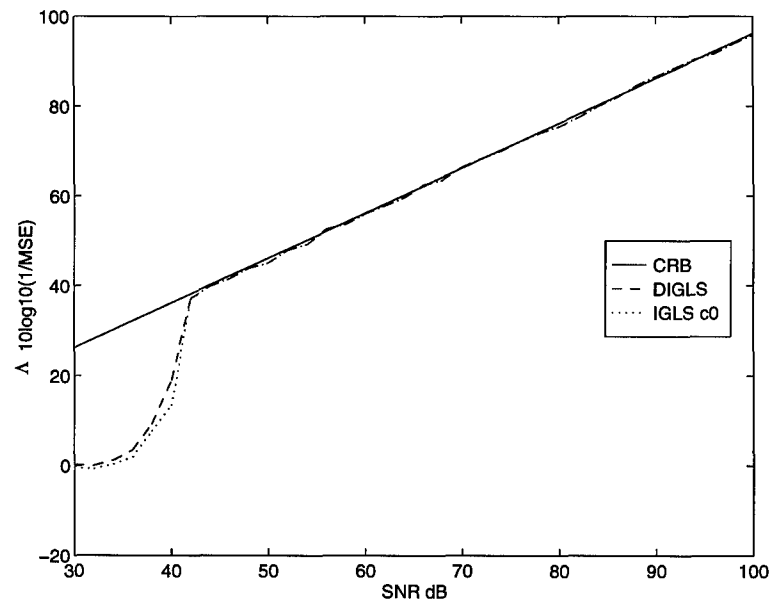
Figure 79. The λ_2 inverse MSE of the DIGLS and IGLS c_0 estimators over a range of SNRs. Also plotted is the CRB.
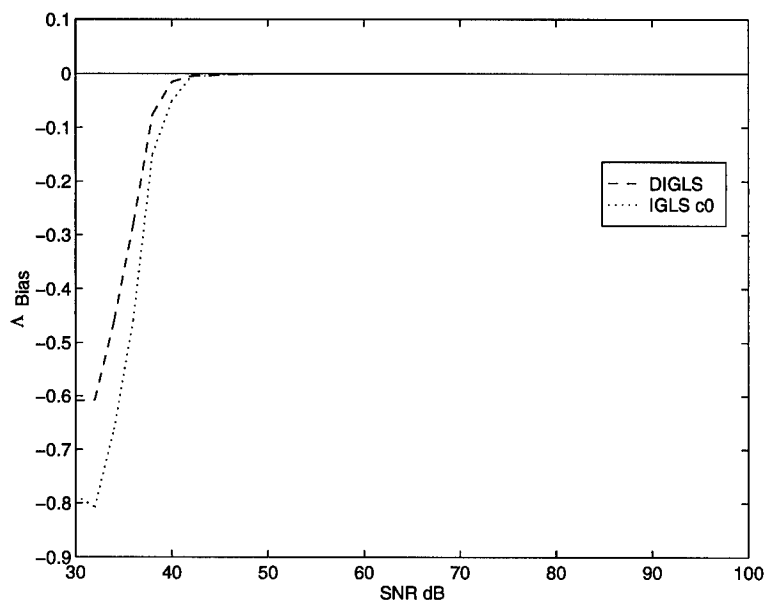
Figure 80. The λ_2 phase bias of the DIGLS and IGLS c_0 estimators over a range of SNRs.

LP methodology brings valuable insight to the superimposed exponential parameter estimation problem.

6.6 Chapter Conclusion

We began the chapter by developing the fundamentals of DLTS, and its relationship with the superimposed exponential model. We discussed immediate estimation applications of the work developed in the previous chapters for DLTS signals with the baseline constant estimated and removed by hardware. Then, we expanded our IGLS algorithm to estimate the baseline constant without the need of DLTS hardware. Our differenced IGLS algorithm was facilitated by our understanding of linear prediction and would be difficult to develop from maximum likelihood. The performance of the DIGLS algorithm was verified with simulations and actual DLTS data. Finally, we concluded the chapter with additional algorithms developed for

168

DLTS and generated by our increased understanding of the superimposed exponential problem.

VII. Conclusion

7.1 Primary Contributions

We began this dissertation by developing the exact maximum likelihood (ML) cost function for estimating the parameters of superimposed exponentials in zero mean, white, Gaussian noise. In the same chapter, Chapter II, we showed how the ML methodology is related to the Cramer-Rao bound (CRB) and why the CRB is useful for evaluating the performance of estimators. In the next chapter, Chapter III, we gave a historical review of parameter estimators developed with the linear prediction (LP) methodology. At the end of the chapter, we introduced critical—previously omitted—statistical modeling considerations for linear prediction. With our statistical considerations, we showed how the exact ML cost function, for the superimposed exponential parameter estimation problem, can be developed with the linear prediction methodology.

Attaining maximum likelihood performance with linear prediction is the most important contribution of our work. Linear prediction can often provide critical insight into the analysis of a problem. Previous to our work, the LP approach was hindered by the erroneous assumption that estimators developed from linear prediction are always inferior to estimators developed from maximum likelihood. With our linear prediction methodology, researchers can utilize the insights and flexibility of linear prediction and achieve maximum likelihood estimation performance.

We demonstrated this capability in our deep level transient spectroscopy (DLTS) application chapter, Chapter VI. In DLTS, the signal is complicated by the presence of a baseline constant. With our linear prediction methodology, the exact ML cost function for the DLTS signal was intuitively developed for minimization. The result-

ing algorithm provided DLTS parameter estimates significantly more accurate than those of the estimators currently in use. Our algorithm was readily accepted by the DLTS community because of their previous exposure to linear prediction and their need for improved capability. In fact, our DLTS contribution has been accepted for publication in the Journal of Applied Physics (30).

7.2 Ancillary Contributions

In our research, we developed additional contributions. In the linear prediction chapter and the cost function minimization chapter—Chapters III and IV, respectively—we proposed and defended an alternative explanation for the moderate success of total least squares (TLS) as a linear prediction based approach for superimposed exponential parameter estimation. We suspect that the readily available minimum norm solution, associated with overmodeling, in TLS is more influential than the "increased perturbation" explanation often provided. In Chapter IV, we showed how TLS, without overmodeling, is actually ill-conditioned at low SNR.

In that same chapter, we showed how TLS, without overmodeling, is equivalent to the eigenvalue-eigenvector optimization technique used extensively for ML cost function minimization. We compared both TLS, without overmodeling, and eigenvalue-eigenvector minimization to the generalized least squares (GLS) approach and concluded GLS is more appropriate for minimization unless additional constraints must be accommodated.

