September 1989 UILU-ENG-89-2224

(2)

COORDINATED SCIENCE LABORATORY
College of Engineering

IDENTIFICATION OF IMPULSIVE INTERFERENCE CHANNELS

Serena M. Zabin

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

REPORT DOCUMENTATION PAGE

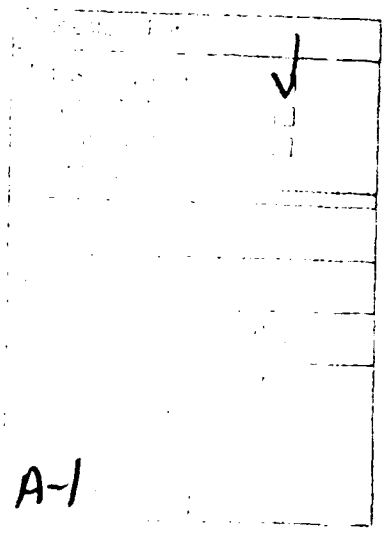
1a REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS None
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) UILU-ENG-89-2224	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Coordinated Science Lab University of Illinois 6b. OFFICE SYMBOL (If applicable) N/A	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code) 1101 W. Springfield Ave. Urbana, IL 61801	7b. ADDRESS (City, State, and ZIP Code) 800 N. Quincy St. Arlington, VA 22217
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Joint Services Electronics Program 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-84-C-0149
8c. ADDRESS (City, State, and ZIP Code) 800 N. Quincy St. Arlington, VA 22217	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.
11 TITLE (Include Security Classification) "Identification of Impulsive Interference Channels"	
12. PERSONAL AUTHOR(S) Serena M. Zabin	
13a. TYPE OF REPORT Technical 13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Year, Month, Day) July 1989 15. PAGE COUNT 126
16. SUPPLEMENTARY NOTATION	
17. COSATI CODES FIELD / GROUP / SUB-GROUP	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) impulsive noise, non-Gaussian channels, Class A Middleton interference model, parameter estimation, identification

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

In this work, the problem of optimum and near-optimum identification of the parameters of the Middleton Class A impulsive interference model is considered. In particular, under the assumption of the availability of a set of independent samples from the Class A envelope distribution, the problems of basic batch estimation of the Class A parameters, recursive identification of the parameters, and efficient estimation of the parameters for small sample sizes, are investigated. Within the context of basic batch estimation, several estimators of the parameters are proposed and their asymptotic performances explored. From this analysis, estimates based on the method of moments are seen to be consistent and computationally desirable but highly inefficient, whereas more efficient likelihood-based estimators are seen to be computationally unwieldy. However, an estimator that initiates likelihood iteration with the method-of-moments estimates is seen to overcome these difficulties in its asymptotic performance. Unfortunately, simulation of this third estimator for moderate sample sizes reveals poor performance under these conditions. To overcome this lack of moderate-sample-size efficiency, a similar estimator that initiates likelihood iteration

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code) / 22c. OFFICE SYMBOL

with physically motivated (but nonoptimal) estimates is also proposed. Simulation of this latter estimator for moderate sample sizes indicates that near-optimal performance is obtained by this technique. Within the context of recursive estimation, a recursive decision-directed estimator for on-line identification of the parameters of the Class A model is proposed. This estimator is based on an adaptive, Bayesian classification of each of a sequence of Class A envelope samples as either an impulsive sample or as a background sample. The performance characteristics of this algorithm are investigated, and an appropriately modified version is found to yield a global, recursive estimator of the parameters that performs very well for all parameter vectors in the parameter set of interest. Within the context of efficient estimation for small sample sizes, an algorithm that has the potential of providing efficient estimates of the Class A parameters for small sample sizes is proposed. For the single-parameter estimation problem, it is shown that the sequence of estimates obtained via this algorithm converges, and a characterization of the point to which the sequence converges is given. For both the single-parameter and two-parameter estimation problems, it is also seen, via an extensive simulation study, that the proposed estimator yields excellent estimates of the parameters for small sample sizes. It is anticipated that the results of this research will have widespread impact in the areas of communications, radar, and sonar due to the common occurrence of impulsive noise channels in these systems.



IDENTIFICATION OF IMPULSIVE INTERFERENCE CHANNELS

BY

SERENA M. ZABIN

B.S., University of Illinois, 1983
M.S., University of Illinois, 1985

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1989

Urbana, Illinois

IDENTIFICATION OF IMPULSIVE INTERFERENCE CHANNELS

Serena M. Zabin, Ph.D.
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, 1989

In this work, the problem of optimum and near-optimum identification of the parameters of the Middleton Class A impulsive interference model is considered. In particular, under the assumption of the availability of a set of independent samples from the Class A envelope distribution, the problems of basic batch estimation of the Class A parameters, recursive identification of the parameters, and efficient estimation of the parameters for small sample sizes, are investigated. Within the context of basic batch estimation, several estimators of the parameters are proposed and their asymptotic performances explored. From this analysis, estimates based on the method of moments are seen to be consistent and computationally desirable but highly inefficient, whereas more efficient likelihood-based estimators are seen to be computationally unwieldy. However, an estimator that initiates likelihood iteration with the method-of-moments estimates is seen to overcome these difficulties in its asymptotic performance. Unfortunately, simulation of this third estimator for moderate sample sizes reveals poor performance under these conditions. To overcome this lack of moderate-sample-size efficiency, a similar estimator that initiates likelihood iteration with physically motivated (but nonoptimal) estimates is also proposed. Simulation of this latter estimator for moderate sample sizes indicates that near-optimal performance is obtained by this technique. Within the context of recursive estimation, a recursive decision-directed estimator for on-line identification of the parameters of the Class A model is proposed. This estimator is based on an adaptive, Bayesian classification of each of a sequence of Class A envelope samples as either an impulsive sample or as a background sample. The performance characteristics of this algorithm are investigated, and an appropriately modified version is found to yield a global, recursive estimator of the parameters that performs very well for all parameter vectors in the parameter set of interest. Within the context of efficient estimation for small sample sizes, an algorithm that has the

potential of providing efficient estimates of the Class A parameters for small sample sizes is proposed. For the single-parameter estimation problem, it is shown that the sequence of estimates obtained via this algorithm converges, and a characterization of the point to which the sequence converges is given. For both the single-parameter and two-parameter estimation problems, it is also seen, via an extensive simulation study, that the proposed estimator yields excellent estimates of the parameters for small sample sizes. It is anticipated that the results of this research will have widespread impact in the areas of communication, radar, and sonar due to the common occurrence of impulsive noise channels in these systems.

To my mother

TABLE OF CONTENTS

1. INTRODUCTION

Communication systems are seldom interfered with by white Gaussian noise alone, yet receiving systems in general use are those which are optimum for white Gaussian noise. The man-made electromagnetic environment, and much of the natural one, is basically "impulsive," i.e., it has a highly structured form characterized by significant probabilities of large interference levels. In addition to the man-made electromagnetic environment, there many other different, common and widely-used communications and remote-sensing type channels where impulsive noise dominates, e.g., extra-low-frequency (ELF) channels, urban radio networks, underwater acoustic channels, and telephone line channels. This is in contrast to the Gaussian noise processes inherent in transmitting and receiving elements. Since the conventional receivers are effectively linear, the impulsive character of the interference can drastically degrade the performance of conventional systems. In fact, it has been well established [1]-[5] that the performance of communications, radar, and sonar systems operating in impulsive channels can be greatly enhanced if the true statistics of the channel are known and exploited. Consequently, the problem of identifying impulsive noise channels is an important one in the development of systems that can approach the performance limits set by such channels. To do so, one first needs to develop a model for the interference that fits available measurements, is physically meaningful when the nature of the noise sources, their distributions in time and space, propagation, etc., are taken into account, is directly relatable to the physical mechanisms giving rise to the interference, and is tractable for signal detection problems.

A physically-meaningful and widely-used model for impulsive interference that satisfies the above requirements is the so-called Class A Middleton model [6]-[8]. This model is parametric, with parameters that can be adjusted to fit a great variety of non-Gaussian noise phenomena arising in practice. The parametric nature of this model makes it amenable to identification through point estimation techniques. Furthermore, this model, which features a non-Gaussian impulsive component, superimposed on a Gaussian background component, has found wide application in several problems of practical interest (see, e.g., [5],[9]). A complete description of Middleton's Class A noise model, including its derivation, further motivation, and taxonomy, can be found in [5],[6],[8].

This study is devoted to the problem of obtaining global optimal and near-optimal identification procedures for the Class A interference model. The problem of estimating the parameters of the Class A model was first considered by Middleton in [10] and [11]. In these studies, Middleton outlines three methods for determining the model parameters. The first is an empirical procedure whereby the parameters are obtained graphically from the experimentally determined distribution function. Expressions for the parameters in terms of the moments of the Class A probability distribution function are given in the second method. The third procedure requires that experimental values of the distribution function and its slope at vanishingly small thresholds be available. The parameters are then determined via two relations involving these measurements. Other work on the Class A estimation problem includes that of Powell and Wilson [12], wherein standard batch estimation techniques are used to estimate the parameters for a restricted range of parameter values.

We begin this study with an overview of the Class A interference model, which is given in Chapter 2. In Chapter 3, the problem of basic batch estimation of the Class A parameters from an independent sequence of Class A samples is considered. In particular, within the context of batch estimation, our goal is two-fold : (i) to obtain a consistent and asymptotically efficient estimator of the parameters and, (ii) to obtain a practical estimator of these parameters which performs well for moderate sample sizes. The problem of recursive identification of the Class A parameters is addressed in Chapter 4. Our objective here is to obtain a global recursive estimator of the parameters which performs well for all parameter vectors in the parameter set of interest. In Chapter 5, we develop an efficient estimator of the paramters with good small-sample-size performance globally. A summary of the research results is given in Chapter 6.

2. THE CLASS A INTERFERENCE MODEL

In this study, we will focus our attention on the model defined in [8] as the "strictly canonical" Class A interference model. An overview of the model will be given in this section. Further details of this model can be found in [6]-[8].

In Middleton's strictly canonical Class A noise model, the received interference is assumed to be a process having two independent components :

$$Z(t) = Z_P(t) + Z_G(t) \quad .$$

The first component, Z_P, is represented by

$$Z_P(t) = \sum_j U_j(t, \gamma) \quad .$$

where U_j denotes the j-th waveform from an interfering source and γ represents a set of random parameters which describes the waveform scale and structure. It is next assumed that only one type of waveform, U, is generated, with variations in the individual waveforms accounted for by appropriate statistical treatment of the parameters in γ, and the generic waveform $U(t)$ is obtained explicitly from the underlying physical mechanisms [6]. Under the additional assumption that the sources emit their waveforms independently according to the Poisson distribution in time, the first-order characteristic function for Z_P is given by (see, e.g., [7])

$$F(i\xi)_P = \exp[<AJ_0(B_0\xi) - A>]$$

where B_0 denotes the envelope of U when U is written in envelope and phase form, J_0 is the Bessel function of order zero, and $<\cdot>$ denotes required statistical averages over the random epoch representing the time at which the typical j-th source emits, Doppler velocities (if any), and the random signal parameters in γ. The second component, Z_G, is an additive stationary Gaussian background process attributable either to receiver noise or to the limit of a high density Poisson process representing the contributions of unresolvable background sources, or both. Hence, under the assumption that this background component has zero mean and variance σ_G^2, its first-order

characteristic function is

$$F(i\xi)_G = e^{-\xi^2\sigma_G^2/2}.$$

and the overall characteristic function for the process is then given by

$$F(i\xi)_{P+G} = F(i\xi)_P F(i\xi)_G .$$

which can be approximated canonically as follows [7],[8]:

$$F(i\xi)_{P+G} \doteq e^{-A} \sum_{m=0}^{\infty} \frac{A^m}{m!} e^{-c_m^2\xi^2/2}, \tag{2.1}$$

where

$$c_m^2 = m <B_o^2>/2 + \sigma_G^2 \quad .$$

and A is a positive parameter to be defined below. For computational purposes, it is convenient to consider the normalized variable

$$Z = X/(<X_G^2> + <X_P^2>)^{1/2}.$$

Transforming (2.1) for the normalized variable Z yields the desired probability density function (pdf):

$$p_Z(z) \doteq e^{-A} \sum_{m=0}^{\infty} \frac{A^m}{m!\sqrt{2\pi}\sigma_m} e^{-z^2/2\sigma_m^2} \quad , \tag{2.2}$$

where

$$\sigma_m^2 \triangleq \frac{\frac{m}{A} + \Gamma}{1 + \Gamma} \quad ,$$

and where Γ is a second parameter (also to be defined below). The corresponding Class A envelope pdf is given by

$$w(z) = \begin{cases} 2e^{-A} \sum_{m=0}^{\infty} \frac{A^m}{m!\sigma_m^2} z\, e^{-z^2/\sigma_m^2} & z \geq 0 \\ 0 & z < 0 \end{cases} \tag{2.3}$$

It is the envelope statistics which will be used in the estimation problem addressed here. Note that this envelope pdf consists of an infinite mixture of weighted Rayleigh densities. The $m = 0$ term is attributable to the background component of the input interference,[1] whereas all terms indexed by $m \geq 1$ are attributable to the impulsive (Poisson) component of the input interference plus an appropriate contribution from the background component of the interference.

The two basic and traditional parameters of the model are A and Γ. Let us consider their definitions and physical significance:

i) A is the "Overlap Index" or "Nonstructure Index." Specifically,

$$A \triangleq \nu \bar{T}_s \tag{2.4}$$

where ν is the average number of emission events impinging on the receiver per second and $\bar{T}_s$ is the mean duration of a typical interfering source emission. Note that ν is simply the rate of the Poisson process underlying the impulsive part of the interference. Thus, A is a measure of the amount of temporal overlap among the interfering signals. The smaller A is, the fewer the number of emission "events" and/or their durations so that the (instantaneous) noise properties are dominated by the waveform characteristics of individual events. As A is made larger, the noise becomes less structured, i.e., the statistics of the instantaneous amplitude approach the Gaussian distribution (asymptotically as $A \to \infty$, although $A \cong 10$ is considered a large value for A).

---

[1]Note that $\sigma_0^2 = \Gamma/(1+\Gamma)$. It follows from the definition of Γ (given in (2.5)) that $\sigma^2(\Gamma/(1+\Gamma))$ is simply the intensity of the Gaussian component of the input interference. Thus, even though the $m = 0$ term appears to depend on the Class A model parameters, via some simple manipulations, it can be seen that the only quantity it actually depends on is the intensity of the Gaussian component of the input interference. Consequently, the $m = 0$ term is entirely attributable to this background component.

ii) Γ is called the "Gaussian factor." It is the ratio of the intensity of the independent Gaussian component of the input interference, σ_G^2, to the intensity Ω_{2A} of the non-Gaussian component, i.e.,

$$\Gamma \triangleq \langle Z_G^2 \rangle / \langle Z_P^2 \rangle = \frac{\sigma_G^2}{\Omega_{2A}} \quad \text{where} \quad \Omega_{2A} \triangleq A \langle B_o^2 \rangle / 2. \tag{2.5}$$

By adjusting the parameters A and Γ, the density in (2.2) can be made to fit a great variety of non-Gaussian noise densities. In particular, the Class A model is appropriate for interference caused by intentionally radiated signals (e.g., as in the crowded HF band) and has also found considerable application in various acoustical (e.g., sonar) problems. Examples of Class A interference include (depending on the receiver bandwidth) the emissions of various man-made devices such as radio frequency dielectric heaters, soldering machines, plastic welders, etc., as well as natural phenomena such as grinding arctic ice plates. Typical values for the parameters include $(A = 10^{-4}, \Gamma = 50)$ for narrowband interference from ore-crushing machinery and $(A = 0.35, \Gamma = 5.0 \times 10^{-4})$ for power-line emissions.

Although A and Γ are the traditional parameters of the Class A model, instead of estimating A and Γ, we will consider the problem of estimating A and K, where

$$K \triangleq A\,\Gamma \ , \tag{2.6}$$

i.e.,

$$K = 2\,\sigma_G^2 / \langle B_o^2 \rangle .$$

This reparametrization proves useful since it increases the analytical tractability of the estimation problem. Finally, throughout this study, where specific values of A and K are considered, we will restrict our attention to the parameter set

$$\Lambda \triangleq \{(A, K)^T \in \mathbb{R}^2 : \ 10^{-2} \leq A \leq 1 \ \text{ and } \ 10^{-6} \leq K \leq 10^{-2}\}$$

since this is the range of usual practical interest for these parameters (see, e.g., [9],[10].)

3. BASIC BATCH ESTIMATION

3.1. Introduction

In this chapter, we will consider the problem of basic batch estimation of the Class A parameters from an independent sequence of Class A envelope samples. Within the context of batch estimation, our goal is two-fold: (i) the first goal is to provide an asymptotically optimal estimator of the parameters of Middleton's strictly canonical Class A noise model; and (ii) the second goal is to provide a practical estimator for these parameters that performs well for moderate sample sizes.

The starting point in this study is an estimator proposed in [10] based on the method of moments, in which parameter estimates are chosen to make population moments agree with sample moments. In Section 3.2 we provide an analysis of the asymptotic performance of this estimator. We show that, although this estimator is strongly consistent, its asymptotic variance for one parameter can be unacceptably high due to the insensitivity of population moments to this parameter. We then turn, in Section 3.3, to the problem of asymptotically efficient estimation. We first analyze the estimation potential in the Class A model by considering the inverse of Fisher's information matrix for the model in the parameter ranges of practical interest. It is seen from this analysis that the Class A model can, in fact, afford good estimates of all of its parameters if efficiency can be achieved.

We then consider two estimators that can achieve efficiency. The first of these is maximum-likelihood, which proves to be unwieldy due to root multiplicity problems in the likelihood equation and to poorly behaved gradients. The second estimator is one which corrects these two difficulties by initiating likelihood search with the method-of-moments estimates considered in Section 3.2. Because of the consistency of the moments estimator, this second estimator retains the efficiency of maximum-likelihood without the associated computational problems. Unfortunately, simulations show that this estimator does not perform well for moderate sample sizes for most parameter values of interest due to the extremely low efficiency of the initiating estimator at these

parameter values. Thus, although this estimator is attractive from a theoretical viewpoint, its use as a practical estimator is limited. However, as we show in Section 3.4, its basic feature of doing likelihood iteration to improve an initial estimator can be used to develop a very good practical estimator. In particular, in Section 3.4 we explore (via simulation) the moderate-sample-size performance of such an estimator initiated with a physically-based estimator motivated by a procedure proposed in [10]. Our simulation of this estimator indicates that it achieves practical efficiency for moderate sample sizes.

Some concluding remarks are contained in Section 3.5.

3.2. A Method-of-Moments Estimator

The method of moments is a simple, intuitively appealing, and computationally expedient estimation technique introduced by K. Pearson in 1894. The problem of estimating the parameters of the Class A model via this method has been considered by Middleton in [10],[11]. In this section, the asymptotic properties of these estimates are analyzed. In particular, it will be shown that the method of moments yields estimates of the Class A model parameters which are strongly consistent and aymptotically normal. Furthermore, explicit expressions for the asymptotic variances of the normalized estimates will be obtained and computed for a broad range of parameter values. The performance of the estimator will then be evaluated on the basis of these computations.