In Chapter V we carefully analyzed the one and two mode *real* exponential problem. We developed unique algorithms as alternatives to iterative generalized least squares (IGLS) and exercised IGLS at alternative initial conditions. We found the IGLS algorithm to be consistently effective and relatively invariant to starting

position. We endorsed the IGLS algorithm as the most versatile tool for superimposed real exponential parameter estimation. However, when the problem was restricted to the estimation of a single real exponential, we found our application of the golden section search to be superior. This fact was facilitated by our observation that the exact ML cost function, for the single real exponential problem, is unimodal in λ—accept under very low SNR scenarios.

7.3 Immediate Additional Work

Programming the differenced iterative generalized least squares algorithm is straight forward. Nevertheless, the algorithm needs to be packaged into a user friendly program for wide spread DLTS applications.

Also, the single mode analysis for estimating the λ of a real exponential in Chapter V should extend to other single mode single parameter exponential problems such as the frequency of a pure sinusoid. The potential of directly minimizing the exact ML cost function for a pure sinusoid with such techniques as a golden section search needs to be investigated. Accurate, efficient single sinusoid frequency estimation is a high interest subject (34, 66).

Finally, the two mode analysis for estimating the λ_n of two real exponentials in Chapter V should extend to other two mode single parameter exponential problems such as two superimposed sinusoids at a known frequency with different unknown phases. This problem arises in communication theory and might by effectively resolved by constraining IGLS for phase only estimation. As always, further investigation is desired.

Appendix A. Properties of Special Matrices

A.1 Vandermonde Matrix V

The $M \times N$ matrix V, defined

$$(277) \qquad V = \left[\begin{array}{cccc} \bar{v}(\lambda_1) & \bar{v}(\lambda_2) & \cdots & \bar{v}(\lambda_N) \end{array} \right]$$

where

$$(278) \qquad \bar{v}(\lambda_n) = \left[\begin{array}{ccccc} 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{M-1} \end{array} \right]^T$$

and $M > N$, has Vandermonde structure. When each λ_n is distinct, each column of V is linearly independent. One way to prove this is to consider an $M \times M$ Vandermonde matrix V_m with M distinct λ_m. The determinant of V_m is

$$(279) \qquad |V_m| = \prod_{\substack{m_1,m_2=1 \\ m_1 > m_2}}^{M} \left(\lambda_{m_1} - \lambda_{m_2} \right).$$

For the determinant of V_m to equal zero, λ_{m_1} must equal λ_{m_2}. Since we know each λ_m is distinct the determinant of V_m is not equal to zero. Therefore V_m is nonsingular and each column is linearly independent (25:29). Because any subset of linear independent vectors is also linear independent the matrix V formed from any N columns of V_m has full column rank.

A.2 Toeplitz Data Matrix Y and $\tilde{Y}$

The $(M - N) \times (N + 1)$ data matrix Y, defined

$$(280) \qquad Y = \left[\begin{array}{cccc} \bar{y}_0 & \bar{y}_1 & \cdots & \bar{y}_N \end{array}\right]$$

where

$$(281) \qquad \bar{y}_n = \left[\begin{array}{cccc} y[N - n] & y[N + 1 - n] & \cdots & y[M - 1 - n] \end{array}\right]^T$$

and $M > N$, has Toeplitz structure. Because Y is comprised of stochastic elements, its properties have statistical connotations. First consider the underlying deterministic elements, $s[m]$, under the same construct of Y so that

$$(282) \qquad S = \left[\begin{array}{cccc} \bar{s}_0 & \bar{s}_1 & \cdots & \bar{s}_N \end{array}\right]$$

where

$$(283) \qquad \bar{s}_n = \left[\begin{array}{cccc} s[N - n] & s[N + 1 - n] & \cdots & s[M - 1 - n] \end{array}\right]^T .$$

It was noted by Tufts and Kumaresan (72:977) that the rank of S is N. This is justified by the relationship

$$(284) \qquad \bar{s} = V\bar{c}.$$

Every column of S is a linear combination of the first $M - N$ elements of the N linearly independent columns of V. Therefore, the rank of S is the rank of V which

174

is N. From the linear prediction equation, we also know,

$$(285) \qquad\qquad S\bar{b} = \bar{0}$$

where $\bar{b}$ is a vector of the linear prediction coefficients $b_0 \ldots b_N$. Therefore, since b_0 can be normalized to 1 without altering the relationship,

$$(286) \qquad\qquad \bar{s}_0 + \tilde{S}\tilde{b} = \bar{0}$$

where

$$(287) \qquad\qquad \tilde{S} = \left[\begin{array}{cccc} \bar{s}_1 & \bar{s}_2 & \cdots & \bar{s}_N \end{array}\right]$$

and

$$(288) \qquad\qquad \tilde{b} = \left[\begin{array}{cccc} b_1 & b_2 & \cdots & b_N \end{array}\right]^T.$$

Because $\tilde{b}$ is a solution to $\bar{s}_0 + \tilde{S}\tilde{b} = \bar{0}$, the rank of $\tilde{S}$ must equal the rank of S which is N (21:140). With dimension $(M - N) \times N$, $\tilde{S}$ has full column rank. Therefore, each column of $\tilde{S}$ is linearly independent.

In our research,

$$(289) \qquad\qquad \bar{y}_n = \bar{s}_n + \bar{w}$$

where the stochastic $\bar{w}$ has mean vector zero. When we consider the $(M - N) \times N$ data matrix

$$(290) \qquad\qquad \tilde{Y} = \left[\begin{array}{cccc} \bar{y}_1 & \bar{y}_2 & \cdots & \bar{y}_N \end{array} \right]$$

we know the underlying signal matrix has rank N. We do not expect the addition of stochastic elements to the underlying signal matrix in $\tilde{Y}$ to cause linear dependence in the columns of $\tilde{Y}$. Therefore, $\tilde{Y}$ is assumed to have full column rank, N. In stochastic theory, every column of Y is considered to be linearly independent as well (56:275). Therefore, Y is assumed to have full column rank, $N + 1$, even though its underlying signal matrix rank is N. This theory extends to all noisy data matrices with any number of columns. We can assume all noisy data matrices have full column rank.

Appendix B. Complex Gradients

B.1 Derivation

Let $\bar{b}$ and $\bar{\theta}$ be complex $N \times 1$ vectors and let A be a Hermitian $N \times N$ matrix. Recall the gradient definitions

$$(291) \qquad \frac{\partial L}{\partial \bar{\theta}} = \frac{1}{2} \left(\frac{\partial L}{\partial \bar{\theta}_r} - j \frac{\partial L}{\partial \bar{\theta}_i} \right)$$

and

$$(292) \qquad \frac{\partial L}{\partial \bar{\theta}^*} = \frac{1}{2} \left(\frac{\partial L}{\partial \bar{\theta}_r} + j \frac{\partial L}{\partial \bar{\theta}_i} \right)$$

where L is a scalar function and $\bar{\theta} = \bar{\theta}_r + j\bar{\theta}_i$ ($\bar{\theta}_r$ and $\bar{\theta}_i$ are real vectors).