3.2.1. Parameter estimates

Let $Z_1,\ldots,Z_n$ be a random sample from the Class A envelope distribution $w(z)$ with unknown parameter vector $\underline{\theta} = (A, K)^T$ to be estimated. In the sequel, we assume that the observations $Z_i, i = 1,\ldots,n$ are independent. The method of moments consists of equating an appropriate number of sample moments to the corresponding moments of the distribution, which are functions of the unknown parameters. By considering as many moments as there are parameters to be estimated, and solving the resulting equations with respect to the parameters, estimates of the latter are obtained.

For the Class A envelope distribution, the most promising moments to use in this context are the fourth and sixth moments. These are the lowest-ordered even moments of interest (the second moment is constrained to unity), and no odd-ordered moments are obtainable in closed form. Use of higher-ordered even moments can result in multiple solutions when equating sample and population moments. Let μ_j denote the j-th moment of $w(z)$. Then, (see, e.g., [13]),

$$\mu_4 = \frac{2A}{(A+K)^2} + 2 \quad \text{and} \quad \mu_6 = \frac{6A}{(A+K)^3} + \frac{18A}{(A+K)^2} + 6. \tag{3.1}$$

For $A \neq 0$, hence for all $\underline{\theta} \in \Lambda$, inversion of (3.1) yields unique expressions for the parameters A and K in terms of μ_4 and μ_6. Specifically,

$$A = \frac{\left(\frac{\mu_4}{2} - 1\right)^3}{\left(\frac{\mu_6}{6} - \frac{3\mu_4}{2} + 2\right)^2} \triangleq f_1(\mu_4, \mu_6) \tag{3.2a}$$

and

$$K = \frac{\left(\frac{\mu_4}{2} - 1\right)}{\left(\frac{\mu_6}{6} - \frac{3\mu_4}{2} + 2\right)} - \frac{\left(\frac{\mu_4}{2} - 1\right)^3}{\left(\frac{\mu_6}{6} - \frac{3\mu_4}{2} + 2\right)^2} \triangleq f_2(\mu_4, \mu_6). \tag{3.2b}$$

The method-of-moments (MM) estimators based on these two moments, $\underline{\hat{\theta}}_n = (\hat{A}_n, \hat{K}_n)^T$, are then given by

$$\hat{A}_n = \frac{\left(\frac{\hat{m}_4}{2} - 1\right)^3}{\left(\frac{\hat{m}_6}{6} - \frac{3\hat{m}_4}{2} + 2\right)^2} \tag{3.3a}$$

and

$$\hat{K}_n = \frac{\left|\frac{\hat{m}_4}{2} - 1\right|}{\left|\frac{\hat{m}_6}{6} - \frac{3\hat{m}_4}{2} + 2\right|} - \frac{\left|\frac{\hat{m}_4}{2} - 1\right|^3}{\left|\frac{\hat{m}_6}{6} - \frac{3\hat{m}_4}{2} + 2\right|^2} \tag{3.3b}$$

where $\hat{m}_4$ and $\hat{m}_6$ denote the 4-th order and 6-th order sample moments, respectively, i.e., $\hat{m}_4 = \frac{1}{n}\sum_{j=1}^{n} Z_j^4$ and $\hat{m}_6 = \frac{1}{n}\sum_{j=1}^{n} Z_j^6$.

Consider (3.2a) and (3.2b). Note that f_1 and f_2 are discontinuous only on $\{(\mu_4, \mu_6)^T \in \mathbb{R}^2 : \frac{\mu_6}{6} - \frac{3\mu_4}{2} + 2 = 0\}$, i.e., on $\{(\mu_4, \mu_6)^T \in \mathbb{R}^2 : \mu_6 = 9\mu_4 - 12\}$. Using the expressions for μ_4 and μ_6 given in (3.1), we have that $\frac{\mu_6}{6} - \frac{3\mu_4}{2} + 2 = \frac{A}{(A+K)^3}$. Hence, if $A > 0$ and $K > 0$, then $\mu_6 > 9\mu_4 - 12$ and $\mu_4 > 2$, where the latter inequality follows from the expression for μ_4 given in (3.1). Thus, via relations (3.1), the parameter set Λ maps *into* the open set $\Omega \triangleq \{(\mu_4, \mu_6)^T \in \mathbb{R}^2 : \mu_4 > 2 \text{ and } \mu_6 > 9\mu_4 - 12\}$ on which f_1 and f_2 are defined and continuous. This fact will be used below.

3.2.2. Asymptotic properties of $\hat{\underline{\theta}}_n$

In this section we consider the asymptotic properties of the estimators of (3.3). These properties are summarized in the following two results.

Theorem 3.1. *(Consistency)* : The *MM* estimator $\hat{\underline{\theta}}_n$ is a strongly consistent estimator of $\underline{\theta}$ for all $\underline{\theta} \in \Lambda$.

Proof. Let $\underline{\mu}_{\underline{\theta}} \triangleq (\mu_4, \mu_6)^T$ and $\hat{\underline{m}} \triangleq (\hat{m}_4, \hat{m}_6)^T$. Since $Z_1, \ldots, Z_n$ are i.i.d. and $\underline{\mu}_{\underline{\theta}} \in \mathbb{R}^2$ for $\underline{\theta} \in \Lambda$, we have that $\hat{\underline{m}} \xrightarrow{a.s.} \underline{\mu}_{\underline{\theta}}$ by the Strong Law of Large Numbers (SLLN), which implies that

$$\lim_{n \to \infty} \hat{\underline{m}} = \underline{\mu}_{\underline{\theta}} \tag{3.4}$$

on a set w.p.1. Now, $A = f_1(\underline{\mu}_{\underline{\theta}})$ and $\hat{A}_n = f_1(\hat{\underline{m}})$. Thus, it follows from continuity of f_1 on Ω

and (3.4) that $\lim_{n\to\infty} \hat{A}_n = A$ on a set w.p.1. Similarly, since $K = f_2(\underline{\mu}_\theta)$ and $\hat{K}_n = f_2(\underline{\hat{m}})$, we have by continuity of f_2 on Ω and (3.4) that $\lim_{n\to\infty} \hat{K}_n = K$ on a set w.p.1. Thus, $\underline{\hat{\theta}}_n \overset{a.s.}{\to} \underline{\theta}$ and the proof of consistency is complete. □

Theorem 3.2. *(Asymptotic Normality)* : For each $\underline{\theta} \in \Lambda$, $\sqrt{n}\,(\underline{\hat{\theta}}_n - \underline{\theta})$ is asymptotically normal with mean zero and covariance matrix $B_{\underline{\theta}} \underline{\Sigma}_\theta B_{\underline{\theta}}^T$, where

$$\underline{\Sigma}_\theta \triangleq \begin{vmatrix} Var_{\underline{\theta}}(Z^4) & Cov_{\underline{\theta}}(Z^4, Z^6) \\ Cov_{\underline{\theta}}(Z^4, Z^6) & Var_{\underline{\theta}}(Z^6) \end{vmatrix}.$$

and

$$B_{\underline{\theta}} \triangleq \left. \begin{vmatrix} \frac{\partial f_1}{\partial \mu_4} & \frac{\partial f_1}{\partial \mu_6} \\ \frac{\partial f_2}{\partial \mu_4} & \frac{\partial f_2}{\partial \mu_6} \end{vmatrix} \right|_{\underline{\mu}_\theta}.$$

Proof. Let $\underline{\mu}_\theta \triangleq (\mu_4, \mu_6)^T$, $\underline{\hat{m}} \triangleq (\hat{m}_4, \hat{m}_6)^T$ and $\underline{X}_\nu = (Z_\nu^4, Z_\nu^6)^T$ where $Z_1, \ldots, Z_n$ is our sequence of i.i.d observations. Then $\{\underline{X}_\nu\}_{\nu=1}^\infty$ is i.i.d. with mean vector $\underline{\mu}_\theta$ and covariance matrix $\underline{\Sigma}_\theta$. (Expressions for the elements of $\underline{\Sigma}_\theta$ will be given below. We note here that all are defined and finite for $\underline{\theta} \in \Lambda$.) Since $\hat{m}_4 = \frac{1}{n}\sum_{\nu=1}^n Z_\nu^4$ and $\hat{m}_6 = \frac{1}{n}\sum_{\nu=1}^n Z_\nu^6$, we have by the multivariate Central Limit Theorem (CLT) (see [14], Thm. 5.1.8) that

$$[\sqrt{n}\,(\hat{m}_4 - m_4), \sqrt{n}\,(\hat{m}_6 - m_6)]^T \overset{D}{\to} N(\underline{0}, \underline{\Sigma}_\theta)$$

as $n \to \infty$. Now, if it can be shown that f_1 and f_2 are real-valued functions of $\underline{\mu}_\theta$, defined and continuously differentiable in a neighborhood $\omega(\underline{\mu}_\theta)$ of $\underline{\mu}_\theta$ and such that the matrix $B_{\underline{\theta}}$ is nonsingular in ω, then (see [14], Thm. 5.1.9)

$$[\sqrt{n}\,(f_1(\hat{\underline{m}}) - f_1(\underline{\mu}_\theta)),\ \sqrt{n}\,(f_2(\hat{\underline{m}}) - f_2(\underline{\mu}_\theta))]^T \xrightarrow{D} N(\underline{0},\ B_{\underline{\theta}} \underline{\Sigma}_\theta B_{\underline{\theta}}^T) \tag{3.5}$$

as $n \to \infty$, whence it follows that

$$[\sqrt{n}\,(\hat{A}_n - A),\ \sqrt{n}\,(\hat{K}_n - K)]^T \xrightarrow{D} N(\underline{0},\ B_{\underline{\theta}} \underline{\Sigma}_\theta B_{\underline{\theta}}^T) \tag{3.6}$$

as $n \to \infty$. Note that we are concerned with the validity of (3.5) only for all $\underline{\mu}_\theta$ corresponding to $\underline{\theta} \in \Lambda$. Hence, we proceed to verify the above conditions for all $\underline{\mu}_\theta$ in the open set Ω.

i) That f_1 and f_2 are real-valued functions on Ω is clear.

ii) To show that f_1 and f_2 are continuously differentiable on Ω, it suffices to show that $\frac{\partial f_i}{\partial \mu_4}$ and $\frac{\partial f_i}{\partial \mu_6}, i = 1,2$, are continuous on this set. Thus, let us consider the form of these partial derivatives: Let $\alpha \triangleq \frac{\mu_4}{2} - 1$, $\beta \triangleq \frac{\mu_6}{6} - \frac{3\mu_4}{2} + 2$. Then,

$$\frac{\partial f_1}{\partial \mu_4} = 3\left[\frac{\alpha}{\beta}\right]^3 + \frac{3}{2}\left[\frac{\alpha}{\beta}\right]^2. \tag{3.7a}$$

$$\frac{\partial f_1}{\partial \mu_6} = -\frac{1}{3}\left[\frac{\alpha}{\beta}\right]^3. \tag{3.7b}$$

$$\frac{\partial f_2}{\partial \mu_4} = \frac{3\alpha}{2\beta^2} + \frac{1}{2\beta} - 3\left[\frac{\alpha}{\beta}\right]^3 - \frac{3}{2}\left[\frac{\alpha}{\beta}\right]^2. \tag{3.7c}$$

and

$$\frac{\partial f_2}{\partial \mu_6} = -\frac{\alpha}{6\beta^2} + \frac{1}{3}\left[\frac{\alpha}{\beta}\right]^3. \tag{3.7d}$$

Note that the partial derivatives are discontinuous only on $\{\underline{\mu}_\theta \in \mathbb{R}^2 : \beta = 0\}$. But $\{\underline{\mu}_\theta \in \mathbb{R}^2 : \frac{\mu_6}{6} - \frac{3\mu_4}{2} + 2 = 0\} \cap \Omega = \phi$ by definition of Ω.

iii) To show that $B_{\underline{\theta}}$ is nonsingular for $\underline{\mu}_\theta \in \Omega$, we must show that $\det B_{\underline{\theta}} \neq 0$ for all such $\underline{\mu}_\theta$. Now,

$$\det B_{\underline{\theta}} = \frac{\partial f_1}{\partial \mu_4}\bigg|_{\underline{\mu}_{\underline{\theta}}} \frac{\partial f_2}{\partial \mu_6}\bigg|_{\underline{\mu}_{\underline{\theta}}} - \frac{\partial f_1}{\partial \mu_6}\bigg|_{\underline{\mu}_{\underline{\theta}}} \frac{\partial f_2}{\partial \mu_4}\bigg|_{\underline{\mu}_{\underline{\theta}}}.$$

Substituting the expressions given for the partial derivatives in (3.7a-d) and simplifying, we have that

$$\det B_{\underline{\theta}} = \frac{-\alpha^3}{12\beta^4}\bigg|_{\underline{\mu}_{\underline{\theta}}}$$

which is nonzero iff $\alpha\big|_{\underline{\mu}_{\underline{\theta}}} \neq 0$. Thus, $B_{\underline{\theta}}$ is nonsingular for all $\underline{\mu}_{\underline{\theta}}$ with $\mu_4 \neq 2$. But $\Omega \subset \{\underline{\mu}_{\underline{\theta}} \in R^2 : \mu_4 \neq 2\}$ by definition of Ω. Thus, $B_{\underline{\theta}}$ is nonsingular for all $\underline{\mu}_{\underline{\theta}} \in \Omega$.

The verification of the conditions is now complete. Consequently, (3.6) holds for all $\underline{\theta} \in \Lambda$, i.e., $\hat{\underline{\theta}}_n$ is asymptotically normal $\left(\underline{\theta}, \frac{1}{n} B_{\underline{\theta}} \ddagger_{\underline{\theta}} B_{\underline{\theta}}^T\right)$ for $\underline{\theta} \in \Lambda$. □

3.2.3. Asymptotic performance of $\hat{A}_n$ and $\hat{K}_n$

The performance of the normalized *MM* estimates $\hat{A}_n/A$ and $\hat{K}_n/K$ will now be considered. Expressions for the asymptotic variances of these estimates will be obtained and computed for a broad range of parameter values in Λ. Note that the normalizations are necessary in order that a meaningful comparison of the computed variances be made since the parameters take on widely varying values. Let $C_1 \triangleq [\frac{1}{A}, 0]$ and $C_2 \triangleq [0, \frac{1}{K}]$. Then, from (3.6), we have that

$$C_1[\sqrt{n}\,(\hat{A}_n - A)\,, \sqrt{n}\,(\hat{K}_n - K)]^T \xrightarrow{D} N(0\,, C_1 B_{\underline{\theta}} \ddagger_{\underline{\theta}} B_{\underline{\theta}}^T C_1^T)$$

and

$$C_2[\sqrt{n}\,(\hat{A}_n - A)\,, \sqrt{n}\,(\hat{K}_n - K)]^T \xrightarrow{D} N(0\,, C_2 B_{\underline{\theta}} \ddagger_{\underline{\theta}} B_{\underline{\theta}}^T C_2^T)\,,$$

i.e.,

$$\sqrt{n}\left(\frac{\hat{A}_n - A}{A}\right) \xrightarrow{D} N(0\,, C_1 B_{\underline{\theta}} \ddagger_{\underline{\theta}} B_{\underline{\theta}}^T C_1^T) \tag{3.8a}$$

and

$$\sqrt{n}\left(\frac{\hat{K}_n - K}{K}\right) \xrightarrow{D} N(0, C_2 B_{\underline{\theta}} \Sigma_{\underline{\theta}} B_{\underline{\theta}}^T C_2^T). \tag{3.8b}$$

Let ϵ_A and ϵ_K denote the asymptotic variances of $\hat{A}_n/A$ and $\hat{K}_n/K$, respectively. Then,

$$\epsilon_A = C_1 B_{\underline{\theta}} \Sigma_{\underline{\theta}} B_{\underline{\theta}}^T C_1^T \tag{3.9a}$$

and

$$\epsilon_K = C_2 B_{\underline{\theta}} \Sigma_{\underline{\theta}} B_{\underline{\theta}}^T C_2^T . \tag{3.9b}$$

We seek expressions for these variances in terms of the parameters A and K. To this end, we begin by substituting the corresponding matrices for $B_{\underline{\theta}}, \Sigma_{\underline{\theta}}, C_1$, and C_2. From (3.9a) we have that

$$\epsilon_A = \left[\frac{1}{A} \quad 0\right] \begin{bmatrix} \frac{\partial f_1}{\partial \mu_4} & \frac{\partial f_1}{\partial \mu_6} \\ \frac{\partial f_2}{\partial \mu_4} & \frac{\partial f_2}{\partial \mu_6} \end{bmatrix}\Bigg|_{\underline{\mu}_{\underline{\theta}}} \begin{bmatrix} Var_{\underline{\theta}}(Z^4) & Cov_{\underline{\theta}}(Z^4, Z^6) \\ Cov_{\underline{\theta}}(Z^4, Z^6) & Var_{\underline{\theta}}(Z^6) \end{bmatrix} \begin{bmatrix} \frac{\partial f_1}{\partial \mu_4} & \frac{\partial f_1}{\partial \mu_6} \\ \frac{\partial f_2}{\partial \mu_4} & \frac{\partial f_2}{\partial \mu_6} \end{bmatrix}\Bigg|_{\underline{\mu}_{\underline{\theta}}} \begin{bmatrix} \frac{1}{A} \\ 0 \end{bmatrix} \tag{3.10a}$$

$$= \frac{1}{A^2}\left[\left(\frac{\partial f_1}{\partial \mu_4}\right)^2\Big|_{\underline{\mu}_{\underline{\theta}}} Var_{\underline{\theta}}(Z^4) + 2\, Cov_{\underline{\theta}}(Z^4, Z^6)\, \frac{\partial f_1}{\partial \mu_4}\Big|_{\underline{\mu}_{\underline{\theta}}} \frac{\partial f_1}{\partial \mu_6}\Big|_{\underline{\mu}_{\underline{\theta}}} + \left(\frac{\partial f_1}{\partial \mu_6}\right)^2\Big|_{\underline{\mu}_{\underline{\theta}}} Var_{\underline{\theta}}(Z^6)\right].$$