Notice that under the gradient definitions,

$$
\begin{aligned}
\frac{\partial L}{\partial \bar{\theta}^*} &= \frac{\partial L}{\partial \left(\bar{\theta}_r + j(-\bar{\theta}_i) \right)} \\
&= \frac{1}{2} \left(\frac{\partial L}{\partial \bar{\theta}_r} - j \frac{\partial L}{\partial (-\bar{\theta}_i)} \right) \\
&= \frac{1}{2} \left(\frac{\partial L}{\partial \bar{\theta}_r} + j \frac{\partial L}{\partial \bar{\theta}_i} \right) \\
(293) \qquad &= \left(\frac{\partial L}{\partial \bar{\theta}} \right)^* .
\end{aligned}
$$

Therefore, $\frac{\partial L}{\partial \theta^*}$ is the complex conjugate of $\frac{\partial L}{\partial \theta}$.

Next, consider the gradient $\frac{\partial L}{\partial \theta}$ when the vector complex variable, $\bar{\theta}$, is reduced to a scalar complex variable, θ, so that

$$(294) \qquad \frac{\partial L}{\partial \theta} = \frac{1}{2} \left(\frac{\partial L}{\partial \theta_r} - j \frac{\partial L}{\partial \theta_i} \right) .$$

177

With this definition,

$$\frac{\partial \theta}{\partial \theta} = \frac{1}{2}\left(\frac{\partial(\theta_r + j\theta_i)}{\partial\theta_r} - j\frac{\partial(\theta_r + j\theta_i)}{\partial\theta_i}\right)$$

$$= \frac{1}{2}\left(\frac{\partial\theta_r}{\partial\theta_r} + j\frac{\partial\theta_i}{\partial\theta_r} - j\frac{\partial\theta_r}{\partial\theta_i} + \frac{\partial\theta_i}{\partial\theta_i}\right)$$

$$= \frac{1}{2}(1 + 0 - 0 + 1)$$

$$\text{(295)} \qquad\qquad = 1$$

and

$$\frac{\partial\theta^*}{\partial\theta} = \frac{1}{2}\left(\frac{\partial(\theta_r - j\theta_i)}{\partial\theta_r} - j\frac{\partial(\theta_r - j\theta_i)}{\partial\theta_i}\right)$$

$$= \frac{1}{2}\left(\frac{\partial\theta_r}{\partial\theta_r} - j\frac{\partial\theta_i}{\partial\theta_r} - j\frac{\partial\theta_r}{\partial\theta_i} - \frac{\partial\theta_i}{\partial\theta_i}\right)$$

$$= \frac{1}{2}(1 - 0 - 0 - 1)$$

$$\text{(296)} \qquad\qquad = 0.$$

Now, we can consider the gradient of $\bar{b}^H\bar{\theta}$ with respect to a vector complex variable $\bar{\theta}$. Let

$$L = \bar{b}^H\bar{\theta}$$

$$\text{(297)} \qquad\qquad = \sum_{x=1}^{N} b_x^*\theta_x.$$

Utilizing the definition of $\frac{\partial L}{\partial \bar{\theta}}$, defined in Equation 294, for each scalar complex variable θ_n in $\bar{\theta}$, we know

$$
\begin{aligned}
\frac{\partial \left(\sum_{x=1}^{N} b_x^* \theta_x \right)}{\partial \theta_n} &= \sum_{x=1}^{N} b_x^* \frac{\partial \theta_x}{\partial \theta_n} \\
&= b_n^* \frac{\partial \theta_n}{\partial \theta_n} \\
&= b_n^*.
\end{aligned}
$$

(298)

Therefore,

(299)
$$
\frac{\partial \left(\bar{b}^H \bar{\theta} \right)}{\partial \bar{\theta}} = \bar{b}^*.
$$

Similarly, let

$$
\begin{aligned}
L &= \bar{\theta}^H \bar{b} \\
&= \sum_{x=1}^{N} \theta_x^* b_x.
\end{aligned}
$$

(300)

From the previous methodology, we know

$$
\begin{aligned}
\frac{\partial \left(\sum_{x=1}^{N} \theta_x^* b_x \right)}{\partial \theta_n} &= \sum_{x=1}^{N} b_x \frac{\partial \theta_x^*}{\partial \theta_n} \\
&= b_n \frac{\partial \theta_n^*}{\partial \theta_n} \\
&= 0.
\end{aligned}
$$

(301)

Therefore,

(302)
$$
\frac{\partial \left(\bar{\theta}^H \bar{b} \right)}{\partial \bar{\theta}} = \bar{0}.
$$

Finally, let

$$
\begin{aligned}
L & = \bar{\theta}^H A \bar{\theta} \\
& = \sum_{y=1}^{N} \sum_{x=1}^{N} \theta_x^* a_{xy} \theta_y.
\end{aligned}
$$

(303)

Again with the established methodology, we know

$$
\begin{aligned}
\frac{\partial \left(\sum_{y=1}^{N} \sum_{x=1}^{N} \theta_x^* a_{xy} \theta_y \right)}{\partial \theta_n} & = \sum_{y=1}^{N} \sum_{x=1}^{N} \left(a_{xy} \theta_y \frac{\partial \theta_x^*}{\partial \theta_n} + a_{xy} \theta_x^* \frac{\partial \theta_y}{\partial \theta_n} \right) \\
& = \sum_{y=1}^{N} \sum_{x=1}^{N} \left(a_{xy} \theta_x^* \frac{\partial \theta_y}{\partial \theta_n} \right) \\
& = \sum_{x=1}^{N} a_{xn} \theta_x^* \frac{\partial \theta_n}{\partial \theta_n} \\
& = \sum_{x=1}^{N} a_{xn} \theta_x^*.
\end{aligned}
$$

(304)

Therefore,

$$
\frac{\partial \left(\bar{\theta}^H A \bar{\theta} \right)}{\partial \bar{\theta}} = A^T \bar{\theta}^* = \left(A \bar{\theta} \right)^*.
$$

(305)

Consequently, from the complex conjugate relationship

$$
\frac{\partial L}{\partial \bar{\theta}^*} = \left(\frac{\partial L}{\partial \bar{\theta}} \right)^*,
$$

(306)

we know

$$\frac{\partial\left(\bar{b}^H\bar{\theta}\right)}{\partial\bar{\theta}*} = \bar{b}$$

$$\frac{\partial\left(\bar{\theta}^H\bar{b}\right)}{\partial\bar{\theta}*} = \bar{0}$$

$$(307)\qquad\frac{\partial\left(\bar{\theta}^H A\bar{\theta}\right)}{\partial\bar{\theta}*} = A\bar{\theta}.$$