Similarly, (3.9b) yields

$$\epsilon_K = \left[0 \quad \frac{1}{K}\right] \begin{bmatrix} \frac{\partial f_1}{\partial \mu_4} & \frac{\partial f_1}{\partial \mu_6} \\ \frac{\partial f_2}{\partial \mu_4} & \frac{\partial f_2}{\partial \mu_6} \end{bmatrix}\Bigg|_{\underline{\mu}_{\underline{\theta}}} \begin{bmatrix} Var_{\underline{\theta}}(Z^4) & Cov_{\underline{\theta}}(Z^4, Z^6) \\ Cov_{\underline{\theta}}(Z^4, Z^6) & Var_{\underline{\theta}}(Z^6) \end{bmatrix} \begin{bmatrix} \frac{\partial f_1}{\partial \mu_4} & \frac{\partial f_1}{\partial \mu_6} \\ \frac{\partial f_2}{\partial \mu_4} & \frac{\partial f_2}{\partial \mu_6} \end{bmatrix}\Bigg|_{\underline{\mu}_{\underline{\theta}}} \begin{bmatrix} 0 \\ \frac{1}{K} \end{bmatrix} \tag{3.10b}$$

$$= \frac{1}{K^2}\left[\left(\frac{\partial f_2}{\partial \mu_4}\right)^2\Big|_{\underline{\mu}_{\underline{\theta}}} Var_{\underline{\theta}}(Z^4) + 2\, Cov_{\underline{\theta}}(Z^4, Z^6)\, \frac{\partial f_2}{\partial \mu_4}\Big|_{\underline{\mu}_{\underline{\theta}}} \frac{\partial f_2}{\partial \mu_6}\Big|_{\underline{\mu}_{\underline{\theta}}} + \left(\frac{\partial f_2}{\partial \mu_6}\right)^2\Big|_{\underline{\mu}_{\underline{\theta}}} Var_{\underline{\theta}}(Z^6)\right].$$

Expressions for the partial derivatives $\frac{\partial f_i}{\partial \mu_4}, \frac{\partial f_i}{\partial \mu_6}, i = 1, 2$, have been given in (3.7a-d) in terms of μ_4 and μ_6, which in turn can be expressed in terms of A and K via (3.1). Expressions for the eighth, tenth, and twelfth moments of $w(z)$ can be obtained in a manner completely analogous to that used in obtaining μ_4 and μ_6. Thus, $Var_{\underline{\theta}}(Z^4)$, $Var_{\underline{\theta}}(Z^6)$, and $Cov_{\underline{\theta}}(Z^4, Z^6)$ are readily expressed in terms of A and K. The final results are stated here:

$$\left.\frac{\partial f_1}{\partial \mu_4}\right|_{\underline{\mu_\theta}} = 3(A+K)^3 + \frac{3}{2}(A+K)^2. \tag{3.11a}$$

$$\left.\frac{\partial f_1}{\partial \mu_6}\right|_{\underline{\mu_\theta}} = -\frac{1}{3}(A+K)^3. \tag{3.11b}$$

$$\left.\frac{\partial f_2}{\partial \mu_4}\right|_{\underline{\mu_\theta}} = \frac{3}{2}\frac{(A+K)^4}{A} + \frac{(1-6A)}{2A}(A+K)^3 - \frac{3}{2}(A+K)^2. \tag{3.11c}$$

$$\left.\frac{\partial f_2}{\partial \mu_6}\right|_{\underline{\mu_\theta}} = -\frac{1}{6}\frac{(A+K)^4}{A} + \frac{1}{3}(A+K)^3. \tag{3.11d}$$

$$Var_{\underline{\theta}}(Z^4) = 20 + \frac{136A}{(A+K)^2} + \frac{(24A - 4A^2)}{(A+K)^4} + \frac{24A(7A+4K)}{(A+K)^4}. \tag{3.11e}$$

$$\begin{aligned} Var_{\underline{\theta}}(Z^6) = \frac{720}{(A+K)^6}[&(A^6 + 15A^5 + 65A^4 + 90A^3 + 31A^2 + A) \\ &+ (6A^5K + 60A^4K + 150A^3K + 90A^2K + 6AK) \\ &+ (15A^4K^2 + 90A^3K^2 + 105A^2K^2 + 15AK^2) \\ &+ (20A^3K^3 + 60A^2K^3 + 20AK^3) \\ &+ (15A^2K^4 + 15AK^4) + (6AK^5 + K^6)] \\ &- \left[\frac{6A}{(A+K)^3} + \frac{18A}{(A+K)^2} + 6\right]^2. \end{aligned} \tag{3.11f}$$

$$
\begin{aligned}
Cov_{\underline{\theta}}(Z^4, Z^6) = \frac{120}{(A+K)^5}\,[&(A^5 + 10A^4 + 25A^3 + 15A^2 + A) \\
&+ (5A^4K + 30A^3K + 35A^2K + 5AK) \\
&+ (10A^3K^2 + 30A^2K^2 + 10AK^2) \\
&+ (10A^2K^3 + 10AK^3) + (5AK^4) + K^5] \\
&- \left[\frac{2A}{(A+K)^2} + 2\right]\left[\frac{6A}{(A+K)^3} + \frac{18A}{(A+K)^2} + 6\right].
\end{aligned}
\tag{3.11g}
$$

The asymptotic variances ϵ_A and ϵ_K have been computed for $\{(A, K)^T \in \Lambda \mid \log A \in \{0,-1,-2\}$ and $\log K \in \{-2,-3,-4,-5,-6\}\}$ using (3.10a), (3.10b), and (3.11a-g) (see Tables 3.1 and 3.2). In addition, values for the asymptotic mean-square norm relative error (MSNRE) ϵ_r, $\epsilon_r \triangleq \epsilon_A + \epsilon_K$, are given in Table 3.3. Note that the primary contribution to ϵ_r is from ϵ_K. In fact, for $K << A$, the contribution to ϵ_r from ϵ_A is negligible and it is for this case that ϵ_r becomes extremely large. Hence, consider ϵ_K.

For A fixed, $\epsilon_K(A, K)$ increases as K decreases and takes on very large values when $K << A$. For K fixed, $\epsilon_K(A, K)$ increases as A increases and again takes on very large values when $A >> K$. Note specifically that when $A = 1$ the asymptotic variance ϵ_K becomes extremely large. These observations are easily explained when one considers the form of μ_4 and μ_6 (see (3.1)).

Table 3.1. ASYMPTOTIC VARIANCE OF $\hat{A}_n / A$ (ϵ_A)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	2.550337×10^{3}	2.468429×10^{3}	2.460431×10^{3}	2.459633×10^{3}	2.459554×10^{3}
10^{-1}	1.472224×10^{3}	1.448790×10^{3}	1.446479×10^{3}	1.446248×10^{3}	1.446225×10^{3}
1	4.792669×10^{3}	4.750635×10^{3}	4.746463×10^{3}	4.746046×10^{3}	4.746005×10^{3}

Table 3.2. ASYMPTOTIC VARIANCE OF $\hat{K}_n/K$ (ϵ_K)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	6.663680×10^{3}	5.264450×10^{4}	5.963216×10^{6}	6.041263×10^{8}	6.049149×10^{10}
10^{-1}	2.449795×10^{4}	2.951534×10^{6}	3.005101×10^{8}	3.010491×10^{10}	3.011030×10^{12}
1	1.073910×10^{7}	1.083890×10^{9}	1.084889×10^{11}	1.084989×10^{13}	1.084999×10^{15}

Table 3.3. ASYMPTOTIC MSNRE FOR *MM* ESTIMATOR (ϵ_r)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	3.216705×10^{3}	5.511293×10^{4}	5.965677×10^{6}	6.041288×10^{8}	6.049150×10^{10}
10^{-1}	2.597017×10^{4}	2.952983×10^{6}	3.005115×10^{8}	3.010491×10^{10}	3.011030×10^{12}
1	1.074389×10^{7}	1.083895×10^{9}	1.084889×10^{11}	1.084989×10^{13}	1.084999×10^{15}

Note that μ_4 and μ_6 depend inversely on powers of $(A+K)$, and K appears only in this way. Thus, when $K << A$, μ_4 and μ_6 are insensitive to changes in K. This is evident in Tables 3.4 and 3.5. For A fixed, it becomes increasingly difficult to resolve $\mu_4\,[\mu_6]$ as K decreases, particularly for $K << A$, hence the increasing values for ϵ_K. The second observation stated above is accounted for by the fact that, for K fixed, increasing A makes K relatively small in comparison to A, which in turn makes the moments and the estimator less sensitive to K. For $A = 1$, $\mu_4\,[\mu_6]$ are nearly the same for all K, since $K << A$ for all values of K under consideration. In fact, $\mu_4\,[\mu_6]$ for $K = 10^{-5}$ equals $\mu_4\,[\mu_6]$ for $K = 10^{-6}$ up to five significant digits. Thus, in order that a reasonable estimate of K be obtained, the number of samples used must be large enough so that a resolution of the moments up to five decimal places is achieved. (It should be noted that this insensitivity of the moments to changes in the value of K when $K << A$ is also evident in the higher-ordered moments.)

Consider the worst-case error for ϵ_r, which occurs for $A = 1$, $K = 10^{-6}$. A comparison of Tables 3.2 and 3.3 indicates that this quantity is essentially the asymptotic variance ϵ_K. Given that $A = 1$, $K = 10^{-6}$, suppose we want the probability that $(\hat{A}_n/A\,,\,\hat{K}_n/K)$ lies within a circle of radius 0.1 with center at (1,1) to be 0.9. Let

$$C \triangleq \begin{bmatrix} \frac{1}{A} & 0 \\ 0 & \frac{1}{K} \end{bmatrix}. \tag{3.12}$$

Since, from (3.6), we have that

$$\left[\sqrt{n}\left[\frac{\hat{A}_n - A}{A}\right], \sqrt{n}\left[\frac{\hat{K}_n - K}{K}\right]\right]^T \xrightarrow{D} N(\underline{0}, CB_{\underline{\theta}} \sharp_{\underline{\theta}} B_{\underline{\theta}}^T C^T),$$

a straightforward calculation shows that

$$P\left[\left[\frac{(\hat{A}_n - A)^2}{A^2} + \frac{(\hat{K}_n - K)^2}{K^2}\right] \leq n(0.1)^2\right] \sim 0.9$$

Table 3.4. FOURTH-ORDER POPULATION MOMENT (μ_4)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	5.200000×10^{1}	1.672893×10^{2}	1.980592×10^{2}	2.016006×10^{2}	2.019600×10^{2}
10^{-1}	1.852893×10^{1}	2.160592×10^{1}	2.196006×10^{1}	2.199600×10^{1}	2.199960×10^{1}
1	3.960592	3.996006	3.999600	3.999960	3.999996

Table 3.5. SIXTH-ORDER POPULATION MOMENT (μ_6)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	7.9560000×10^{3}	4.6572491×10^{4}	6.0005942×10^{4}	6.1622765×10^{4}	6.1787644×10^{4}
10^{-1}	6.0554921×10^{2}	7.6480738×10^{2}	7.8384413×10^{2}	7.8578404×10^{2}	7.8597840×10^{2}
1	2.9468870×10^{1}	2.9946090×10^{1}	2.9994601×10^{1}	2.9999460×10^{1}	2.9999946×10^{1}

requires $n(0.1)^2$ to be approximately 2.94×10^{15}, or n to be approximately 2.94×10^{17}. This is an unrealistically large sample size for most applications and thus reveals the moments estimator to be highly inefficient in this sense. The potential poor performance of the moments estimator has been verified via simulation for a wide range of parameter values.[1] Whether this inefficiency is inherent in the Class A model or is a property of the method-of-moments estimator will be determined in Section 3.3.

3.3. Asymptotically Efficient Estimation

We have seen in the previous section that the method of moments yields a strongly consistent and asymptotically normal estimator of the parameters of the Class A model. However, the *MM* estimator has a serious shortcoming. Specifically, for values of $K << A$, the asymptotic MSNRE is astoundingly large. This is due to the fact that the *MM* estimator is highly insensitive to changes in the parameter K when $K << A$. A natural question that arises is whether this insensitivity is a property of the *MM* estimator or is an inherent feature of the Class A model, i.e., is it possible to improve on the performance of the *MM* estimator? To answer this question, let us examine the Cramer-Rao Lower Bound (CRLB).

3.3.1. The Cramer-Rao Lower Bound

We begin with the following assertion. Let $\underline{\theta}_n^* = (A_n^*, K_n^*)^T$ denote an estimator of $\underline{\theta} = (A, K)^T$ based on n i.i.d. observations. Under regularity conditions on the class of estimators $\underline{\theta}_n^*$ under consideration [15], it may be asserted that if $\underline{\theta}_n^*$ is asymptotically normal with mean vector $\underline{\theta}$ and covariance matrix $n^{-1}\mathbf{\Psi}_{\underline{\theta}}'$, then the condition

$$C\left[\mathbf{\Psi}_{\underline{\theta}}' - [I(\underline{\theta})]^{-1}\right] C^T \geq 0 \tag{3.13}$$

must hold, where C is the nonsingular, symmetric matrix defined in (3.12) and $I(\underline{\theta})$ denotes Fisher's information matrix:

[1] This inefficiency is also corroborated for a restricted range of parameter values by experimental results presented in [12].

$$I(\underline{\theta}) = E_{\underline{\theta}}\left[\begin{array}{cc} \left[\frac{\partial}{\partial A}\log w(Z)\right]^2 & \left[\frac{\partial}{\partial A}\log w(Z)\right]\left[\frac{\partial}{\partial K}\log w(Z)\right] \\ \left[\frac{\partial}{\partial A}\log w(Z)\right]\left[\frac{\partial}{\partial K}\log w(Z)\right] & \left[\frac{\partial}{\partial K}\log w(Z)\right]^2 \end{array}\right]. \quad (3.14)$$

It follows from (3.13) that

$$tr\left[C\left[\Sigma'_{\underline{\theta}} - [I(\underline{\theta})]^{-1}\right]C^T\right] \geq 0$$

which, in turn, yields the following lower bound on the asymptotic MSNRE of $\underline{\theta}_n^*$:

$$tr\left[C^2 \Sigma'_{\underline{\theta}}\right] \geq tr\left[C^2 [I(\underline{\theta})]^{-1}\right]. \quad (3.15)$$

Now, if $\underline{\theta}_n^*$ is an asymptotically efficient (AE) estimator of $\underline{\theta}$, then $\Sigma'_{\underline{\theta}} = [I(\underline{\theta})]^{-1}$ and (3.15) holds with equality. Thus, let ϵ_r' denote the asymptotic MSNRE for an AE estimator. Then,

$$\epsilon_r' = tr\left[C^2 [I(\underline{\theta})]^{-1}\right]. \quad (3.16)$$

We will now consider the contribution to ϵ_r' due to estimating each parameter in the two-parameter estimation scheme. Let ϵ_A' denote the asymptotic relative variance due to estimating A via an AE estimator and let ϵ_K' denote the corresponding quantity for the parameter K. Since $\Sigma'_{\underline{\theta}} = [I(\underline{\theta})]^{-1}$ for an AE estimator, it follows that

$$\epsilon_A' = V_1\left(C^2 \Sigma'_{\underline{\theta}}\right) V_1^T = V_1\left[C^2 [I(\underline{\theta})]^{-1}\right] V_1^T \quad (3.17a)$$

and

$$\epsilon_K' = V_2\left(C^2 \Sigma'_{\underline{\theta}}\right) V_2^T = V_2\left[C^2 [I(\underline{\theta})]^{-1}\right] V_2^T \quad (3.17b)$$

where $V_1 \triangleq [1,0]$ and $V_2 \triangleq [0,1]$. Thus, (3.17a) and (3.17b) imply that

$$\epsilon_A' = [I(\underline{\theta})]_{11}^{-1} / A^2 \quad (3.18a)$$

and

$$\epsilon_K' = [I(\underline{\theta})]_{22}^{-1} / K^2 \quad (3.18b)$$

where $[I(\underline{\theta})]_{jj}^{-1}$ denotes the jj-th element of the inverse of the matrix $I(\underline{\theta})$.

Examining the theoretical lower bounds on the asymptotic MSNRE and asymptotic relative variances as given by (3.16) and (3.18a,b), we can determine whether an improvement in performance over the method-of-moments estimator is possible. Moreover, the degree of improvement possible can be ascertained by a comparison of ϵ_A and ϵ_A', ϵ_K and ϵ_K', ϵ_r and ϵ_r'. The quantities ϵ_A', ϵ_K', and ϵ_r' have been computed and tabulated for $\{(A,K)^T \in \Lambda \mid \log A \in \{0,-1,-2\}$ and $\log K \in \{-2,-3,-4,-5,-6\}\}$ (see Tables 3.6 - 3.8). A discussion of the results follows.

We note by a comparison of ϵ_A and ϵ_A' (Tables 3.1 and 3.6), ϵ_K and ϵ_K' (Tables 3.2, 3.7), and ϵ_r and ϵ_r' (Tables 3.3, 3.8) that the values for the asymptotic variances and asymptotic MSNRE dictated by an *AE* estimator are significantly lower than the corresponding values for the *MM* estimator for all parameter pairs under consideration. Moreover, whereas for the *MM* estimator, the primary contribution to ϵ_r is from ϵ_K, for the *AE* estimator, neither ϵ_A' nor ϵ_K' dominates.

Consider the asymptotic MSNRE. Since $\epsilon_r \neq \epsilon_r'$, the moments estimator is not only inefficient in the sense described in Section 3.2.3, but it is not asymptotically efficient. Moreover, the apparent improvement in performance yielded by an *AE* estimator is tremendous, particularly for values of $K << A$, i.e., in the region where the *MM* estimator is most inefficient. Note also

Table 3.6. ASYMPTOTIC RELATIVE VARIANCE DUE TO ESTIMATING A VIA AN AE ESTIMATOR (ϵ_A')

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	6.2440×10^{1}	5.1972×10^{1}	5.0865×10^{1}	5.0735×10^{1}	5.0720×10^{1}
10^{-1}	6.7196	5.7920	5.6860	5.6732	5.6716
1	1.5909	1.3509	1.3224	1.3188	1.3184

Table 3.7. ASYMPTOTIC RELATIVE VARIANCE DUE TO ESTIMATING K VIA AN AE ESTIMATOR (ϵ_K')

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	6.2945×10^{1}	5.2961×10^{1}	5.1869×10^{1}	5.1744×10^{1}	5.1730×10^{1}
10^{-1}	7.5054	6.8492	6.7839	6.7773	6.7767
1	4.8083	4.2365	4.0854	4.0470	4.0385

Table 3.8. ASYMPTOTIC MSNRE FOR AN AE ESTIMATOR (ϵ_r')

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	1.2539×10^{2}	1.0493×10^{2}	1.0273×10^{2}	1.0248×10^{2}	1.0245×10^{2}
10^{-1}	1.4225×10^{1}	1.2641×10^{1}	1.2470×10^{1}	1.2451×10^{1}	1.2448×10^{1}
1	6.3991	5.5874	5.4078	5.3658	5.3569

that the maximum and minimum values for ϵ_r are 1.0850×10^{15} and 3.2167×10^{3}, respectively, as opposed to a maximum and minimum value of 1.2539×10^{2} and 5.3569, respectively, for ϵ_r'. Furthermore, ϵ_r achieves its maximum value at the same point at which ϵ_r' achieves its minimum value, namely at $A = 1$, $K = 10^{-6}$. In fact, the improvement at this point is on the order of 10^{15}. Thus, there is a dramatic improvement in performance in the area where it is most needed.

3.3.2. Likelihood-based estimators

The question posed at the beginning of the section has now been answered. In particular, the high insensitivity of the *MM* estimator to changes in the parameter K is a feature of the estimator. It is not a feature of the model. Moreover, we can expect a significant improvement in performance given that an asymptotically efficient estimator can be found. Naturally, the search for such an estimator begins with maximum likelihood.