Appendix C. Theory of the Cramer-Rao Bound

C.1 Property One

Recall from the first property of the theory of the CRB that if a PDF, $f(\bar{y}; \bar{\theta})$, satisfies the "regularity" condition, then the error covariance matrix, C, of any unbiased estimator, $g(\bar{y})$, must satisfy the condition that $C - F(\bar{\theta})^{-1}$ is positive semidefinite. The matrix, $F(\bar{\theta})$, is the Fisher information matrix and is defined as the covariance matrix of the gradient of the log-likelihood function

$$(308) \qquad F(\bar{\theta}) = E\left\{ \left(\frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \right) \left(\frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \right)^{H} \right\}.$$

For the difference of matrices, $C - F(\bar{\theta})^{-1}$, to be positive semi-definite, each element of the diagonal of C must be greater than or equal to each element of the diagonal of $F(\bar{\theta})^{-1}$. This implies

$$(309) \qquad C_{nn} = E\left\{ (g(\bar{y})_n - \bar{\theta}_n)(g(\bar{y})_n - \bar{\theta}_n)^{H} \right\} \geq F(\bar{\theta})_{nn}^{-1}.$$

We now prove that, with the "regularity" condition satisfied and the existence of an unbiased estimator assumed, the difference of matrices, $C - F(\bar{\theta})^{-1}$, must be positive semi-definite. The proof follows the delivery of Scharf (61:221–229).

By definition, the area under a PDF must equal 1. Therefore,

$$(310) \qquad \int_{-\infty}^{\infty} f(\bar{y}; \bar{\theta}) d\bar{y} = 1$$

182

and

$$(311) \qquad \frac{\partial}{\partial \bar{\bar{\theta}}} \int_{-\infty}^{\infty} f(\bar{y}; \bar{\theta}) d\bar{y} = \frac{\partial 1}{\partial \bar{\bar{\theta}}} = \bar{0}.$$

When we assume the "regularity" condition, we assume

$$(312) \qquad \frac{\partial}{\partial \bar{\bar{\theta}}} \int_{-\infty}^{\infty} f(\bar{y}; \bar{\theta}) d\bar{y} = \int_{-\infty}^{\infty} \frac{\partial f(\bar{y}; \bar{\theta})}{\partial \bar{\bar{\theta}}} d\bar{y}.$$

In our case, this is a valid assumption for the modeled PDF. Therefore,

$$(313) \qquad \int_{-\infty}^{\infty} \frac{\partial f(\bar{y}; \bar{\theta})}{\partial \bar{\bar{\theta}}} d\bar{y} = \bar{0}.$$

Through the derivative, we know

$$(314) \qquad \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\bar{\theta}}} = \frac{1}{f(\bar{y}; \bar{\theta})} \frac{\partial f(\bar{y}; \bar{\theta})}{\partial \bar{\bar{\theta}}}.$$

Therefore,

$$(315) \qquad \int_{-\infty}^{\infty} \frac{\partial f(\bar{y}; \bar{\theta})}{\partial \bar{\bar{\theta}}} d\bar{y} = \int_{-\infty}^{\infty} \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\bar{\theta}}} f(\bar{y}; \bar{\theta}) d\bar{y} = \bar{0}.$$

The middle expression in Equation 315 is the definition of the expected value for the gradient of the log-likelihood function. This implies that the mean vector of the gradient of the log-likelihood function equals $\bar{0}$ or

$$(316) \qquad E\left\{ \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\bar{\theta}}} \right\} = \bar{0}.$$

With the mean vector equal to $\bar{0}$, the covariance matrix of the gradient of the log-likelihood function is

$$(317) \qquad F(\bar{\theta}) = E\left\{ \left(\frac{\partial \ln f(\bar{y};\bar{\theta})}{\partial \bar{\theta}} \right) \left(\frac{\partial \ln f(\bar{y};\bar{\theta})}{\partial \bar{\theta}} \right)^H \right\}.$$

The matrix, $F(\bar{\theta})$, is the Fisher information matrix. Because $F(\bar{\theta})$ is constructed from a covariance matrix and because each of the $y[m]$ elements in $\bar{y}$ are stochastically independent, the Fisher information matrix is positive definite (54:190). Thus $F(\bar{\theta})^{-1}$ exists. Also, by construction, $F(\bar{\theta})$ is Hermitian.

Continuing the proof, when we assume the existence of an unbiased estimator, we assume

$$(318) \qquad E\left\{g(\bar{y})\right\} = \bar{\theta}.$$

Since $E\{\bar{\theta}\} = \int_{-\infty}^{\infty} \bar{\theta} f(\bar{y};\bar{\theta}) d\bar{y} = \bar{\theta}$, we can say

$$(319) \qquad \begin{aligned} E\left\{g(\bar{y})\right\} - E\left\{\bar{\theta}\right\} &= \bar{0} \\ \Rightarrow \quad E\left\{g(\bar{y}) - \bar{\theta}\right\} &= \bar{0} \\ \Rightarrow \quad \int_{-\infty}^{\infty} \left(g(\bar{y}) - \bar{\theta}\right) f(\bar{y};\bar{\theta}) d\bar{y} &= \bar{0} \\ \Rightarrow \quad \int_{-\infty}^{\infty} f(\bar{y};\bar{\theta}) \left(g(\bar{y}) - \bar{\theta}\right)^H d\bar{y} &= \bar{0}^H. \end{aligned}$$

With the "regularity" condition and the chain rule, we know

$$
\frac{\partial}{\partial \bar{\theta}} \int_{-\infty}^{\infty} f(\bar{y}; \bar{\theta}) \left(g(\bar{y}) - \bar{\theta} \right)^H d\bar{y} =
$$

$$
= \int_{-\infty}^{\infty} \frac{\partial \left(f(\bar{y}; \bar{\theta}) \left(g(\bar{y}) - \bar{\theta} \right)^H \right)}{\partial \bar{\theta}} d\bar{y}
$$

$$
= \int_{-\infty}^{\infty} \left[\frac{\partial f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \left(g(\bar{y}) - \bar{\theta} \right)^H + f(\bar{y}; \bar{\theta}) \frac{\partial \left(g(\bar{y}) - \bar{\theta} \right)^H}{\partial \bar{\theta}} \right] d\bar{y}
$$

$$
= \int_{-\infty}^{\infty} \frac{\partial f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \left(g(\bar{y}) - \bar{\theta} \right)^H d\bar{y} - \int_{-\infty}^{\infty} f(\bar{y}; \bar{\theta}) \mathrm{I} d\bar{y}
$$

$$
= \int_{-\infty}^{\infty} \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \left(g(\bar{y}) - \bar{\theta} \right)^H f(\bar{y}; \bar{\theta}) d\bar{y} - \mathrm{I}
$$

$$
= E \left\{ \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \left(g(\bar{y}) - \bar{\theta} \right)^H \right\} - \mathrm{I}
$$

$$
(320) \qquad = \frac{\partial \bar{0}^H}{\partial \bar{\theta}}.
$$

Therefore,

$$
(321) \qquad E \left\{ \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \left(g(\bar{y}) - \bar{\theta} \right)^H \right\} = \mathrm{I}.
$$

In words, the cross-covariance of the gradient of the log-likelihood function and the estimator error vector is equal to the identity matrix.

Let us propose a vector comprised of the estimator error vector stacked over the gradient of the log-likelihood function represented by

$$
(322) \qquad \begin{bmatrix} g(\bar{y}) - \bar{\theta} \\ \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \end{bmatrix}_{2N \times 1}
$$

where N is the number of parameters to be estimated. With the assumptions and derivations just presented, we know the mean of such a vector is $\bar{0}$ and the covariance

185

matrix is

$$Q_1 = E\left\{ \begin{bmatrix} g(\bar{y}) - \bar{\theta} \\ \frac{\partial \ln f(\bar{y};\bar{\theta})}{\partial \theta} \end{bmatrix} \begin{bmatrix} \left(g(\bar{y}) - \bar{\theta} \right)^H & \frac{\partial \ln f(\bar{y};\bar{\theta})}{\partial \theta}^H \end{bmatrix} \right\}$$

$$(323) \qquad = \begin{bmatrix} C & I \\ I & F(\bar{\theta}) \end{bmatrix}.$$

Because Q_1 is a covariance matrix, it is at least positive semi-definite (54:190). Therefore, for any vector $\bar{u}$,

$$(324) \qquad\qquad\qquad \bar{u}^H Q_1 \bar{u} \geq 0.$$

If we define the vector $\bar{u} = W\bar{v}$, where W is any matrix of acceptable dimensions, then for any vector $\bar{v}$,

$$(325) \qquad\qquad\qquad \bar{v}^H W^H Q_1 W \bar{v} \geq 0.$$

Therefore, $W^H Q_1 W$ is also positive semi-definite (25:399).

Let

$$(326) \qquad\qquad\qquad W = \begin{bmatrix} I & 0 \\ -F(\bar{\theta})^{-1} & I \end{bmatrix}$$

186

so that

$$Q_2 = W^H Q_1 W$$

$$= \begin{bmatrix} I & -F(\bar{\theta})^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} C & I \\ I & F(\bar{\theta}) \end{bmatrix} \begin{bmatrix} I & 0 \\ -F(\bar{\theta})^{-1} & I \end{bmatrix}$$

$$(327) \qquad = \begin{bmatrix} C - F(\bar{\theta})^{-1} & 0 \\ 0 & F(\bar{\theta}) \end{bmatrix}$$

and Q_2 is positive semi-definite.

To characterize the properties of the $C - F(\bar{\theta})^{-1}$ submatrix of Q_2, let $\bar{v}$ be a $2N \times 1$ vector where the first N elements are arbitrary and the last N elements are zero. Also, let $\tilde{v}$ be an $N \times 1$ vector of just the first N arbitrary elements of $\bar{v}$, and let $\tilde{Q}_2$ equal the difference of matrices $C - F(\bar{\theta})^{-1}$. With these definitions, we know

$$(328) \qquad \tilde{v}^H \tilde{Q}_2 \tilde{v} = \bar{v}^H Q_2 \bar{v} \geq 0.$$

Since the elements of $\tilde{v}$ are arbitrary, $\tilde{Q}_2 = C - F(\bar{\theta})^{-1}$ must also be positive semi-definite. If we go one step further and let only one of the first N elements of $\tilde{v}$ and $\bar{v}$ be arbitrary and non-zero, then

$$(329) \qquad \tilde{v}^H \tilde{Q}_2 \tilde{v} = \bar{v}^H Q_2 \bar{v} \geq 0$$

implies the elements on the diagonal of the difference of matrices, $\tilde{Q}_2 = C - F(\bar{\theta})^{-1}$ must be greater than or equal to zero. Therefore, the elements on the diagonal of C must be greater than or equal to the elements on the diagonal of $F(\bar{\theta})^{-1}$. Thus proving the first property of the CRB.

C.2 Property Two

The second property from the theory of the CRB states that an unbiased estimator, $g(\bar{y})$, may be found that attains the CRB, in that $C = F(\bar{\theta})^{-1}$, if and only if

$$(330) \qquad \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} = F(\bar{\theta}) \left(g(\bar{y}) - \bar{\theta} \right).$$

We prove the if portion of this statement by assuming

$$(331) \qquad \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} = F(\bar{\theta}) \left(g(\bar{y}) - \bar{\theta} \right).$$

This implies

$$(332)$$

$$
\begin{aligned}
\left(\tfrac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \right) \left(\tfrac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \right)^H &= F(\bar{\theta}) \left(g(\bar{y}) - \bar{\theta} \right) \left(g(\bar{y}) - \bar{\theta} \right)^H F(\bar{\theta})^H \\
\Rightarrow \quad E\left\{ \left(\tfrac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \right) \left(\tfrac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \right)^H \right\} &= E\left\{ F(\bar{\theta}) \left(g(\bar{y}) - \bar{\theta} \right) \left(g(\bar{y}) - \bar{\theta} \right)^H F(\bar{\theta})^H \right\} \\
\Rightarrow \quad F(\bar{\theta}) &= F(\bar{\theta}) C F(\bar{\theta})^H \\
\Rightarrow \quad F(\bar{\theta})^{-1} &= C \\
\Rightarrow \quad F(\bar{\theta})^{-1} &= E\left\{ \left(g(\bar{y}) - \bar{\theta} \right) \left(g(\bar{y}) - \bar{\theta} \right)^H \right\}
\end{aligned}
$$

Recall that a parameter error covariance matrix of the form

$$(333) \qquad E\left\{ \left(g(\bar{y}) - \bar{\theta} \right) \left(g(\bar{y}) - \bar{\theta} \right)^H \right\}$$

is predicated on an unbiased estimator defined by

$$(334) \qquad E\left\{ g(\bar{y}) \right\} = \bar{\theta}.$$

Therefore, if

(335)
$$\frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} = F(\bar{\theta}) \left(g(\bar{y}) - \bar{\theta} \right),$$

then $g(\bar{y})$ is the unbiased estimator that attains $C = F(\bar{\theta})^{-1}$.

For the only if portion of the proof, assume $g(\bar{y})$ exists and is unbiased. Recall that when an estimator is unbiased,

(336)
$$E\left\{ \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \left(g(\bar{y}) - \bar{\theta} \right)^H \right\} = \mathrm{I}.$$

This implies

(337)
$$\left[E\left\{ \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \left(g(\bar{y}) - \bar{\theta} \right)^H \right\} \right]^2 = \mathrm{I}^2 = \mathrm{I}.$$

By Schwartz' inequality for random variables (61:227),

(338)
$$\left[E\left\{ \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \left(g(\bar{y}) - \bar{\theta} \right)^H \right\} \right]^2 \leq$$
$$E\left\{ \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}}^H \right\} E\left\{ \left(g(\bar{y}) - \bar{\theta} \right) \left(g(\bar{y}) - \bar{\theta} \right)^H \right\}$$

which can also be represented as

(339)
$$\mathrm{I} \leq F(\bar{\theta})C.$$

The matrix inequality of Equations 338 and 339 implies that the difference of matrices, $F(\bar{\theta})C - \mathrm{I}$, is positive semi-definite and transforms to an equality if and only

189

if

$$(340) \qquad \frac{\partial \ln f(\bar{y}; \bar{\theta})}{\partial \bar{\theta}} = F(\bar{\theta}) \left(g(\bar{y}) - \bar{\theta} \right).$$

Thus proving the second property from the theory of the CRB.