Unfortunately, maximum-likelihood estimation for the Class A model turns out to be unwieldy. Numerical experimentation with the likelihood equation (LE) reveals that the LE cannot be readily solved for the maximum-likelihood estimator. In particular, the likelihood function has steep gradients, the LE has multiple roots for finite sample sizes, etc. Moreover, closer examination of the LE in the context of estimating a single parameter (fixing K and considering the problem of estimating A only) reveals that the LE does not have a unique root asymptotically for all $(A, K)^T \in \Lambda$. For example, for $(A, K) = (10^{-2}, 10^{-4})$, the LE has roots at $A = 10^{-2}$ and at $A = 0.308312$ asymptotically. This multiplicity of roots implies the existence of inconsistent sequences of roots to the LE when the problem of estimating A only with K known is considered and a similar phenomenon may be the source of difficulty in the two-parameter situation.

Thus, we have a consistent estimator (the *MM* estimator) which is highly inefficient and we have a potentially efficient estimator which is computationally difficult (and possibly inconsistent). However, the consistency of the moments estimator can be combined with the potential efficiency of likelihood-based estimation to yield a consistent and asymptotically efficient estimator of the parameters of the Class A model. This is done via a standard procedure [14] whereby Newton's

root-finding method on the LE is initiated with a $\sqrt{n}$ -consistent estimator and, in so doing, both the consistency of the consistent estimator and the efficiency of likelihood-based estimation are retained. A natural initial estimator to use in such a scheme for the Class A estimation problem is the *MM* estimator, which was shown to be consistent in Section 3.2.2. That the *MM* estimator is, in fact, $\sqrt{n}$ -consistent follows from its asymptotic normality. Thus, provided certain regularity conditions are satisfied by the densities w (see [14]), Newton iteration on the LE initiated with the *MM* estimate will be consistent and efficient.

As a practical matter, this type of estimator will work only if the initial estimate (the *MM* estimate) is reasonably close to the consistent root of the LE. Unfortunately, simulation studies using several thousand samples indicate that this closeness is not achieved for most values of A and K of interest. In fact, the *MM* estimator frequently produces invalid (e.g., negative) initial estimates for these sample sizes. Thus, even though this procedure performs well asymptotically, there are limitations as to how well it can perform for a moderate number of samples, the source of this poor performance being the high inefficiency of the moments estimator.

However, as we shall see in the following section, by starting with an initial estimator with enhanced efficiency and iterating around it until Newton's method converges, we can obtain a better estimator. Hence, we turn now to an estimator which, despite the fact that it lacks the asymptotic optimality properties of the *Moment /Likelihood* procedure described above, appears to be much more efficient than this procedure for moderate sample sizes.

3.4. Threshold-Comparison/Likelihood Estimator

In this section, we consider a practical estimator based on the idea of using likelihood iteration initiated with a physically motivated, but nonoptimal, estimator. In particular, we consider a procedure which uses as its initial estimator a scheme motivated by Middleton's approximate empirical procedure. This latter estimator, described in [10],[11], is a graphical procedure based on features of the finite sample size distribution. The threshold comparison estimates are motivated in the following way: Note that we can decompose the Class A envelope pdf w (2.3) into two components.

The first of these corresponds to the $m = 0$ term and is attributable to the Gaussian background noise component; the second corresponds to all terms indexed by $m \geq 1$ and is attributable primarily to the impulsive noise component. As it happens, these two terms can be clearly distinguished for most A and K values of practical interest if we consider the envelope distribution function $P(Z > z_o)$. Typical envelope distributions are shown in Figs. 3.1 and 3.2. We note from the form of this function that it divides the z_o-axis into three regions: the first corresponds to smaller values of z_o, in which the Gaussian background noise component dominates; the second corresponds to larger values of z_o, wherein the impulsive noise component dominates; and the third corresponds to intermediate values of z_o, for which $P(Z > z_o)$ is virtually constant. The portion of the distribution corresponding to these intermediate values of z_o will be termed the "null region." We note that for values of $A \geq 10^{-1}$ and $\Gamma \leq 10^{-3}$, the departure from the straight-line, Gaussian (actually, Rayleigh, since the envelope distribution is being considered) portion of the distribution is abrupt and extensive. In this case, the null region is clearly identifiable. Thus, we can set a threshold α^* at any value of the abscissa (z_o) corresponding to a point in this null region (see Fig. 3.1) so that, with high probability, samples falling below α^* can be attributed to the Gaussian background component and samples exceeding α^* can be attributed to the impulsive component. For values of $A < 10^{-1}$ or $\Gamma > 10^{-3}$, the departure from the straight-line, Rayleigh portion is gradual and/or less extensive, in which case such an α^* can be chosen to be the abscissa value (z_o) at which the distribution begins to depart, observably, from its straight-line (Rayleigh) behavior (see Fig. 3.2).

The above feature of the envelope distribution can be used to obtain estimates of the parameters A and K. In particular, from an i.i.d. sequence, $Z_1, \ldots, Z_n$, of Class A envelope samples, we can determine an estimate of the threshold α^* from the sample distribution function. We then divide the observations $Z_1, \ldots, Z_n$ into a set $\{\overline{Z}_1, \overline{Z}_2, \ldots, \overline{Z}_{n_I}\}$ consisting of those lying above α^* and a set $\{\underline{Z}_1, \underline{Z}_2, \ldots, \underline{Z}_{n_B}\}$ consisting of those falling below α^*. Since A is approximately the expected fraction of impulses in a random sample, it can be estimated as

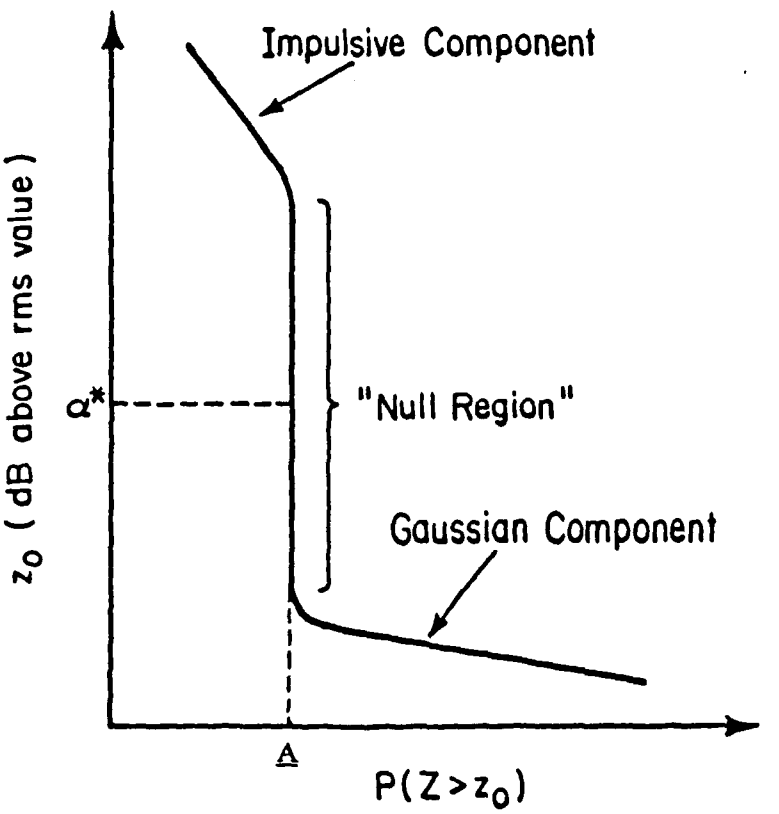


Fig. 3.1. Typical envelope distribution for $A \geq 10^{-1}$ and $\Gamma \leq 10^{-3}$. Note the abrupt and extensive departure from the straight-line (Rayleigh) portion of the curve. (The distribution is plotted on linear (for z_o) by $0.5 \log_{10} (-\log_e P)$ coordinates.)

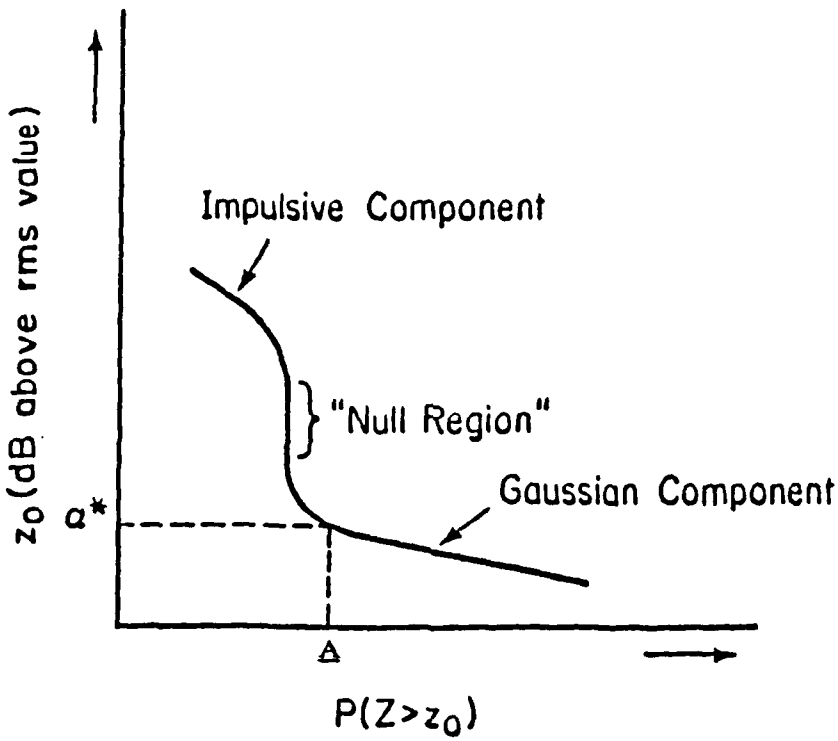


Fig. 3.2. Typical envelope distribution for $A < 10^{-1}$ and $\Gamma > 10^{-3}$. Note the gradual and brief departure from the straight-line (Rayleigh) portion of the curve. (The distribution is plotted on linear (for z_o) by $0.5 \log_{10} (-\log_e P)$ coordinates.)

$$\underline{A}_n = n_I / n \ . \tag{3.19}$$

Similarly, since K is approximately the average energy in a background sample relative to that in an impulsive sample, it can be estimated as

$$\underline{K}_n = \left[\frac{1}{n_B} \sum_{i=1}^{n_B} \underline{Z}_i^2 \right] \Big/ \left[\frac{1}{n_I} \sum_{i=1}^{n_I} \overline{Z}_i^2 \right] . \tag{3.20}$$

These estimates will be referred to as the *threshold comparison* estimates of A and K. Note that for values of $A \geqslant 10^{-1}$ and $\Gamma \leqslant 10^{-3}$, $\underline{A}_n$ is simply the probability value corresponding to the point where the sharp rise in the sample distribution begins, and for values of $A < 10^{-1}$ or $\Gamma > 10^{-3}$, $\underline{A}_n$ is simply the value of P where the sample distribution begins to depart, observably, from its straight-line (Rayleigh) behavior. (The threshold comparison estimator is based on Middleton's approximate empirical procedure, which is described in [10],[11]. We find that the threshold comparison estimator is computationally easier to implement and leads to more accurate results than the approximate empirical procedure.)

In order to assess the performance of the above estimator, an extensive simulation study was performed wherein the normalized sample MSNRE ($\underline{\Delta}\, n$ (sample MSNRE)) for the threshold comparison estimator was computed for values of A and K throughout their practical ranges. For each parameter pair, the computation was made using 100 data sets, each containing 3000 observations randomly generated from the corresponding Class A envelope pdf. The values for the normalized MSNRE are tabulated in Table 3.9. Note that the threshold comparison estimator performs very well. In particular, from a comparison of Tables 3.3 and 3.9, one can infer that the threshold comparison estimator performs significantly better than the *MM* estimator for moderate sample sizes. Thus, the threshold comparison estimator is a good candidate for use as an initial estimator in a scheme whereby a solution to the likelihood equation is determined iteratively. The iterative determination of a solution to the LE after initiating with the threshold comparison estimator will be referred to as the *Threshold - Comparison /Likelihood* Estimator. Let us examine the performance of this estimator.

Table 3.9. NORMALIZED MSNRE FOR THRESHOLD COMPARISON ESTIMATOR (3000 SAMPLES, 100 RUNS)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	2.3885×10^{2}	2.1233×10^{2}	2.1253×10^{2}	2.1332×10^{2}	2.1310×10^{2}
10^{-1}	1.0812×10^{2}	3.7563×10^{1}	3.3097×10^{1}	3.2785×10^{1}	3.2810×10^{1}
1	7.7198×10^{2}	7.4617×10^{2}	8.1020×10^{2}	8.1468×10^{2}	8.1504×10^{2}

Table 3.10. NORMALIZED MSNRE FOR THRESHOLD-COMPARISON/LIKELIHOOD ESTIMATOR (3000 SAMPLES, 100 RUNS)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	1.1841×10^{2}	9.2304×10^{1}	8.7421×10^{1}	8.8302×10^{1}	8.8320×10^{1}
10^{-1}	1.2818×10^{1}	1.1732×10^{1}	1.1723×10^{1}	1.1747×10^{1}	1.1736×10^{1}
1	6.5348	5.1473	5.3332	5.2676	5.2281

As before, a simulation study was performed wherein the normalized MSNRE for the Threshold-Comparison/Likelihood estimator was computed for a range of A and K values. Again, the computation for each parameter pair was made using 100 data sets, each containing 3000 observations. The results are tabulated in Table 3.10. We noted above that the threshold comparison estimator performs very well. However, a comparison of Tables 3.9 and 3.10 reveals that the Threshold-Comparison/Likelihood estimator performs even better. In particular, for $A = 1$, there is a reduction in the normalized MSNRE on the order of 10^2 for all values of K under consideration. Moreover, a comparison of Tables 3.8 and 3.10 indicates that the normalized MSNRE for the Threshold-Comparison/Likelihood estimator is very close to the Cramer-Rao Lower Bound for all values of A and K under consideration.

In addition to computation of the normalized MSNRE, the relative biases for the threshold comparison and Threshold-Comparison/Likelihood estimator were computed (3000 samples, 100 runs) (see Tables 3.11, 3.12). Note that, in addition to lowering the MSNRE, the likelihood iteration step also serves as a bias reduction technique for most values of A and K. In particular, a substantial reduction in the magnitude of the relative bias is observed for $A = 1$ and $A = 10^{-1}$ for all values of K under consideration.

Table 3.11. RELATIVE BIAS FOR THRESHOLD COMPARISON ESTIMATOR (3000 SAMPLES, 100 RUNS)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	-2.0086×10^{-1}	-3.4793×10^{-2}	-9.5308×10^{-3}	-1.3236×10^{-2}	-1.2456×10^{-2}
10^{-1}	-2.3777×10^{-1}	-1.0400×10^{-1}	-8.8897×10^{-2}	-8.7665×10^{-2}	-8.7864×10^{-2}
1	-7.0614×10^{-1}	-6.9930×10^{-1}	-7.3395×10^{-1}	-7.3603×10^{-1}	-7.3620×10^{-1}

Table 3.12. RELATIVE BIAS FOR THRESHOLD-COMPARISON/LIKELIHOOD ESTIMATOR (3000 SAMPLES, 100 RUNS)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	-3.9698×10^{-2}	-4.3362×10^{-2}	-4.1480×10^{-2}	-4.0226×10^{-2}	-4.0494×10^{-2}
10^{-1}	7.7754×10^{-3}	8.0087×10^{-3}	8.1027×10^{-3}	8.0989×10^{-3}	7.9131×10^{-3}
1	1.1470×10^{-3}	2.0072×10^{-3}	1.4641×10^{-3}	1.4592×10^{-3}	1.1912×10^{-3}

In summary of the above, the Threshold-Comparison/Likelihood estimator has many desirable features: (i) it performs very well from a practical viewpoint (substantially better than MM-based estimators); (ii) the normalized MSNRE for the estimator is very close to the CRLB;

and (iii) it serves as a bias reduction technique. Unfortunately, the Threshold-Comparison/Likelihood estimator apparently lacks the asymptotic optimality properties of the Moment/Likelihood procedure. In fact, it appears from examination of the properties of the population distribution function that this estimator is asymptotically biased and inconsistent. However, as we have seen, it works quite well for moderate sample sizes, and thus is an attractive estimator for use in applications.

3.5. Conclusions

In this chapter, we have proposed and investigated several batch estimators for the parameters of the Middleton Class A noise model in its strictly canonical form. These estimators include the method-of-moments estimator, which is computationally attractive but is unattractive in terms of performance; likelihood-based estimators, which are potentially efficient but which have undesirable computational properties; and the Moment/Likelihood estimator which, in its asymptotic performance, combines the desirable features of these two approaches. In response to the poor moderate-sample-size performance of the Moment/Likelihood estimator observed via simulation, a similar estimator that initiates likelihood iteration with the threshold comparison estimator has also been considered. Analysis of the moderate-sample-size performance of this scheme shows it to be an effective estimator for practical use (i.e., sample sizes on the order of 10^3).

4. RECURSIVE IDENITIFCATION

4.1. Introduction

In Chapter 3, we addressed the problem of basic batch estimation of the Class A parameters. In this chapter, we will develop a *recursive* algorithm for on-line identification of these parameters. In particular, our objective is to provide a global recursive estimator of the parameters of the Class A model which performs well for all parameter vectors in the parameter set of interest. We begin by proposing a basic, physically-motivated, decision-directed algorithm. This decision-directed scheme is based on an adaptive Bayesian classification of each Class A envelope sample as being either impulsive or background. As each sample is so classified, recursive updates of the estimates of the following three quantities are obtained: the second moment of the impulsive component of the interference envelope density, the second moment of the background component of the interference envelope density, and the probability with which the impulsive component occurs. From these estimates, estimates of the parameters of the model are readily obtained, since closed-form expressions for the parameters exist in terms of these three quantities. Examination of the performance of the proposed estimator via simulation reveals two major shortcomings of the scheme, which adversely affect its performance even in a local setting. However, by appropriately modifying the basic algorithm and imposing the necessary restrictions on the form of one of its initiation vectors, a global recursive estimator of the parameters of the Class A model with excellent performance characteristics can be obtained.

The chapter is organized as follows. In Section 4.2, the basic decision-directed algorithm is developed and its performance examined. Via a probability-of-error analysis, it is seen that the degradation in the performance of the algorithm arising from its two basic shortcomings can be alleviated if the necessary restrictions are imposed and the appropriate modifications are incorporated. Consequently, in Section 4.3, an initiation procedure for each parameter which yields initial estimates of that parameter satisfying the necessary restrictions is presented. Upon using the estimates obtained from these procedures as initial estimates of the parameters, a modified version

of the basic algorithm is then proposed which incorporates the modifications suggested by the two flaws of the algorithm and some additional modifications which are deemed necessary for improving its performance in a global framework. In Section 4.4, the moderate-sample-size performance of the modified algorithm is explored extensively via simulation. From these simulations, it is seen that the proposed global decision-directed scheme does, in fact, provide a global estimator of the parameters that performs very well for all parameter vectors in the parameter set of interest. Some concluding remarks are contained in Section 4.5.

4.2. A Basic Decision-Directed (BDD) Algorithm

The problem of recursive estimation of the Class A parameters from an independent sequence of Class A envelope samples will now be considered. The procedure which will be proposed in this section is a recursive version of the Threshold-Comparison estimator, which was seen in Section 3.4 to provide good estimates of the parameters. The objective in the batch scheme and its recursive version is optimally to discriminate between background and impulsive samples, the optimality criterion being minimization of the probability of an incorrect classification. Given that the samples can be so classified, accurate estimates of the parameters can then be obtained. Of course, the optimum decision statistic in this case is given by the likelihood ratio test (LRT). Fortunately, as will be seen later, for each parameter vector in the parameter set of interest, the likelihood ratio (LR) is strictly monotone increasing in the envelope sample. Thus, to each parameter vector in the parameter set of interest, we can associate a unique threshold so that, for a given observation, the LRT is equivalent to comparing that observation to this optimum threshold. The problem of optimum discrimination of the samples has then been transformed to the problem of locating the optimum threshold corresponding to the true parameter vector. This is what the proposed algorithm attempts to do.

4.2.1. Formulation of the algorithm

The proposed recursive scheme is a basic decision-directed (BDD) algorithm based on an adaptive Bayesian classification of each of a sequence of independent Class A envelope samples as

background or impulsive. The mathematical formulation for the algorithm is given as follows: Let w_{un} denote the unnormalized Class A envelope pdf, i.e.,

$$w_{un} = \frac{1}{\sigma} w\left(\frac{z}{\sigma}\right), \qquad \sigma^2 > 0.$$

The parameter σ^2, the second moment of the interference envelope, was assumed to be unity in the previously considered batch estimation problem, since, in a batch setting, the data can be easily normalized to have unit second moment. As stated in Section 3.4, we can decompose the Class A envelope pdf w_{un} into two components. The first of these corresponds to the $m = 0$ term and the second corresponds to all terms indexed by $m \geq 1$, i.e.,

$$w_{un}(z) = e^{-A}\left[\frac{2z}{\sigma^2\sigma_0^2} e^{-z^2/\sigma^2\sigma_0^2}\right] + (1-e^{-A})\left[\frac{2e^{-A}\sum_{m=1}^{\infty} \frac{A^m}{m!\sigma_m^2} z\, e^{-z^2/\sigma^2\sigma_m^2}}{\sigma^2(1-e^{-A})}\right], \qquad z \geq 0. \tag{4.1}$$

$$= (1-\pi_1)\, p_0(z) + \pi_1\, p_1(z)$$

where

$$\pi_1 \triangleq 1 - e^{-A}. \tag{4.2a}$$

$$p_0(z) \triangleq \frac{2z}{\sigma^2\sigma_0^2} e^{-z^2/\sigma^2\sigma_0^2}. \tag{4.2b}$$

and

$$p_1(z) \triangleq \frac{2e^{-A}}{\sigma^2\pi_1} \sum_{m=1}^{\infty} \frac{A^m}{m!\sigma_m^2} z\, e^{-z^2/\sigma^2\sigma_m^2}. \tag{4.2c}$$

In the sequel, p_0 will be referred to as "the background component of w_{un}" and p_1 as "the impulsive

component of w_{un}."[1] Thus, Class A envelope samples attributable to p_0 will be referred to as "background samples" and those attributable to p_1 will be referred to as "impulsive samples." Let σ_B^2 denote the second moment of w_{un} conditioned on the event that p_0 occurred, and let σ_I^2 denote the second moment of w_{un} conditioned on the event that p_1 occurred, i.e.,

$$\sigma_B^2 = \int_0^\infty z^2 p_0(z)\,dz \quad \text{and} \quad \sigma_I^2 = \int_0^\infty z^2 p_1(z)\,dz \ . \tag{4.3}$$

From (4.2a)-(4.2c) and (4.3), we then have that

$$\pi_1 = 1 - e^{-A} , \tag{4.4a}$$

$$\sigma_B^2 = \sigma^2 \left[\frac{K}{A + K} \right] , \tag{4.4b}$$

and

$$\sigma_I^2 = \sigma^2 \left[\frac{A + \pi_1 K}{\pi_1 (A + K)} \right] . \tag{4.4c}$$

These equations can be readily inverted to yield unique, closed-form expressions for the parameters of the Class A model in terms of π_1, σ_B^2, and σ_I^2. Specifically,

$$A = -\ln(1 - \pi_1) , \tag{4.5a}$$

$$K = \left[-\ln(1 - \pi_1) \right] \left[\frac{\sigma_B^2}{\pi_1 \sigma_I^2 - \pi_1 \sigma_B^2} \right] , \tag{4.5b}$$

and

$$\sigma^2 = (1 - \pi_1)\sigma_B^2 + \pi_1 \sigma_I^2 \tag{4.5c}$$

for all $(A, K, \sigma^2) \in \Lambda'$, where $\Lambda' \triangleq \{(A, K, \sigma^2) : (A, K)^T \in \Lambda \text{ and } \sigma^2 > 0\}$.

---

[1]That p_0 can justifiably be referred to as the "background component of w" follows from the discussion given in Chapter 2. Furthermore, since p_1 is primarily attributable to the impulsive component of the input noise and, since the variance of p_1 is significantly larger than that of p_0 for parameter vectors in the parameter set of interest, p_1 can also justifiably be referred to as "the impulsive component of w." Again, this terminology is consistent with that given, e.g., in [16].

Given that we can distinguish between envelope samples attributable to the background component of w_{un} and those corresponding to the impulsive component, then, based on the definitions of σ_B^2, σ_I^2, and π_1, we can use as estimates of these quantities the sample second moment of those samples classified as background, the sample second moment of those samples classified as impulsive, and the frequency with which the impulsive samples occur, respectively. From these estimates, we can then obtain estimates of A, K, and σ^2 using the relations given in (4.5a)-(4.5c). Since the classification of each sample as "impulsive" or "background" can be performed using a likelihood ratio test based on that sample and on the estimate of (A, K, σ^2) obtained at the previous iteration of the algorithm, we are now in a position to propose the following decision-directed recursive scheme for estimating the parameters of the Class A model, wherein an adaptive Bayesian classification of each sample as impulsive or background is performed :

Basic Decision-Directed (BDD) Algorithm

Step 1 : Choose the initiation vectors. Choose $(\tilde{\pi}_1(0), \tilde{\sigma}_B^2(0), \widetilde{\pi_1\sigma_I^2}(0))$ arbitrarily and $(\tilde{A}_0, \tilde{K}_0, \tilde{\sigma}_0^2) \in \Lambda'$. (A tilde above a given quantity denotes the estimate of that quantity for the iteration shown after that estimate either parenthetically, when estimates of π_1, σ_B^2, or $\pi_1\sigma_I^2$ are being considered, or as a subscript, when estimates of A, K, or σ^2 are being considered. Note that we are considering estimates of $\pi_1\sigma_I^2$ instead of σ_I^2 since σ_I^2 always appears in conjunction with π_1 in the expressions for the parameters given in (4.5a)-(4.5c).) The initiation vector $(\tilde{\pi}_1(0), \tilde{\sigma}_B^2(0), \widetilde{\pi_1\sigma_I^2}(0))$ is chosen arbitrarily since, as will be seen from the form of the update equations given in Step 3, the performance of the algorithm is independent of this choice. At the n-th iteration $(n \geq 1)$, we have the estimate vectors $(\tilde{\pi}_1(n-1), \tilde{\sigma}_B^2(n-1), \widetilde{\pi_1\sigma_I^2}(n-1))$ and $(\tilde{A}_{n-1}, \tilde{K}_{n-1}, \tilde{\sigma}_{n-1}^2)$ and we observe the n-th sample Z_n.

Step 2 : Classify Z_n as an impulsive sample or as a background sample using a likelihood ratio test based on the estimate of (A, K, σ^2) obtained at the $(n-1)$-st iteration.

Let

$$\phi_n = \begin{cases} 1 & \text{if } f(Z_n : \tilde{A}_{n-1}, \tilde{K}_{n-1}, \tilde{\sigma}^2_{n-1}) > 1 \\ 0 & \text{if } f(Z_n : \tilde{A}_{n-1}, \tilde{K}_{n-1}, \tilde{\sigma}^2_{n-1}) \leqslant 1 \end{cases},$$

where f is the likelihood ratio function normalized so that the threshold of the LRT has unity value, i.e.,

$$\begin{aligned} f(z; A, K, \sigma^2) &\triangleq \frac{(1-e^{-A})\, p_1(z; A, K, \sigma^2)}{e^{-A}\, p_0(z; A, K, \sigma^2)} \qquad , z > 0, \\ &= \sum_{m=1}^{\infty} \frac{A^m}{m!} \left| \frac{K}{m+K} \right| e^{\frac{z^2}{\sigma^2}\left|\frac{m}{m+K}\right|\left|\frac{A+K}{K}\right|} \qquad , z > 0, \end{aligned} \tag{4.6}$$

and where

$\phi_n = 1 \Rightarrow Z_n$ is classified as an impulsive sample,

$\phi_n = 0 \Rightarrow Z_n$ is classified as a background sample.

(The functions p_0 and p_1 have been defined in (4.2b) and (4.2c), respectively. Here the dependence of these functions on the parameters is made explicit.) In the sequel, the function f will be referred to as the normalized likelihood ratio (NLR) function.

Step 3 : Update recursively the estimates of π_1, σ_B^2, and $\pi_1\sigma_I^2$. (These three parameters will be referred to as the update parameters.) : For $n \geqslant 1$, let

$$\tilde{\pi}_1(n) = \tilde{\pi}_1(n-1) + \frac{1}{n}(\phi_n - \tilde{\pi}_1(n-1)) \quad . \tag{4.7a}$$

$$\tilde{\sigma}_B^2(n) = \begin{cases} \tilde{\sigma}_B^2(n-1) + \dfrac{(1-\phi_n)}{\sum_{l=1}^{n}(1-\phi_l)}\,(Z_n^2 - \tilde{\sigma}_B^2(n-1)) & \text{if } \sum_{l=1}^{n}(1-\phi_l) \neq 0 \\ \tilde{\sigma}_B^2(n-1) & \text{if } \sum_{l=1}^{n}(1-\phi_l) = 0 \end{cases} \tag{4.7b}$$

and

$$\widetilde{\pi_1\sigma_I^2}(n) = \widetilde{\pi_1\sigma_I^2}(n-1) + \frac{1}{n}\left(\phi_n Z_n^2 - \widetilde{\pi_1\sigma_I^2}(n-1)\right) . \tag{4.7c}$$

Note that at the first iteration of the algorithm ($n = 1$), execution of this step results in cancellation of the terms involving the initial estimate of π_1 in (4.7a) and those involving the initial estimate of $\pi_1\sigma_I^2$ in (4.7c). Furthermore, at the first iteration n' of the algorithm for which a given sample is classified as a background sample, the first portion of (4.7b) becomes effective and execution of this step results in cancellation of the terms involving the initial estimate of σ_B^2 in this portion of (4.7b). Since, as will be seen in the next step of the algorithm, the classificiation of each sample as background or impulsive depends only on the value of the initiation vector $(\tilde{A}_0, \tilde{K}_0, \tilde{\sigma}_0^2)$ when $n \leq n'$, it follows that for a given sample sequence the values of $\tilde{\pi}_1(1)$, $\tilde{\sigma}_B^2(n')$, and $\widetilde{\pi_1\sigma_I^2}(1)$ are unaffected by the choice of $(\tilde{\pi}_1(0), \tilde{\sigma}_B^2(0), \widetilde{\pi_1\sigma_I^2}(0))$. Moreover, the estimate of each update parameter obtained at a given iteration depends on the estimate of that parameter obtained at the previous iteration only. Consequently, for a given sample sequence, initiation vector $(\tilde{A}_0, \tilde{K}_0, \tilde{\sigma}_0^2)$, and fixed value of $n \geq n'$, the estimate vector $(\tilde{\pi}_1(n), \tilde{\sigma}_B^2(n), \widetilde{\pi_1\sigma_I^2}(n))$ will have the same value independent of the choice of the initiation vector $(\tilde{\pi}_1(0), \tilde{\sigma}_B^2(0), \widetilde{\pi_1\sigma_I^2}(0))$. Thus, as claimed in Step 1, the performance of the algorithm is unaffected by the choice of the initial estimates of π_1, σ_B^2, and $\pi_1\sigma_I^2$.

Note, in addition, that the estimates of the update parameters given by recursions (4.7a)-(4.7c) are motivated entirely by their definitions: For $n \geq 1$, $\tilde{\pi}_1(n)$ as given by (4.7a) is simply the proportion of samples that have been classified as impulsive by the n-th iteration; for $n \geq n'$, $\tilde{\sigma}_B^2(n)$ is the sample second moment of those samples that have been classified as background by the n-th iteration; and lastly, for $n \geq 1$, $\widetilde{\pi_1\sigma_I^2}(n)$ is the proportion of those samples that have been classified as impulsive by the n-th iteration times the sample second moment of those samples.

Step 4 : Obtain estimates of the parameters of the Class A model. First, let

$$\alpha_n \triangleq \left(\sum_{l=1}^{n} \phi_l \right) \left(\sum_{l=1}^{n} (1-\phi_l) \right)$$

and let n^* denote the minimum value of n for which α_n is nonzero. Then, for $n < n^*$, set $(\tilde{A}_n , \tilde{K}_n , \tilde{\sigma}_n^2) = (\tilde{A}_{n-1}, \tilde{K}_{n-1}, \tilde{\sigma}_{n-1}^2)$. For $n \geqslant n^*$, obtain estimates of the Class A parameters in terms of $\tilde{\pi}_1$, $\tilde{\sigma}_B^2$, and $\widetilde{\pi_1 \sigma_I^2}$ using the expressions for these parameters given in (4.5a)-(4.5c):

$$\tilde{A}_n = -\ln(1-\tilde{\pi}_1(n)) \ , \tag{4.8a}$$

$$\tilde{K}_n = [-\ln(1-\tilde{\pi}_1(n))] \left[\frac{\tilde{\sigma}_B^2(n)}{\widetilde{\pi_1 \sigma_I^2}(n) - \tilde{\pi}_1(n)\, \tilde{\sigma}_B^2(n)} \right] , \tag{4.8b}$$

and

$$\tilde{\sigma}_n^2 = (1-\tilde{\pi}_1(n))\, \tilde{\sigma}_B^2(n) + \widetilde{\pi_1 \sigma_I^2}(n) \ . \tag{4.8c}$$

(Note that (4.8a)-(4.8c) do not yield valid estimates of the parameters for $n < n^*$.)

Step 5 : Constrain $(\tilde{A}_n , \tilde{K}_n , \tilde{\sigma}_n^2)$ to lie in the parameter set of interest. First, extend the boundary of Λ' slightly (by a factor of 1.1) in two of its coordinates to obtain the following set Λ^* containing Λ':

$$\Lambda^* \triangleq \left\{ (A , K , \sigma^2) : 9.\overline{09} \times 10^{-3} \leqslant A \leqslant 1.1, \ 9.\overline{09} \times 10^{-7} \leqslant K \leqslant 1.1 \times 10^{-2}, \ \sigma^2 > 0 \right\} .$$

Secondly, let

$$\beta_n^1 \triangleq \tilde{A}_n - 9.\overline{09} \times 10^{-3},$$

$$\beta_n^2 \triangleq 1.1 - \tilde{A}_n ,$$

$$\beta_n^3 \triangleq \tilde{K}_n - 9.\overline{09} \times 10^{-7},$$

$$\beta_n^4 \triangleq 1.1 \times 10^{-2} - \tilde{K}_n .$$

Thirdly, modify $(\tilde{A}_n , \tilde{K}_n , \tilde{\sigma}_n^2)$ as follows to obtain the constrained (to lie in Λ^*) estimate $(\bar{A}_n , \bar{K}_n , \bar{\sigma}_n^2)$ of (A , K , σ^2) for the n-th iteration:

$$\bar{A}_n = \begin{cases} \tilde{A}_n & \text{if } \beta_n^1 \beta_n^2 \geq 0 \\ (1 - \max\{sgn\ \beta_n^1, 0\})(9.\overline{09} \times 10^{-3}) & \\ \quad + (1 - \max\{sgn\ \beta_n^2, 0\})(1.1) & \text{if } \beta_n^1 \beta_n^2 < 0 \end{cases}, \tag{4.9a}$$

$$\bar{K}_n = \begin{cases} \tilde{K}_n & \text{if } \beta_n^3 \beta_n^4 \geq 0 \\ (1 - \max\{sgn\ \beta_n^3, 0\})(9.\overline{09} \times 10^{-7}) & \\ \quad + (1 - \max\{sgn\ \beta_n^4, 0\})(1.1 \times 10^{-2}) & \text{if } \beta_n^3 \beta_n^4 < 0 \end{cases}, \tag{4.9b}$$

$$\bar{\sigma}_n^2 = \tilde{\sigma}_n^2 . \tag{4.9c}$$

Now, if $(\bar{A}_n, \bar{K}_n, \bar{\sigma}_n^2) = (\tilde{A}_n, \tilde{K}_n, \tilde{\sigma}_n^2)$, then execution of this step is complete. However, if $(\bar{A}_n, \bar{K}_n, \bar{\sigma}_n^2) \neq (\tilde{A}_n, \tilde{K}_n, \tilde{\sigma}_n^2)$, then proceed as follows: Using the inverse relations given in (4.4a)-(4.4c), modify $\tilde{\pi}_1(n)$, $\tilde{\sigma}_B^2(n)$, and $\widetilde{\pi_1\sigma_I^2}(n)$ at the n-th iteration to reflect the above changes in the estimates of the Class A parameters. This modified estimate will be denoted by $(\bar{\pi}_1(n), \bar{\sigma}_B^2(n), \overline{\pi_1\sigma_I^2}(n))$:

$$\bar{\pi}_1(n) = 1 - e^{-\bar{A}_n} , \tag{4.10a}$$

$$\bar{\sigma}_B^2(n) = \bar{\sigma}_n^2 \left[\frac{\bar{K}_n}{\bar{A}_n + \bar{K}_n} \right] , \tag{4.10b}$$

and

$$\overline{\pi_1\sigma_I^2}(n) = \bar{\sigma}_n^2 \left[\frac{\bar{A}_n + \bar{\pi}_1(n)\bar{K}_n}{\bar{A}_n + \bar{K}_n} \right] . \tag{4.10c}$$

Finally, set $(\tilde{\pi}_1(n), \tilde{\sigma}_B^2(n), \widetilde{\pi_1\sigma_I^2}(n)) = (\bar{\pi}_1(n), \bar{\sigma}_B^2(n), \overline{\pi_1\sigma_I^2}(n))$ and $(\tilde{A}_n, \tilde{K}_n, \tilde{\sigma}_n^2) = (\bar{A}_n, \bar{K}_n, \bar{\sigma}_n^2)$.