Appendix D. Sequential Quadratic Programming Gradients

D.1   Cost Function Gradient

As developed in Chapter II, the exact maximum likelihood cost function in terms of the linear prediction coefficients, $\bar{b}$, is

$$(341) \qquad\qquad L(\bar{b}) = \bar{y}^H B^H (BB^H)^{-1} B\bar{y}$$

where

$$(342) \qquad B = \begin{bmatrix} b_N & b_{N-1} & \cdots & b_0 & 0 & \cdots & 0 \\ 0 & b_N & b_{N-1} & \cdots & b_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & b_N & b_{N-1} & \cdots & b_0 \end{bmatrix}_{(M-N)\times M} .$$

Because L is a scalar function, the gradient of L with respect to $\bar{b}$ takes the form

$$(343) \qquad \frac{\partial L(\bar{b})}{\partial \bar{b}} = \begin{bmatrix} \frac{\partial L(\bar{b})}{\partial b_1} \\ \frac{\partial L(\bar{b})}{\partial b_2} \\ \vdots \\ \frac{\partial L(\bar{b})}{\partial b_N} \end{bmatrix} .$$

Let B^+ represent the pseudo-inverse $B^H(BB^H)^{-1}$ so that

$$(344) \qquad\qquad L(\bar{b}) = \bar{y}^H B^+ B\bar{y}$$

and

$$\begin{aligned}
\frac{\partial L(\bar{b})}{\partial \bar{b}} &= \frac{\partial L(\bar{y}^H B^+ B \bar{y})}{\partial \bar{b}} \\
&= \bar{y}^H \frac{\partial L(B^+ B)}{\partial \bar{b}} \bar{y}.
\end{aligned}$$

(345)

In our application, we programmed the SQP algorithm for the two mode problem, but because the projection matrix, $B^+ B$, is idempotent and Hermitian, we can use the chain rule to express each partial derivative of the gradient, $\frac{\partial(B^+ B)}{\partial \bar{b}}$, in a generalized form. This insight is conveyed by Magnus and Neudecker in (49). Since

(346)
$$B^+ B = B^+ B B^+ B$$

then

$$\begin{aligned}
\frac{\partial(B^+ B)}{\partial b_n} &= \frac{\partial(B^+ B B^+ B)}{\partial b_n} \\
&= B^+ B \frac{\partial(B^+ B)}{\partial b_n} + \frac{\partial(B^+ B)}{\partial b_n} B^+ B \\
&= B^+ B \frac{\partial(B^+ B)}{\partial b_n} + \left(B^+ B \frac{\partial(B^+ B)}{\partial b_n} \right)^H.
\end{aligned}$$

(347)

Also from Equation 346

(348)
$$\begin{aligned}
B &= B B^+ B \\
\Rightarrow \frac{\partial B}{\partial b_n} &= \frac{\partial(B B^+ B)}{\partial b_n} \\
\Rightarrow \frac{\partial B}{\partial b_n} &= \frac{\partial B}{\partial b_n} B^+ B + B \frac{\partial(B^+ B)}{\partial b_n}.
\end{aligned}$$

192

Rearranging Equation 348 we see

$$
\begin{aligned}
B\frac{\partial (B^+ B)}{\partial b_n} &= \frac{\partial B}{\partial b_n} - \frac{\partial B}{\partial b_n} B^+ B \\
&= \frac{\partial B}{\partial b_n}(\mathrm{I} - B^+ B).
\end{aligned}
$$

(349)

Finally, Equation 349 can be substituted back into Equation 347 so that

$$
(350) \qquad \frac{\partial (B^+ B)}{\partial b_n} = B^+ \frac{\partial B}{\partial b_n}(\mathrm{I} - B^+ B) + \left(B^+ \frac{\partial B}{\partial b_n}(\mathrm{I} - B^+ B) \right)^H .
$$

Therefore, calculating the gradient of the log-likelihood function, $L(\bar{b})$, is reduced to determining $\frac{\partial B}{\partial b_n}$ for each mode and inserting it back into Equations 350 and 345. With B defined by equation 342, the partial derivative of B with respect to each b_n is the identity matrix shifted for the specific n.

D.2 Boundary Gradients

The gradients of the boundaries of the feasibility region are also needed for the SQP algorithm. The boundary gradients for the two mode problem are straight forward.

D.2.1 Boundary A.

$$
\begin{aligned}
G_A(\bar{b}) &= -b_2 \\
\frac{\partial G_A(\bar{b})}{\partial \bar{b}} &= \begin{bmatrix} 0 \\ -1 \end{bmatrix}
\end{aligned}
$$

(351)

D.2.2 Boundary B.

$$
\begin{aligned}
G_A(\bar{b}) &= -b_1 - b_2 - 1 \\
\frac{\partial G_A(\bar{b})}{\partial \bar{b}} &= \begin{bmatrix} -1 \\ -1 \end{bmatrix}
\end{aligned}
$$

(352)

D.2.3 Boundary C.

$$
\begin{aligned}
G_A(\bar{b}) &= -\frac{1}{4}b_1^2 + b_2 \\
\frac{\partial G_A(\bar{b})}{\partial \bar{b}} &= \begin{bmatrix} -\frac{1}{2}b_1^2 \\ 1 \end{bmatrix}
\end{aligned}
$$

(353)

Bibliography

1. Abatzoglou, T. J., et al. "The Constrained Total Least Squares Technique and its Applications to Harmonic Superresolution," *IEEE Transactions on Signal Processing*, *39*(5):1070–1087 (May 1991).

2. Bernasconi, C. F. *Relaxation Kinetics*. New York, NY: Academic Press, 1976.

3. Bohme, A. "Least Squares Methods." *Handbook of Numerical Analysis* edited by P. G. Ciarlet and J. L. Lions, 467–647, Amsterdam, Holland: North Holland, 1992.

4. Bresler, Y. and A. Macovski. "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise," *IEEE Transactions on Acoustics, Speech and Signal Processing*, *34*(5):1081–1089 (October 1986).

5. Buchwald, W. R. and N. M. Johnson. "Revised Role for the Poole-Frenkel Effect in Deep-Level Characterization," *Journal of Applied Physics*, *64*(2):958–961 (1988).

6. Cadzow, J. A. "Total Least Squares, Matrix Enhancement, and Signal Processing," *Digital Signal Processing*, *4*:21–39 (August 1969).

7. Cheney, E. W. and D. R. Kincaid. *Numerical Mathematics and Computing*. Monterey, CA: Brooks/Cole Publishing, 1985.

8. Clark, M. P. and L. L. Scharf. "On the Complexity of IQML Algorithms," *IEEE Transactions on Signal Processing*, *40*(4):1811–1813 (July 1992).

9. de Prony, G. R. "Essai Experimantal et Analytique: sur les Lois de la Dilatabilite de Fluides Elastiques et Sur Celles de la Force Expansive de la Vapeur de l'Eau et de la Vapeur de l'Alcool, a Differentes Temperatures," *Journal de l'Ecole Polytechnique*, *1*(2):24–76 (1795).

10. Doolittle, W. A. and A. Rohatgi. "A Novel Computer Based Pseudo-Logarithmic Capacitance/Conductance DLTS System Specifically Designed for Transient Analysis," *Review of Scientific Instrumentation*, *63*(12):5733–5741 (December 1992).

11. Doolittle, W. A. and A. Rohatgi. "Comparison of Covariance Linear Predictive Modeling to the Modulation Function Method for Use in Deep Level Transient Spectroscopy," *Journal of Applied Physics*, *75*(9):4560–4569 (May 1994).

12. Doolittle, W. A. and A. Rohatgi. "A New Figure of Merit and Methodology for Quantitatively Determining Defect Resolution Capabilities in Deep Level Transient Spectroscopy Analysis," *Journal of Applied Physics*, *75*(9):4570–4575 (May 1994).

13. Dowling, E. M. and R. D. DeGroat. "The Equivalence of the Total Least Squares and Minimum Norm Methods," *IEEE Transactions on Signal Processing*, *39*(8) (August 1991).

195

14. Elsaesser, D. W. *Excitation and De-Excitation Mechanisms of Er-Doped GaAs and AlGaAs*. PhD dissertation, Air Force Institute of Technology, Wright Patterson Air Force Base, OH, 1992.

15. Evans, A. G. and R. Fischl. "Optimal Least Squares Time-Domain Synthesis of Recursive Digital Filters," *IEEE Transactions on Audio and Electroacoustics*, *21*(1):61–65 (February 1973).

16. Evwaraye, A. O., et al. "Kinetics of Slow Buildup of Photoconductance in 6H-SiC," *Applied Physics Letters*, *66*(20):2691 (1995).

17. Forsythe, G. E., et al. *Computer Methods for Mathematical Computations*. Englewood Cliffs, NJ: Prentice Hall, 1977.

18. Golub, G. H. and C. F. Van Loan. "An Analysis of the Total Least Squares Problem," *SIAM Journal on Numerical Analysis*, *17*(6):883–93 (1980).

19. Golub, G. H. and C. F. Van Loan. *Matrix Computations*. Baltimore, MD: Johns Hopkins, 1993.

20. Grace, A. *Optimization Toolbox for Use with MATLAB*. Natick, MA: Math Works, 1992.

21. Graybill, F. A. *Introduction to Matrices with Applications in Statistics*. Belmont, CA: Wadsworth Publishing Company, 1969.

22. Hildebrand, F. B. *Introduction to Numerical Analysis*. New York, NY: McGraw Hill, 1974.

23. Hinshaw, W. S. and A. H. Lent. "An Introduction to NMR Imaging: From the Bloch Equation to the Imaging Equation," *Proceedings of the IEEE*, *71*(3):338–350 (March 1983).

24. Hogg, R. V. and A. T. Craig. *Introduction to Mathematical Statistics*. New York, NY: Macmillan Publishing, 1978.

25. Horn, R. A. and C. A. Johnson. *Matrix Analysis*. New York, NY: Cambridge University Press, 1985.

26. Hua, Y. "The Most Efficient Implementation of the IQML Algorithm," *IEEE Transactions on Signal Processing*, *42*(8):2203–2204 (August 1994).

27. Hua, Y. and T. K. Sarkar. "On the Total Least Squares Linear Prediction Method for Frequency Estimation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, *38*(12):2186–2189 (December 1990).

28. Huffel, S. Van. *The Total Least Squares Problem: Computational Aspects and Analysis*. Philadelphia, PA: SIAM, 1991.

29. Huffel, S. Van and J. Vandewalle. "Algebraic Connections Between the Least Squares and Total Least Squares Problems," *Numerische Mathematik*, *55*:431–449 (1989).

30. Ingham, E. A., et al. "Deriving the Maximum Likelihood Cost Function with Linear Prediction for the Summed Exponentials Problem," *submitted to IEEE Transactions on Signal Processing* (1995).

31. Ingham, E. A., et al. "Improved Linear Prediction for Deep Level Transient Spectroscopy Analysis," *to appear in the Journal of Applied Physics* (1996).

32. Isenberg, I., et al. "Studies on Analysis of Fluorescence Decay by the Method of Moments," *Biophysical Journal*, *13*:1090–1115 (1973).

33. Kay, S. M. *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice Hall, 1988.

34. Kay, S. M. "A Fast and Accurate Single Frequency Estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, *37*(12):1987–1989 (December 1989).

35. Kay, S. M. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

36. Kumaresan, R. "On the Zeros of the Linear Prediction-Error Filter for Deterministic Signals," *IEEE Transactions on Acoustics, Speech and Signal Processing*, *31*(1):217–220 (February 1983).

37. Kumaresan, R. "On a Frequency Domain Analog of Prony's Method," *IEEE Transactions on Acoustics, Speech and Signal Processing*, *38*(1):168–170 (January 1990).

38. Kumaresan, R. "Spectral Analysis." *Handbook for Digital Signal Processing* edited by S. K. Mitra and J. F. Kaiser, 1143–1242, New York, NY: John Wiley & Sons, 1993.

39. Kumaresan, R., et al. "An Algorithm for Pole-Zero Modeling and Spectral Analysis," *IEEE Transactions on Acoustics, Speech and Signal Processing*, *34*(3):637–640 (June 1986).

40. Kumaresan, R. and A. K. Shaw. "Superresolution by Structured Matrix Approximation," *IEEE Transactions on Antennas and Propagation*, *36*(1):34–44 (January 1988).

41. Kumaresan, R. and D. W. Tufts. "Estimating the Parameters of Exponentially Damped Sinusoids and Pole-Zero Modeling in Noise," *IEEE Transactions on Acoustics, Speech and Signal Processing*, *30*(6):833–840 (December 1982).

42. Kumaresan, R. and D. W. Tufts. "Estimating the Angles of Arrival of Multiple Plane Waves," *IEEE Transactions on Aerospace and Electronic Systems*, *19*(1):134–139 (January 1983).