The steps of the basic decision-directed algorithm are now complete. From a graphical standpoint, this is what the BDD algorithm attempts to do: Let (A, K, σ^2) denote a parameter vector in Λ' with corresponding envelope pdf w_{un} as shown in Fig. 4.1. Now, for each $(\underline{A}, \underline{K}, \underline{\sigma}^2) \in \Lambda'$,

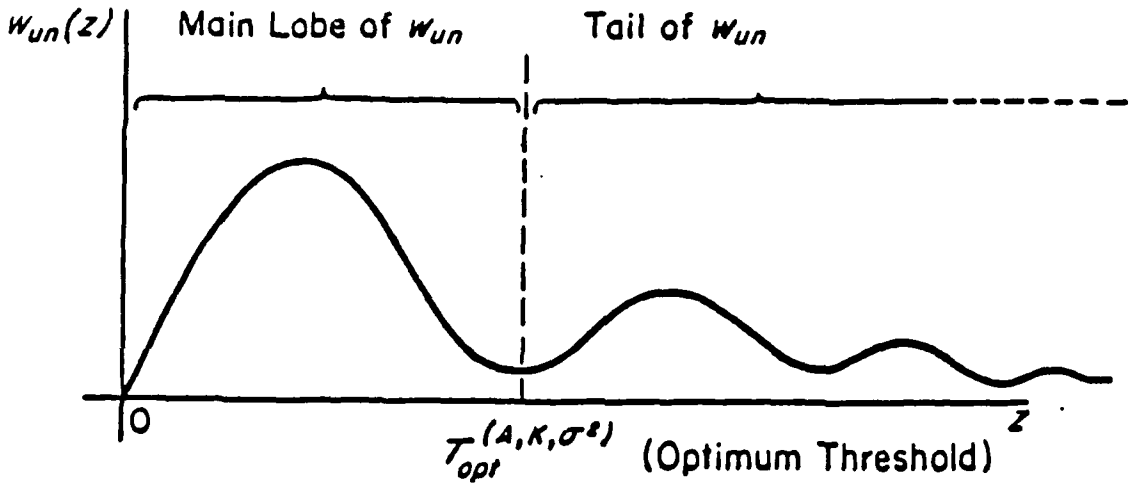


Fig. 4.1. Envelope pdf for typical parameter vector (A, K, σ^2) in Λ' .

$f(z; \underline{A}, \underline{K}, \underline{\sigma}^2)$ as defined in (4.6) is a strictly increasing continuous function of z, with $\lim_{z \to 0} f(z; \underline{A}, \underline{K}, \underline{\sigma}^2) < 1$ and $\lim_{z \to \infty} f(z; \underline{A}, \underline{K}, \underline{\sigma}^2) = \infty$. Thus, for each $(\underline{A}, \underline{K}, \underline{\sigma}^2) \in \Lambda'$, there exists a unique $\tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)} \in (0, \infty)$ for which $f(\tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)}; \underline{A}, \underline{K}, \underline{\sigma}^2) = 1$. It then follows that for each $(\underline{A}, \underline{K}, \underline{\sigma}^2) \in \Lambda'$, the corresponding likelihood ratio test for a given observation is equivalent to comparing the given observation to the optimum threshold $\tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)}$, i.e., the decision regions which minimize the probability of error in the classification process for each $(\underline{A}, \underline{K}, \underline{\sigma}^2) \in \Lambda'$ consist of the intervals $(0, \tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)})$, $(\tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)}, \infty)$. (If the observation lies in $(0, \tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)}]$, it is classified as a background sample, whereas if the observation lies in $(\tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)}, \infty)$, it is classified as an impulse.) Consequently, since (A, K, σ^2) lies in Λ', we can associate with the parameter vector (A, K, σ^2) the optimum threshold $\tau_{opt}^{(A, K, \sigma^2)}$ in the manner just described. Similarly, since the sequence of estimate vectors $(\tilde{A}_n, \tilde{K}_n, \tilde{\sigma}_n^2)$ of the true parameter vector (A, K, σ^2) lies in Λ', we can associate with this sequence of estimate vectors a corresponding sequence of threshold estimates $\tau_{opt}^{(\tilde{A}_n, \tilde{K}_n, \tilde{\sigma}_n^2)}$. Implicitly, via this sequence of threshold estimates, the BDD algorithm attempts to locate $\tau_{opt}^{(A, K, \sigma^2)}$ and, hence, the corresponding optimum decision regions $(0, \tau_{opt}^{(A, K, \sigma^2)}]$, $(\tau_{opt}^{(A, K, \sigma^2)}, \infty)$. In so doing, it can then, with minimum error probability, discriminate between those samples corresponding to the main lobe of the envelope pdf and those corresponding to the tail of the pdf (see Fig. 4.1). Given that the background samples (those corresponding to the main lobe of the envelope pdf) can be optimally discriminated from the impulsive samples (those corresponding to the tail of the pdf), accurate estimates of the Class A model parameters can then be obtained.

We see then that the BDD algorithm is physically motivated, easy to implement, and is a recursive version of a batch procedure which is known to provide good estimates of the parameters. In the next section, we will examine the performance of this BDD algorithm.

4.2.2. Performance of BDD algorithm

The behavior of the BDD algorithm has been examined for an extensive range of true parameter vectors $(A, K, \sigma^2) \in \Lambda'$ and an equally extensive range of initiation vectors $(\tilde{A}_0, \tilde{K}_0, \tilde{\sigma}_0^2) \in \Lambda'$ for each true parameter vector. A few major difficulties have been observed, these difficulties being equally apparent when $\sigma^2 = 1$ and $\tilde{\sigma}_n^2$ is fixed to have unity value for all $n \geq 0$. Thus, for the sake of simplicity, we will now cite these difficulties as they pertain to the situation when $\tilde{\sigma}_n = \sigma^2 = 1$ $(n \geq 0)$. We note that σ^2 is taken to have unity value since the *absolute* value of σ^2 has no bearing on the estimation problem at hand (see Appendix A).

Drawbacks of BDD Algorithm

(i) For values of A close to 10^{-2} and arbitrary values of K, the convergence of the BDD algorithm to the true parameter vector is sensitive to the distribution of impulses over values of $n < O(500)$. That is, despite the fact that the algorithm may correctly distinguish between impulsive and background samples in its initial stages, if the percentage of impulses over the sequence of samples classified correctly exceeds the expected percentage (which, e.g., would be 0.995% for $A = 10^{-2}$ since, for $A = 10^{-2}$, $\pi_1 = 1 - e^{-A} \approx 9.95 \times 10^{-3}$), then, with significant probability, the algorithm will not converge to the true parameter vector. Furthermore, for fixed A, the frequency with which the algorithm does not converge to the true parameter vector due to its sensitivity to the distribution of impulses increases with increasing K, becoming relatively high for values of K close to 10^{-2}.

(ii) For values of $\tilde{\Gamma}_0 \triangleq \tilde{K}_0 / \tilde{A}_0 \geq \Gamma$, the frequency with which the algorithm converges to the true parameter vector is relatively high, whereas for values of $\tilde{\Gamma}_0 < \Gamma$, the frequency with which the algorithm converges to the *wrong* parameter vector is relatively high. The former portion of this statement is not valid for values of A close to 10^{-2}, since, as explained in (i), the distribution of impulses forces the algorithm to converge to the wrong parameter vector for a significant percentage of the runs when A is close to 10^{-2}. Moreover, for values of A close to 1, the frequency with which the algorithm converges to the wrong parameter vector once

again becomes relatively high for values of $\tilde{\Gamma}_0 \sim O(\geq 10^{-1})$. Thus, even when the set of vectors from which the initiation vectors are chosen is a small neighborhood of the true parameter vector, the performance of the BDD algorithm can vary drastically over that neighborhood depending on the location of the initiation vector. This highly nonuniform behavior of the algorithm even when the initiation vectors lie close to the true parameter vector makes it undesirable for use even as a local tracking scheme.

A careful and thorough analysis of the source of these two shortcomings of the BDD algorithm has been made, and is given in Appendix A. We note here that the ensuing adverse effects on the performance of the algorithm arising from its two basic drawbacks can be eliminated by placing certain restrictions on the form of its initiation vector and by incorporating the appropriate modifications into its framework. In particular, the following restrictions (on $(\tilde{A}_0, \tilde{K}_0, \tilde{\sigma}_0^2)$) and modifications must be imposed (the derivation of these conditions can be found in Appendix A):

(R1) $\tilde{A}_0$ must either provide an accurate estimate of A, or, $\tilde{A}_0$ must provide an estimate of A for which $\tilde{A}_0 < A$ and not less by an order of magnitude or more.

(R2) $\tilde{K}_0$ must either provide an accurate estimate of K, or, $\tilde{K}_0$ must provide an estimate of K for which $\tilde{K}_0 > K$ and not greater by two orders of magnitude or more.

(R3) $\tilde{\sigma}_0^2$ must provide an accurate estimate of σ^2.

(M1) The estimator of σ^2 given by (4.8c) must be replaced by an estimator of σ^2 consisting of an update equation for $\tilde{\sigma}_0^2$.

(M2) The estimate of A must be fixed to its initial value $\tilde{A}_0$ in the initial stages of the algorithm, with only the estimates of K and σ^2 being updated.

In the next section, we will consider a modified BDD algorithm which incorporates modifications (M1) and (M2) and whose initiation vector satisfies conditions (R1)-(R3).

4.3. A Global Decision-Directed Algorithm

In this section, we will develop a global recursive estimator of the Class A parameters by appropriately modifying the BDD algorithm. First, we will present an initiation procedure for each parameter which yields initial estimates of the parameter satisfying the corresponding restriction as given in (R1)-(R3) of the previous section. Then, we will propose a modified BDD algorithm which incorporates the changes described under (M1) and (M2) and additional changes which are deemed necessary either for the sake of simplifying the BDD algorithm at a given step or for the sake of improving its performance in a global framework.

Initiation Procedure for A

We need to locate an initiation procedure which yields estimates $\tilde{A}_0$ satisfying restriction (R1). The search for such an estimator can be decomposed into two steps: (i) First, we will locate a procedure which provides us with a reasonable estimate of A. (ii) Secondly, we will construct a quantizer whose input is this estimate and whose output is the estimate quantized in the direction of small A. This quantized estimate will then be used for $\tilde{A}_0$. By decomposing the search in this manner, the determination of a procedure which yields initial estimates of A with the desired property is greatly simplified.

Let us focus our attention on the first step: Which estimator will provide us with a reasonable estimate of A? One procedure which suggests itself from the batch estimation problem discussed in the previous chapter is the method of moments. It was seen in Section 3.2 that the *MM* estimator based on the fourth and sixth moments was highly inefficient in estimating the parameters of the model. However, this high inefficiency was due to the insensitivity of the moments to changes in the parameter K. In fact, the computed asymptotic variances for the normalized *MM* estimate of A (given in Table 3.1) suggest that a relative error in the estimate of A on the $O(10^{-1})$ can be attained using the *MM* estimator if a sample size of 10000 is used.[2] Thus, given the many

[2] The computed variances were based on the assumption of a fixed envelope second moment. However, they should not change significantly when the second moment is unknown since, as will be seen in the sequel, the sample second moment is extremely accurate for 10000 samples.

desirable features associated with the MM estimator, among which are its recursivity and computational expediency, we will use the MM estimator based on 10000 samples to obtain a reasonable estimate of the parameter A.

In some cases, the estimates of A obtained via the method of moments may exceed the true A by a significant amount. Unfortunately, restriction (R1) does not admit such estimates of A for $\tilde{A}_0$. Thus, instead of using the MM estimate of A for $\tilde{A}_0$, we will instead use a quantized MM estimate of A, wherein the MM estimate of A is quantized in the direction of small A.

Let m_2, m_4, and m_6 denote the second-order, fourth-order, and sixth-order sample moments, respectively, of a sequence of 10000 independent, unnormalized Class A envelope samples and let A_{MM} denote the MM estimate of A based on these moments. Then, (see 3.3a),

$$A_{MM} = \frac{\left(\dfrac{m_4}{2(m_2)^2} - 1\right)^3}{\left(\dfrac{m_6}{6(m_2)^3} - \dfrac{3m_4}{2(m_2)^2} + 2\right)^2} . \tag{4.11}$$

Furthermore, let $\epsilon_A(A, K)$ denote the asymptotic variance of the normalized MM estimate of A based on the fourth and sixth moments, the expression for which was derived in Section 3.2.3. Consider the following simple quantization scheme:

(i) Divide $(-\infty, \infty)$ into the subintervals $(-\infty, 1.1 \times 10^{-2})$, $[A_j, A_j^{\max})$ $(1 \leqslant j \leqslant 990)$, and $[A_{990}^{\max}, \infty)$, where

$$A_j \triangleq j\, 10^{-3} + 10^{-2} \tag{4.12a}$$

and

$$A_j^{\max} \triangleq A_j + \left\lfloor \left(\frac{\epsilon_A(A_j, 10^{-2})}{10000} \right)^{1/2} \times A_j \right\rfloor . \tag{4.12b}$$

(ii) Then, take

$$\tilde{A}_0 = \begin{cases} 10^{-2} & \text{if } A_{MM} \in (-\infty, 1.1 \times 10^{-2}) \\ \left[\min_{\substack{1 \le j \le 990 \\ s.t.\ I_{[A_j, A_j^{\max})}(A_{MM}) = 1}} A_j \right] - 10^{-3} & \text{if } A_{MM} \in [1.1 \times 10^{-2}, A_{990}^{\max}) \\ 1 & \text{if } A_{MM} \in [A_{990}^{\max}, \infty) \end{cases} , \tag{4.13}$$

where I_B denotes the indicator function of the set B.

Note that $A_j^{\max}$ is an upper bound which, in the worst case, is correct approximately 85% of the time (based on a Gaussian approximation for the distribution of A_{MM}; see Section 3.2.2). Thus, the above quantization scheme achieves the goal of quantizing toward smaller values of A without undue distortion. This statement is further supported by simulation results wherein the proposed initiation procedure was in fact seen to yield estimates $\tilde{A}_0$ satisfying restriction (R1). (Note that the quantization scheme must be independent of the parameter K since K is unknown. Now, from Table 3.1, it is evident that for fixed A and arbitrary K, $\epsilon_A(A, K)$ is largest for $K = 10^{-2}$. Thus, the choice of $K = 10^{-2}$ in (4.12b) yields the most conservative bound.)

Initiation Procedure for K

We need to locate an initiation procedure for $\tilde{K}_0$ which satisfies restriction (R2). Again, we decompose the search for such a procedure into two steps: First, we will locate an estimator which yields an estimate of K which differs from K by less than an order of magnitude. Then, we will multiply this estimate by a factor of 10 and use the resulting value for $\tilde{K}_0$. In so doing, we will obtain values for $\tilde{K}_0$ which satisfy (R2).

Now, we need an estimator which will approximate K to within an order of magnitude. It was seen in Section 3.4 that estimators which correctly distinguish between impulsive and background samples provide good estimates of the Class A parameters. Moreover, it was seen in our earlier discussion that correct discrimination of the samples involves the determination of the optimum threshold corresponding to the true parameter vector. Thus, by considering a recursive

version of the Threshold-Comparison estimator (given in Section 3.4) which utilizes a simple, heuristic scheme for approximating the threshold, perhaps we can obtain an estimator which yields estimates of K having the desired property. With this in mind, consider the following initiation procedure for obtaining $\tilde{K}_0$:

Let $Z_{-999}, Z_{-998}, \ldots, Z_0, Z_1, \ldots, Z_{2000}$ denote a sequence of 3000 independent Class A envelope samples.

Step 1: Choose $(\tilde{\gamma}(0), \tilde{\sigma}_a^2(0), \tilde{\gamma}\tilde{\sigma}_b^2(0))$ arbitrarily and let

$$X_0 \triangleq \max\{Z_{-999}, \ldots, Z_0\} .$$

$$Y_0 \triangleq \min\{Z_{-999}, \ldots, Z_0\} .$$

Step 2: For $0 \leq k \leq 1999$, update the threshold τ_k using the following system of equations:

$$\tau_k = (X_k Y_k)^{1/2}$$

$$X_{k+1} = X_k + \frac{\lambda(Z_{k+1} - \tau_k)}{[\sum_{l=0}^{k} \lambda(Z_{l+1} - \tau_l)] + 1} [Z_{k+1} - X_k]$$

$$Y_{k+1} = Y_k + \frac{[1 - \lambda(Z_{k+1} - \tau_k)]}{[\sum_{l=0}^{k} (1 - \lambda(Z_{l+1} - \tau_l))] + 1} [Z_{k+1} - Y_k] ,$$

where

$$\lambda(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} ,$$

and

$\lambda(Z_{k+1} - \tau_k) = 1 \Rightarrow Z_{k+1}$ is classified as an impulsive sample.

$\lambda(Z_{k+1} - \tau_k) = 0 \Rightarrow Z_{k+1}$ is classified as a background sample.

Step 3: For $1000 \leqslant k \leqslant 1999$, update recursively the estimates of γ, σ_a^2, and $\gamma\sigma_b^2$:

$$\tilde{\gamma}(k-999) = \tilde{\gamma}(k-1000) + \frac{1}{k-999}[\lambda(Z_{k+1} - \tau_k) - \tilde{\gamma}(k-1000)] \,. \quad (4.14a)$$

$$\tilde{\sigma}_a^2(k-999) = \begin{cases} \tilde{\sigma}_a^2(k-1000) + \dfrac{(1-\lambda(Z_{k+1}-\tau_k))}{\sum_{l=1000}^{k}(1-\lambda(Z_{l+1}-\tau_l))}[Z_{k+1}^2 - \tilde{\sigma}_a^2(k-1000)] & \text{if } \sum_{l=1000}^{k}(1-\lambda(Z_{l+1}-\tau_l)) \neq 0 \\ \tilde{\sigma}_a^2(k-1000) & \text{if } \sum_{l=1000}^{k}(1-\lambda(Z_{l+1}-\tau_l)) = 0 \,. \end{cases} \quad (4.14b)$$

and

$$\widetilde{\gamma\sigma}_b^2(k-999) = \widetilde{\gamma\sigma}_b^2(k-1000) + \frac{1}{k-999}[\lambda(Z_{k+1}-\tau_k)Z_{k+1}^2 - \widetilde{\gamma\sigma}_b^2(k-1000)] \,. \quad (4.14c)$$

Step 4: Compute $\tilde{K}_0$ as follows:

First, let

$$K^* = [-\ln(1-\tilde{\gamma}(1000))]\left|\frac{\tilde{\sigma}_a^2(1000)}{\widetilde{\gamma\sigma}_b^2(1000) - \tilde{\gamma}(1000)\tilde{\sigma}_a^2(1000)}\right| \,. \quad (4.15)$$

Then, let

$$K^{**} = 10\,K^* \,.$$

Finally, take

$$\tilde{K}_0 = \begin{cases} 10^{-2} & \text{if } K^{**} > 10^{-2} \\ K^{**} & \text{if } K^{**} \in [10^{-6}, 10^{-2}] \\ 10^{-6} & \text{if } K^{**} < 10^{-6} \end{cases} \,. \quad (4.16)$$

This initiation procedure attempts to locate the optimum threshold corresponding to the true parameter vector via the sequence of threshold estimates τ_k. Note the simplicity of the scheme used to approximate this optimum threshold: X_k is the average of X_0 and the values of those samples classified as impulsive by the k-th iteration; Y_k is the average of Y_0 and the values of those samples classified as background by the k-th iteration; and τ_k is simply the geometric mean of these two quantities. Note, moreover, that (4.14a)-(4.14c) are based on the update equations for the BDD algorithm given in (4.7a)-(4.7c) and that the relation for K^* given in (4.15) is obtained using relation (4.8b). (The interpretation of the update parameters γ, σ_a^2, and $\gamma\sigma_b^2$ is the same as that for the parameters π_1, σ_B^2, and $\pi_1\sigma_I^2$, respectively.) Now, simulations reveal that for values of the true parameter vectors for which $K \leq 10^{-4}$, K^* approximates K to within an order of magnitude. However, even though K^* is less than K by less than an order of magnitude for parameter vectors for which $K > 10^{-4}$, K^* sometimes *exceeds* K by more than an order of magnitude for these parameter vectors. But, by multiplying K^* by a factor of 10, constraining the resulting value to lie within the set of allowable values for the parameter K, and using this constrained value for $\tilde{K}_0$, it is seen that estimates $\tilde{K}_0$ which satisfy restriction (R2) are obtained.