43. Lagowski, J., et al. "Origin of the .82 *eV* Electron Trap in GaAs and Its Annihilation by Shallow Donors," *Applied Physics Letters*, *40*:342–344 (1982).

44. Lanczos, C. *Applied Analysis*. Englewood Cliffs, NJ: Prentice Hall, 1956.

45. Lang, D. V. "Deep-Level Transient Spectroscopy: A New Method to Characterize Traps in Semiconductors," *Journal of Applied Physics, 45*(7):3023–3032 (July 1974).

46. Lang, D. V. "Space-Charge Spectroscopy in Semiconductors." *Topics in Applied Physics 37*, edited by P. Braunlich, 93–133, New York, NY: Springer Verlag, 1979.

47. Lang, S. W. and J. H. McClellan. "Frequency Estimation with Maximum Entropy Spectral Estimators," *IEEE Transactions on Acoustics, Speech and Signal Processing, 28*(6):716–724 (December 1980).

48. Look, D. C. "The Electrical and Photoelectronic Properties of Semi-Insulating GaAs." *Semiconductors and Semimetals 19*, edited by R. K. Willardson and A. C. Beer, 75–170, New York, NY: Academic Press, 1983.

49. Magnus, J. R. and H. Neudecker. *Matrix Differential Calculus*. New York, NY: John Wiley & Sons, 1988.

50. Makoul, J. "Linear Prediction: A Tutorial Review," *Proceedings of the IEEE, 63*(4):561–580 (April 1975).

51. Mirsky, L. "Symmetric Gauge Functions and Unitarily Invariant Norms," *Quarterly Journal of Mathematics, 11*:50–59 (1960).

52. Nagesha, V and S. Kay. "On Frequency Estimation with the IQML Algorithm," *IEEE Transactions on Signal Processing, 42*(9):2509–2513 (September 1994).

53. Neudeck, G. W. "The PN Junction Diode." *Modular Series on Solid State Devices 2*, edited by G. W. Neudeck and R. F. Pierret, Reading, MA: Addison Wesley, 1983.

54. Papoulis, A. *Probability, Random Variables, and Stochastic Processes*. New York, NY: McGraw Hill, 1991.

55. Pierret, R. F. "Semiconductor Fundamentals." *Modular Series on Solid State Devices 1*, edited by R. F. Pierret and G. W. Neudeck, Reading, MA: Addison Wesley, 1983.

56. Porat, B. *Digital Processing of Random Signals*. Englewood Cliffs, NJ: Prentice Hall, 1994.

57. Rahman, A. and K. Yu. "Total Least Squares Approach for Frequency Estimation Using Linear Prediction," *IEEE Transactions on Acoustics, Speech and Signal Processing, 35*(10):1440–1454 (October 1987).

58. Ransom, C. M., et al. "Modulating Functions Waveform Analysis of Multi-Exponential Transients for Deep-Level Transient Spectroscopy." *Materials Research Society Symposia Proceedings 69*. 337–342. 1986.

59. Sacchini, J. J. *Development of Two-Dimensional Parametric Radar Signal Modeling and Estimation Techniques with Application to Target Identification.* PhD dissertation, The Ohio State University, Columbus, OH, 1992.

60. Sacchini, J. J., et al. "Evaluation of Single and Full-Polarization Two-Dimensional Prony Techniques Applied to Radar Data." *Proceedings of the SPIE.* 91–105. Orlando, FL: SPIE, April 1994.

61. Scharf, L. L. *Statistical Signal Processing.* Reading, MA: Addison Wesley, 1990.

62. Scofield, J. D., et al. "Experimental Observation of Inequivalent Lattice Site Energies for Deep Level Traps in 6H-SiC by Isothermal Deep Level Transient Spectroscopy," *to be submitted to the Journal of Applied Physics* (1996).

63. Scofield, J. D., et al. "Deep Level Investigation of Bulk and Epitaxial 6H-SiC at High Temperatures." *Proceedings of the 22nd ISCS Conference.* Bristol, UK: IOP Publishing, 1995.

64. Seber, G. A. F. *Linear Regression Analysis.* New York, NY: John Wiley & Sons, 1977.

65. Shapiro, F. R., et al. "The Use of Linear Predictive Modeling for the Analysis of Transients from Experiments on Semiconductor Defects," *Journal of Applied Physics,* $55(10){:}3453$–3459 (May 1984).

66. Shaw, A. K. "Maximum Likelihood Estimation of Multiple Frequencies with Constraints to Guarantee Unit Circle Roots," *IEEE Transactions on Signal Processing,* $43(3){:}796$–799 (March 1995).

67. Steedly, W. and R. Moses. "The Cramer-Rao Bound for Pole and Amplitude Coefficient Estimates of Damped Exponential Signals in Noise," *IEEE Transactions on Signal Processing,* $41(3){:}1305$–1318 (March 1993).

68. Stewart, G. W. "On the Invariance of Perturbed Null Vectors Under Column Scaling," *Numerische Mathematik,* $44{:}61$–65 (1984).

69. Sze, S. M. *Semiconductor Devices Physics and Technology.* New York, NY: John Wiley & Sons, 1985.

70. Therrien, C. W. *Discrete Random Signals and Statistical Signal Processing.* Englewood Cliffs, NJ: Prentice Hall, 1992.

71. Tufts, D. W. and R. Kumaresan. "Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood," *Proceedings of the IEEE,* $70(9){:}975$–990 (September 1982).

72. Tufts, D. W. and R. Kumaresan. "Singular Value Decomposition and Improved Frequency Estimation Using Linear Prediction," *IEEE Transactions on Acoustics, Speech and Signal Processing,* $30(4){:}671$–675 (August 1982).

73. Vincent, G., et al. "Electric Field Effect on the Thermal Emission of Traps in Semiconductor Junctions," *Journal of Applied Physics,* $50{:}5484$–5487 (1979).

74. Ziskind, I. and M. Wax. "Maximum Likelihood Localization of Multiple Sources by Alternating Projection," *IEEE Transactions on Acoustics, Speech and Signal Processing*, *36*(10):1553–1560 (October 1988).

Vita

Major Edward A. Ingham ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬, Washington. He graduated from high school in Edmonds, Washington in June of 1980. He then attended the United States Air Force Academy where he obtained a Bachelor of Science in Engineering Sciences on 30 May 1984. He was commissioned a second lieutenant and attended Undergraduate Pilot Training at Williams AFB, Arizona. After training, he continued at Williams as a first assignment instructor pilot in the T-37. In May of 1988, he upgraded to the F-16C and was assigned to the 14th Fighter Squadron, Misawa Air Base, Japan. He entered the Graduate School of Engineering, Air Force Institute of Technology, in May of 1992. At the Air Force Institute of Technology, Major Ingham completed a Masters of Science in Systems Engineering. He has continued to the present in a Ph.D. program with the department of Electrical and Computer Engineering. Major Ingham and his wife Renee were married in 1985. They have two sons: Connan and Dillon.

201