Initiation Procedure for σ^2

For $\tilde{\sigma}_0^2$, we will use the sample second moment of those samples used to obtain A_{MM}. Let m_2 be defined as above. Then, we shall take

$$\tilde{\sigma}_0^2 = m_2 \ . \tag{4.17}$$

Simulations have shown that the sample second moment based on 10000 samples yields accurate estimates of the envelope second moment. Thus, estimates $\tilde{\sigma}_0^2$ obtained using (4.17) do, in fact, satisfy restriction (R3).

The descriptions of the initiation procedures for the parameters are now complete. We will now determine the changes in the BDD algorithm induced by (M1) and (M2). First, consider (M1). Since $\tilde{\sigma}_0^2$ is the sample second moment of a sequence of 10000 Class A envelope samples, the

estimator of σ^2 given by (4.8c) will be replaced by the following update equation for this sample second moment:

$$\tilde{\sigma}_n^2 = \tilde{\sigma}_{n-1}^2 + \frac{1}{n + 10000}(Z_n^2 - \tilde{\sigma}_{n-1}^2), \qquad n \geq 1. \tag{4.18}$$

Now, consider (M2). Let n_f denote the last iteration for which the estimate of A is fixed to its initial value $\tilde{A}_0$. Then, for $n \leq n_f$, the estimator of A given by (4.8a) must be replaced by the following estimator:

$$\tilde{A}_n = \tilde{A}_0 \ . \tag{4.19}$$

Now, from (4.5a) and (4.5b), note that

$$K = \left\lfloor \frac{A}{1 - e^{-A}} \right\rfloor \left\lfloor \frac{\sigma_B^2}{\sigma_I^2 - \sigma_B^2} \right\rfloor .$$

Thus, for $n \leq n_f$, the following estimator of K can be used:

$$\tilde{K}_n = \left\lfloor \frac{\tilde{A}_0}{1 - e^{-\tilde{A}_0}} \right\rfloor \left\lfloor \frac{\tilde{\sigma}_B^2(n)}{\tilde{\sigma}_I^2(n) - \tilde{\sigma}_B^2(n)} \right\rfloor , \tag{4.20}$$

where the update equation for σ_B^2 is given by (4.7b) and the update equation for σ_I^2 is given analogously as follows :

$$\tilde{\sigma}_I^2(n) = \begin{cases} \tilde{\sigma}_I^2(n-1) + \left\lfloor \dfrac{\phi_n}{\sum_{l=1}^{n} \phi_l} \right\rfloor (Z_n^2 - \tilde{\sigma}_I^2(n-1)) & \text{if } \sum_{l=1}^{n} \phi_l \neq 0 \\ \tilde{\sigma}_I^2(n-1) & \text{if } \sum_{l=1}^{n} \phi_l = 0 \end{cases} ; \tag{4.21}$$

$\tilde{\sigma}_I^2(0) \in \mathbb{R}$. (If n''' is defined to be the first iteration for which a given sample is classified as impulsive, then for $n \geq n'''$, $\tilde{\sigma}_I^2(n)$ is simply the sample second moment of those samples classified as impulsive by the n-th iteration.) Note from the estimators of the Class A parameters given in

(4.18)-(4.20) that only the estimates of the two update parameters σ_B^2 and σ_I^2 are required when $n \leqslant n_f$. Consequently, in lieu of the update parameters π_1, σ_B^2, and $\pi_1\sigma_I^2$ used previously, the update parameters σ_B^2 and σ_I^2 will be used when $n \leqslant n_f$.

We are now in a position to present the following modified BDD algorithm which incorporates the above modifications and some additional minor modifications.

4.3.1. Modified BDD (MBDD) algorithm

Step 1′: *Choose the initiation vectors.* Choose $(\tilde{\sigma}_B^2(0), \tilde{\sigma}_I^2(0))$ arbitrarily and obtain $(\tilde{A}_0, \tilde{K}_0, \tilde{\sigma}_0^2)$ using the initiation procedures described above.

Step 2′: *Classify sample as impulsive or background.* Since $m \geqslant 1$ in (4.6) and $10^{-6} \leqslant K \leqslant 10^{-2}$, it follows that $m + K \cong m$. Using this approximation for $(m + K)$ in (4.6), classify $Z_n, n \geqslant 1$, as an impulsive sample or as a background sample using the following simplified test:

Let

$$\phi_n = \begin{cases} 1 & \text{if } Z_n > g(\tilde{A}_{n-1}, \tilde{K}_{n-1}, \tilde{\sigma}_{n-1}^2) \\ 0 & \text{if } Z_n \leqslant g(\tilde{A}_{n-1}, \tilde{K}_{n-1}, \tilde{\sigma}_{n-1}^2) \end{cases},$$

where

$$g(\tilde{A}_{n-1}, \tilde{K}_{n-1}, \tilde{\sigma}_{n-1}^2) \triangleq \left| -\tilde{\sigma}_{n-1}^2 \left(\frac{K_{n-1}}{\tilde{A}_{n-1} + \tilde{K}_{n-1}} \right) \ln \left[\tilde{K}_{n-1} \sum_{m \geqslant 1} \frac{\tilde{A}_{n-1}^m}{m!\, m} \right] \right|^{1/2},$$

and where

$\phi_n = 1 \Rightarrow Z_n$ is classified as an impulsive sample,

$\phi_n = 0 \Rightarrow Z_n$ is classified as a background sample.

Step 3′ : *Update recursively the appropriate update parameters.*

(a) For $n \leqslant n_f$, update recursively the estimates of σ_B^2 and σ_I^2 using (4.7b) and (4.21), respectively.

(b) For $n > n_f$, update recursively the estimates of π_1, σ_B^2, and $\pi_1\sigma_I^2$ using (4.7a), (4.7b), and (4.7c), respectively. Note that these update equations require estimates of π_1, σ_B^2, and $\pi_1\sigma_I^2$ when $n = n_f$. Since σ_B^2 is an update parameter when $n \leqslant n_f$, $\tilde{\sigma}_B^2(n_f)$ is available. Furthermore, using relations (4.4a) and (4.4c), estimates of π_1 and $\pi_1\sigma_I^2$ at the n_f -th iteration can be obtained as follows:

$$\tilde{\pi}_1(n_f) = 1 - e^{-\tilde{A}_{n_f}} = 1 - e^{-\tilde{A}_0} \quad ,$$

$$\widetilde{\pi_1\sigma_I^2}(n_f) = \tilde{\sigma}_{n_f}^2 \left| \frac{\tilde{A}_{n_f} + \tilde{\pi}_1(n_f)\tilde{K}_{n_f}}{\tilde{A}_{n_f} + \tilde{K}_{n_f}} \right| = \tilde{\sigma}_{n_f}^2 \left| \frac{\tilde{A}_0 + (1 - e^{-\tilde{A}_0})\tilde{K}_{n_f}}{\tilde{A}_0 + \tilde{K}_{n_f}} \right| .$$

Step 4′ : *Obtain estimates of the Class A parameters.*

Let α_n and n^* be defined as in Step 4 of the BDD algorithm.

(a) For $n < n^*$, obtain $\tilde{\sigma}_n^2$ using (4.18) and set $(\tilde{A}_n , \tilde{K}_n) = (\tilde{A}_{n-1} , \tilde{K}_{n-1})$.

(b) If $n^* \leqslant n_f$, then for $n^* \leqslant n \leqslant n_f$, obtain estimates for A K, and σ^2 using (4.19), (4.20), and (4.18), respectively.

(c) For $n > \max(n_f , n^* - 1)$, obtain estimates for A , K, and σ^2 using (4.8a), (4.8b), and (4.18), respectively.

Step 5′ : *Constrain estimates of Class A parameters to lie in parameter set of interest.*

Define Λ^* , β_n^1 , β_n^2 , β_n^3 , and β_n^4 as in Step 5 of the BDD algorithm.

(a) For $n \leqslant n_f$, constrain $\tilde{A}_n$, $\tilde{K}_n$, and $\tilde{\sigma}_n^2$ using (4.9a), (4.9b), and (4.9c), respectively. (Let $\bar{A}_n$, $\bar{K}_n$, and $\bar{\sigma}_n^2$ denote the constrained estimates.) If $(\bar{A}_n , \bar{K}_n , \bar{\sigma}_n^2) = (\tilde{A}_n , \tilde{K}_n , \tilde{\sigma}_n^2)$, then execution of this step is complete. However, if $(\bar{A}_n , \bar{K}_n , \bar{\sigma}_n^2) \neq (\tilde{A}_n , \tilde{K}_n , \tilde{\sigma}_n^2)$, then proceed as follows : Using the relations given in (4.4a)-(4.4c), modify $\tilde{\sigma}_B^2(n)$ and $\tilde{\sigma}_I^2(n)$ to reflect the

change in the estimates of A, K, and σ^2 induced by the constraint. (Let $\tilde{\sigma}_B^2(n)$ and $\tilde{\sigma}_I^2(n)$ denote the modified estimates.) :

$$\tilde{\sigma}_B^2(n) = \bar{\sigma}_n^2 \left| \frac{\bar{K}_n}{\bar{A}_n + \bar{K}_n} \right| = \bar{\sigma}_n^2 \left| \frac{\bar{K}_n}{\tilde{A}_0 + \bar{K}_n} \right| ,$$

$$\tilde{\sigma}_I^2(n) = \bar{\sigma}_n^2 \left| \frac{\bar{A}_n + (1-e^{-\bar{A}_n})\bar{K}_n}{(1-e^{-\bar{A}_n})(\bar{A}_n + \bar{K}_n)} \right| = \bar{\sigma}_n^2 \left| \frac{\tilde{A}_0 + (1-e^{-\tilde{A}_0})\bar{K}_n}{(1-e^{-\tilde{A}_0})(\tilde{A}_0 + \bar{K}_n)} \right| .$$

Finally, set $(\tilde{\sigma}_B^2(n), \tilde{\sigma}_I^2(n)) = (\bar{\sigma}_B^2(n), \bar{\sigma}_I^2(n))$ and $(\tilde{A}_n, \tilde{K}_n, \tilde{\sigma}_n^2) = (\bar{A}_n, \bar{K}_n, \bar{\sigma}_n^2)$.

(b) For $n > n_f$, proceed as in Step 5 of the BDD algorithm.

The description of the MBDD algorithm is now complete. Experimentation has shown that setting n_f to 1000 eliminates the difficulties associated with not fixing the estimate of A in the initial stages of the algorithm, without unduly slowing the convergence of the algorithm. Moreover, extensive simulation of the MBDD algorithm (with $n_f = 1000$) has shown that convergence of the algorithm is essentially attained within 5000 iterations (i.e., the relative variation in the estimates of the parameters from iteration to iteration for iteration values near 5000 is very slight) and that good estimates of all parameter vectors in the parameter set of interest can generally be obtained for this iteration value. Occasionally, however, for values of $A \sim O(\geq 10^{-1})$ (K arbitrary), the estimate of A can be somewhat low. Furthermore, for these values of A and values of $K \sim O(\geq 10^{-4})$, the estimate of K is occasionally high. This problem can be easily remedied by noting the following: Even though the estimate of A is low and that of K is high, these estimates are closer to the true values than those given by $\tilde{A}_0$ and $\tilde{K}_0$. Thus, by restarting the BDD algorithm with an initiation vector consisting of these estimates of A and K, better estimates of the parameters can be expected. Now, it was noted that the convergence of the MBDD algorithm can be attained within 5000 iterations. However, even after the 3000-th iteration, the variation in the estimates of the parameters is only slight. Thus, the BDD algorithm can equally well be restarted with the estimates of A, K, and σ^2 obtained at the 3000-th iteration. With this in mind, consider

the following modification to the MBDD algorithm:

If the estimate of A at the 3000-th iteration exceeds 10^{-1}, then introduce as a second estimate of the update parameter π_1 the proportion of samples classified as impulsive after the 3000-th iteration. Denote this second estimate by $\approx{\pi}_1$. At some iteration $n_{max} > 3000$ (and for all iterations thereafter), obtain the estimate of A using relation (4.5a) and $\approx{\pi}_1$ as an estimate of π_1, instead of $\tilde{\pi}_1$. If, in addition, the estimate of K at the 3000-th iteration exceeds 10^{-4}, then also introduce as a second estimate of σ_B^2 the proportion of samples classified as background after the 3000-th iteration and as a second estimate of $\pi_1\sigma_I^2$ the proportion of samples classified as impulsive after the 3000-th iteration times the sample second moment of those samples. Denote these estimates by $\approx{\sigma}_B^2$ and $\approx{\pi_1\sigma}_I^2$, respectively. Then, for all iterations $n \geq n_{max}$, obtain the estimate of K using relation (4.5b) and $\approx{\pi}_1$, $\approx{\sigma}_B^2$, and $\approx{\pi_1\sigma}_I^2$ as estimates of the corresponding update parameters. Furthermore, for iteration values greater than 3000 and less than n_{max}, obtain the estimates of A and K using the original estimates of the update parameters. Also, if the estimate of A at the 3000-th iteration exceeds 10^{-1} but that of K does not exceed 10^{-4}, then for $n \geq n_{max}$, continue to obtain the estimate of K using the original estimates of the update parameters.

The first iteration value n_{max}, for which the second estimates of the update parameters are used to obtain the estimates of the model parameters, is the smallest iteration value which satisfies the following two conditions:

(i) There must exist iterations n_1 and n_2, $3000 < n_1, n_2 \leq n_{max}$, such that Z_{n_1} is classified as an impulse and Z_{n_2} is classified as background since, otherwise, invalid estimates of the Class A parameters would be obtained.

(ii) n_{max} must be greater than 3000 plus an offset δ, specified below as a function of the estimate of A at the 3000-th iteration. For iteration values greater than 3000 and less than or equal to $3000 + \delta$, the original estimates of the update parameters are used to obtain the estimates of the model parameters. Consequently, this experimentally determined offset δ is chosen so that, for iteration values slightly greater than $3000 + \delta$,

(a) the accuracy of the estimate of A obtained using $\tilde{\tilde{\pi}}_1$ is, on the average, higher than that obtained using $\tilde{\pi}_1$, and,

(b) when the second estimates of the update parameters are used to obtain the estimate of K, the accuracy of this estimate is, on the average, close to or higher than that obtained using the original estimates of the update parameters.

Otherwise, it would be more appropriate to continue using the original estimates of the update parameters in obtaining the estimates of the model parameters.

By implementing the above modification to the MBDD algorithm, what is effectively being done is the following: At the 3000-th iteration of the MBDD algorithm, the BDD algorithm is restarted alongside the MBDD algorithm, using as estimates of the model parameters in the initiation vector for the restarted algorithm the estimates of these parameters obtained from the MBDD algorithm at the 3000-th iteration. Then, at some well-defined iteration after the 3000-th iteration (and for all iterations thereafter), the estimates of A and K obtained from the BDD algorithm are sometimes used instead of those obtained from the MBDD algorithm. The estimates of the parameters obtained from the restarted algorithm are used whenever it is expected that these estimates will, on the average, and within a moderate sample size, provide better estimates of the parameters than those offered by the MBDD algorithm.

The proposed modification to the MBDD algorithm should yield an algorithm which provides a global estimator of the Class A parameters for all parameter vectors in the parameter set of interest. The steps of this modified MBDD algorithm will now be given.

4.3.2. Global decision-directed algorithm

Step 1'': *Choose the initiation vectors*. Proceed as in Step 1' of the MBDD algorithm. In addition, choose $\tilde{\tilde{\pi}}_1(0)$, $\tilde{\tilde{\sigma}}_B^2(0)$, and $\tilde{\tilde{\pi_1\sigma}}_I^2(0)$ arbitrarily.

Step 2'': *Classify sample as impulsive or background*. Proceed as in Step 2' of the MBDD algorithm.

Step 3'' : *Update recursively the appropriate update parameters* .

Let

$$\mu = \begin{cases} 1 & \text{if } \tilde{A}_{3000} > 0.1 \\ 0 & \text{if } \tilde{A}_{3000} \leqslant 0.1 \end{cases} ,$$

$$\eta = \begin{cases} 1 & \text{if } \tilde{K}_{3000} > 10^{-4} \\ 0 & \text{if } \tilde{K}_{3000} \leqslant 10^{-4} \end{cases} ,$$

and

$$\delta = \begin{cases} 500 & \text{if } 0.1 < \tilde{A}_{3000} < 0.2 \\ 400 & \text{if } 0.2 \leqslant \tilde{A}_{3000} < 0.3 \\ 300 & \text{if } 0.3 \leqslant \tilde{A}_{3000} < 0.4 \\ 200 & \text{if } 0.4 \leqslant \tilde{A}_{3000} < 0.5 \\ 100 & \text{if } 0.5 \leqslant \tilde{A}_{3000} \leqslant 1.1 \end{cases} .$$

Furthermore, let n^{**} denote the minimum value of n , $n > 3000$, for which $\epsilon_n \neq 0$, where

$$\epsilon_n \triangleq \Big(\sum_{l=3001}^{n} \phi_l \Big) \Big(\sum_{l=3001}^{n} (1-\phi_l) \Big), \qquad n > 3000 .$$

(a) For $n \leqslant 1000$, proceed as in Step 3'(a) of the MBDD algorithm.

(b) For $1000 < n \leqslant 3000$, proceed as in Step 3'(b) of the MBDD algorithm.

(c) For $n > 3000$,

Case (i): $\mu = 0$.

Continue to recursively update $\tilde{\pi}_1$, $\tilde{\sigma}_B^2$, and $\widetilde{\pi_1 \sigma_I^2}$ using (4.7a), (4.7b), and (4.7c), respectively.

Case (*ii*): $\mu = 1$ and $\eta = 0$.

(1) For $3000 < n \leqslant max(\,3000 + \delta\,,\, n^{**} - 1\,)$, continue to recursively update $\tilde{\pi}_1$, $\tilde{\sigma}_B^2$, and $\widetilde{\pi_1\sigma_I^2}$ using (4.7a), (4.7b), and (4.7c), respectively. In addition, update recursively the estimate $\widetilde{\widetilde{\pi}}_1$ as follows :

$$\widetilde{\widetilde{\pi}}_1(n-3000) = \widetilde{\widetilde{\pi}}_1(n-3001) + \frac{1}{n-3000}\left(\phi_n - \widetilde{\widetilde{\pi}}_1(n-3001)\right) . \quad (4.22)$$

(2) For $n > max(\,3000 + \delta\,,\, n^{**} - 1\,)$, continue to recursively update the estimates $\tilde{\pi}_1$, $\tilde{\sigma}_B^2$, $\widetilde{\pi_1\sigma_I^2}$, and $\widetilde{\widetilde{\pi}}_1$ using (4.7a), (4.7b), (4.7c), and (4.22), respectively.

Case (*iii*): $\mu = 1$ and $\eta = 1$.

(1) For $3000 < n \leqslant max(\,3000 + \delta\,,\, n^{**} - 1\,)$, continue to recursively update $\tilde{\pi}_1$, $\tilde{\sigma}_B^2$, and $\widetilde{\pi_1\sigma_I^2}$ using (4.7a), (4.7b), and (4.7c), respectively. In addition, update $\widetilde{\widetilde{\pi}}_1$ using (4.22) and update recursively the estimates $\widetilde{\widetilde{\sigma}}_B^2$ and $\widetilde{\widetilde{\pi_1\sigma_I^2}}$ as follows :

$$\widetilde{\widetilde{\sigma}}_B^2(n-3000) = \begin{cases} \widetilde{\widetilde{\sigma}}_B^2(n-3001) + \dfrac{(1-\phi_n)}{\sum_{l=3001}^{n}(1-\phi_l)}\left(Z_n^2 - \widetilde{\widetilde{\sigma}}_B^2(n-3001)\right) & \text{if } \sum_{l=3001}^{n}(1-\phi_l) \neq 0 \\ \widetilde{\widetilde{\sigma}}_B^2(n-3001) & \text{if } \sum_{l=3001}^{n}(1-\phi_l) = 0 \end{cases} \quad (4.23)$$

and

$$\widetilde{\widetilde{\pi_1\sigma_I^2}}(n-3000) = \widetilde{\widetilde{\pi_1\sigma_I^2}}(n-3001) + \frac{1}{n-3000}\left(\phi_n Z_n^2 - \widetilde{\widetilde{\pi_1\sigma_I^2}}(n-3001)\right) . \quad (4.24)$$

(2) For $n > max(\,3000 + \delta\,,\, n^{**} - 1\,)$, continue to recursively update the estimates $\widetilde{\widetilde{\pi}}_1$, $\widetilde{\widetilde{\sigma}}_B^2$, and $\widetilde{\widetilde{\pi_1\sigma_I^2}}$ using (4.22), (4.23), and (4.24), respectively.

Step $4''$: *Obtain estimates of the Class A parameters.*

Let α_n and n^* be defined as in Step 4 of the BDD algorithm.

(a) For $n < n^*$, proceed as in Step $4'$(a) of the MBDD algorithm.

(b) If $n^* \leqslant 1000$, then for $n^* \leqslant n \leqslant 1000$, proceed as in Step $4'$(b) of the MBDD algorithm.

(c) If $n^* \leqslant 3000$, then for $max(1000, n^* - 1) < n \leqslant 3000$, proceed as in Step $4'$(c) of the MBDD algorithm.

(d) *Case* (i): $\mu = 0$.

For $n > max(3000, n^* - 1)$, proceed as in Step $4'$(c) of the MBDD algorithm.

Case (ii): $\mu = 1$ and $\eta = 0$.

(1) If $n^* \leqslant max(3000 + \delta, n^{**} - 1)$, then for $max(3000, n^* - 1) < n \leqslant max(3000 + \delta, n^{**} - 1)$, proceed as in Step $4'$(c) of the MBDD algorithm.

(2) For $n > max(3000 + \delta, n^{**} - 1)$, obtain estimates of K and σ^2 using (4.8b) and (4.18), respectively. Then, using the relation given in (4.5a), obtain the estimate of A as follows :

$$\tilde{A}_n = -\ln(1 - \tilde{\tilde{\pi}}_1(n - 3000)). \tag{4.25}$$

Case (iii): $\mu = 1$ and $\eta = 1$.

(1) If $n^* \leqslant max(3000 + \delta, n^{**} - 1)$, then for $max(3000, n^* - 1) < n \leqslant max(3000 + \delta, n^{**} - 1)$, proceed as in Step $4'$(c) of the MBDD algorithm.

(2) For $n > max(3000 + \delta, n^{**} - 1)$, obtain estimates of A and σ^2 using (4.25) and (4.18), respectively. Then, using the relation for K given in (4.5b), obtain the estimate of K as follows :

$$\tilde{K}_n = -\ln(1-\tilde{\tilde{\pi}}_1(n-3000)) \left| \frac{\tilde{\tilde{\sigma}}_B^2(n-3000)}{\tilde{\tilde{\pi}}_1\sigma_I^2(n-3000) - \tilde{\tilde{\pi}}_1(n-3000)\,\tilde{\tilde{\sigma}}_B^2(n-3000)} \right| . \quad (4.26)$$

Step 5'' : *Constrain estimates of Class A parameters to lie in parameter set of interest .*

Define Λ^*, β_n^1, β_n^2, β_n^3, and β_n^4 as in Step 5 of the BDD algorithm.

(a) For $n \leq 1000$, proceed as in Step 5'(a) of the MBDD algorithm.

(b) For $1000 < n \leq 3000$, proceed as in Step 5 of the BDD algorithm.

(c) For $n > 3000$,

Case (i): $\mu = 0$.

Proceed as in Step 5 of the BDD algorithm.

Case (ii): $\mu = 1$ and $\eta = 0$.

(1) For $3000 < n \leq max(3000+\delta, n^{**}-1)$, proceed as in Step 5 of the BDD algorithm.

(2) For $n > max(3000+\delta, n^{**}-1)$, constrain $\tilde{A}_n$, $\tilde{K}_n$, and $\tilde{\sigma}_n^2$ using (4.9a), (4.9b), and (4.9c), respectively. (Let $\bar{A}_n$, $\bar{K}_n$, and $\bar{\sigma}_n^2$ denote the constrained estimates.) If $(\bar{A}_n, \bar{K}_n, \bar{\sigma}_n^2) = (\tilde{A}_n, \tilde{K}_n, \tilde{\sigma}_n^2)$, then execution of this step is complete. However, if $(\bar{A}_n, \bar{K}_n, \bar{\sigma}_n^2) \neq (\tilde{A}_n, \tilde{K}_n, \tilde{\sigma}_n^2)$, then proceed as follows : Modify $\tilde{\pi}_1(n)$, $\tilde{\sigma}_B^2(n)$, and $\widetilde{\pi_1\sigma_I^2}(n)$ using (4.10a), (4.10b), and (4.10c), respectively. (Let $\bar{\pi}_1(n)$, $\bar{\sigma}_B^2(n)$, and $\overline{\pi_1\sigma_I^2}(n)$ denote the modified estimates.) In addition, using the relation given in (4.4a), modify $\tilde{\tilde{\pi}}_1(n-3000)$ as follows (let $\bar{\tilde{\pi}}_1(n-3000)$ denote the modified estimate):

$$\bar{\tilde{\pi}}_1(n-3000) = 1 - e^{-\bar{A}_n} .$$

Finally, set $(\tilde{\pi}_1(n), \tilde{\sigma}_B^2(n), \widetilde{\pi_1\sigma_I^2}(n)) = (\bar{\pi}_1(n), \bar{\sigma}_B^2(n), \overline{\pi_1\sigma_I^2}(n))$, $\tilde{\tilde{\pi}}_1(n-3000) = \bar{\tilde{\pi}}_1(n-3000)$, and $(\tilde{A}_n, \tilde{K}_n, \tilde{\sigma}_n^2) = (\bar{A}_n, \bar{K}_n, \bar{\sigma}_n^2)$.

Case (iii): $\mu = 1$ and $\eta = 1$.

(1) For $3000 < n \leq max(3000 + \delta, n^{**} - 1)$, proceed as in Step 5 of the BDD algorithm.

(2) For $n > max(3000 + \delta, n^{**} - 1)$, constrain $\tilde{A}_n$, $\tilde{K}_n$, and $\tilde{\sigma}_n^2$ using (4.9a), (4.9b), and (4.9c), respectively. (let $\bar{A}_n$, $\bar{K}_n$, and $\bar{\sigma}_n^2$ denote the constrained estimates) If $(\bar{A}_n, \bar{K}_n, \bar{\sigma}_n^2) = (\tilde{A}_n, \tilde{K}_n, \tilde{\sigma}_n^2)$, then execution of this step is complete. However, if $(\bar{A}_n, \bar{K}_n, \bar{\sigma}_n^2) \neq (\tilde{A}_n, \tilde{K}_n, \tilde{\sigma}_n^2)$, then proceed as follows : Using the relations given in (4.4a)-(4.4c), modify $\tilde{\tilde{\pi}}_1(n - 3000)$, $\tilde{\tilde{\sigma}}_B^2(n - 3000)$, and $\tilde{\tilde{\pi_1\sigma}}_I^2(n - 3000)$ (let $\bar{\bar{\pi}}_1(n - 3000)$, $\bar{\bar{\sigma}}_B^2(n - 3000)$, and $\bar{\bar{\pi_1\sigma}}_I^2(n - 3000)$ denote the modified estimates) :

$$\bar{\bar{\pi}}_1(n - 3000) = 1 - e^{-\bar{A}_n},$$

$$\bar{\bar{\sigma}}_B^2(n - 3000) = \bar{\sigma}_n^2 \left[\frac{\bar{K}_n}{\bar{A}_n + \bar{K}_n} \right],$$

and

$$\bar{\bar{\pi_1\sigma}}_I^2(n - 3000) = \bar{\sigma}_n^2 \left[\frac{\bar{A}_n + \bar{\bar{\pi}}_1(n - 3000)\bar{K}_n}{\bar{A}_n + \bar{K}_n} \right].$$

Finally, set $(\tilde{\tilde{\pi}}_1(n - 3000), \tilde{\tilde{\sigma}}_B^2(n - 3000), \tilde{\tilde{\pi_1\sigma}}_I^2(n - 3000)) = (\bar{\bar{\pi}}_1(n - 3000), \bar{\bar{\sigma}}_B^2(n - 3000), \bar{\bar{\pi_1\sigma}}_I^2(n - 3000))$, and $(\tilde{A}_n, \tilde{K}_n, \tilde{\sigma}_n^2) = (\bar{A}_n, \bar{K}_n, \bar{\sigma}_n^2)$.

The steps of the global decision-directed algorithm are now complete. In the following section, the performance of the proposed algorithm will be examined via simulation.

4.4. Simulation Results

In order to assess the performance of the proposed global decision-directed algorithm, an extensive simulation study of the algorithm was made. First, the sample mean-square relative error incurred by the initial estimate $\tilde{\sigma}_0^2$ of σ^2 was computed for all (A, K, σ^2) in Ω (defined in

Appendix A) using 100 data sets. The results are tabulated in Table 4.1. It is evident from the values of the relative errors cited in this table that the sample second moment based on 10000 samples provides highly accurate estimates of σ^2. Furthermore, since the estimate of σ^2 at each iteration of the algorithm is obtained by simply updating this sample second moment, the proposed global scheme provides, on the average, highly accurate estimates of σ^2 at each iteration. Consequently, in assessing the ability of this scheme to estimate the remaining two parameters A and K, the parameter σ^2 was assumed to be known.

Using 100 data sets, each containing a sequence of 5000 observations randomly generated from the Class A envelope pdf, the 1%-trimmed mean relative bias of the estimate of A and the 1%-trimmed mean relative bias of the estimate of K were computed for each $(A, K, \sigma^2) \in \Omega$. Let b_A and b_K denote these quantities, respectively, and let δ_j^A and δ_j^K denote the relative errors in the estimates of A and K, respectively, obtained using the j-th data set, i.e.,

$$\delta_j^A \triangleq \frac{\tilde{A}^j - A}{A} \quad \text{and} \quad \delta_j^K \triangleq \frac{\tilde{K}^j - K}{K},$$

where $(\tilde{A}^j, \tilde{K}^j)$ is the estimate of (A, K) obtained using the j-th data set. Then,

Table 4.1. MEAN-SQUARE RELATIVE ERROR OF $\tilde{\sigma}_0^2$ ($\sigma^2 = 1.0$)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	4.7266×10^{-3}	1.5638×10^{-2}	1.8560×10^{-2}	1.8897×10^{-2}	1.8931×10^{-2}
10^{-1}	1.5018×10^{-3}	1.7473×10^{-3}	1.7755×10^{-3}	1.7784×10^{-3}	1.7786×10^{-3}
1	3.2742×10^{-4}	3.3107×10^{-4}	3.3144×10^{-4}	3.3148×10^{-4}	3.3148×10^{-4}

$$b_A \triangleq \frac{1}{98}\left|\left(\sum_{j=1}^{100}\delta_j^A\right) - (\max_j \delta_j^A) - (\min_j \delta_j^A)\right|$$

and

$$b_K \triangleq \frac{1}{98}\left|\left(\sum_{j=1}^{100}\delta_j^K\right) - (\max_j \delta_j^K) - (\min_j \delta_j^K)\right| .$$

The values for these biases are tabulated in Tables 4.2 and 4.3. Note from these tables that the values for b_A and b_K are quite low for all parameter vectors under consideration. Thus, the proposed global decision-directed algorithm yields an essentially unbiased recursive estimator of the parameters A and K of the Class A model.

In addition to the relative biases, the following quantities were computed for each $(A, K, \sigma^2) \in \Omega$ using the aforementioned data sets:

(i) the 1%-trimmed sample mean-square relative error due to estimating A,

(ii) the 1%-trimmed sample mean-square relative error due to estimating K,

and

(iii) the 1%-trimmed sample mean-square-norm relative error (MSNRE) due to estimating A and K.

Let e_A, e_K, and e_{tot} denote the quantities described in (i), (ii), and (iii), respectively. We note that, for each of these quantities, the "trimming" is based on the exclusion of a data set which yields the highest value *for* e_{tot} and a data set which yields the lowest value *for* e_{tot}, i.e., if

$$j^* \triangleq \arg\max_j \left|(\delta_j^A)^2 + (\delta_j^K)^2\right|$$

and

$$j^{**} \triangleq \arg\min_j \left|(\delta_j^A)^2 + (\delta_j^K)^2\right| ,$$

Table 4.2. 1% - TRIMMED RELATIVE BIAS OF ESTIMATE OF A (b_A)
($\sigma^2 = 1.0$)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	3.8241×10^{-2}	6.0110×10^{-2}	6.3540×10^{-2}	6.4057×10^{-2}	6.3781×10^{-2}
10^{-1}	-9.7836×10^{-2}	-5.0248×10^{-2}	-4.2653×10^{-2}	-4.1699×10^{-2}	-4.1542×10^{-2}
1	-4.5694×10^{-2}	-8.8028×10^{-3}	-1.6972×10^{-3}	-8.5670×10^{-4}	-7.3438×10^{-4}

Table 4.3. 1% - TRIMMED RELATIVE BIAS OF ESTIMATE OF K (b_K)
($\sigma^2 = 1.0$)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	-6.4449×10^{-2}	4.6075×10^{-2}	5.2917×10^{-2}	5.3716×10^{-2}	1.1140×10^{-1}
10^{-1}	-3.5629×10^{-2}	3.1080×10^{-2}	3.0820×10^{-2}	3.4567×10^{-2}	4.8787×10^{-2}
1	-3.1074×10^{-2}	1.4006×10^{-2}	-9.2976×10^{-3}	-1.6823×10^{-3}	2.0792×10^{-2}

then

$$e_A = \frac{1}{98}\left[\left(\sum_{j=1}^{100}(\delta_j^A)^2\right) - (\delta_{j^{**}}^A)^2 - (\delta_{j^*}^A)^2\right],$$

$$e_K = \frac{1}{98}\left[\left(\sum_{j=1}^{100}(\delta_j^K)^2\right) - (\delta_{j^{**}}^K)^2 - (\delta_{j^*}^K)^2\right],$$

and

$$e_{tot} = e_A + e_K.$$

(For the computations performed here, j^* and j^{**} were uniquely determined. However, the defining properties for these two quantities does not guarantee this. If either j^* or j^{**} is not uniquely determined, then these indices are chosen arbitrarily among those that satisfy the corresponding defining property.) The computed values for e_A, e_K, and e_{tot} are tabulated in Tables 4.4, 4.5, and 4.6, respectively. As with the relative biases, it is seen from Table 4.6 that the MSNRE is quite low for all parameter vectors under consideration. Moreover, from a comparison of the values given in Tables 4.4 and 4.5, it is evident that neither e_A nor e_K dominates in its contribution to e_{tot}. Thus, in a mean-square error sense, not only does the proposed estimator provide a very good global estimator of both parameters A and K, but, in addition, it has no difficulty in estimating one parameter over the other.

Lastly, the normalized sample MSNRE ($\triangleq 5000 \times e_{tot}$) was computed (see Table 4.7) and compared to the Cramer-Rao Lower Bound (CRLB). (The values for the CRLB were given in Table 3.8.) Note that the values for the normalized sample MSNRE and CRLB are essentially on the same order of magnitude. In fact, for $A = 10^{-2}$, the values for the normalized sample MSNRE are quite close to those of the CRLB. Furthermore, a comparison of the computed values for the normalized sample MSNRE of the proposed recursive estimator and the normalized sample MSNRE of the batch scheme upon which it is based (see Table 3.9) indicates that the proposed recursive scheme performs better than the batch estimator for a sizeable subset of the parameter set under consideration.

Table 4.4. 1% - TRIMMED MEAN-SQUARE RELATIVE ERROR DUE TO ESTIMATING A (e_A)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	1.0896×10^{-2}	1.4162×10^{-2}	1.4684×10^{-2}	1.4761×10^{-2}	1.4702×10^{-2}
10^{-1}	1.3786×10^{-2}	6.4035×10^{-3}	6.0285×10^{-3}	5.8759×10^{-3}	5.7837×10^{-3}
1	3.1014×10^{-3}	1.1524×10^{-3}	1.0628×10^{-3}	1.0558×10^{-3}	1.0589×10^{-3}

Table 4.5. 1% - TRIMMED MEAN-SQUARE RELATIVE ERROR DUE TO ESTIMATING K (e_K)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	1.2353×10^{-2}	2.1480×10^{-2}	2.3808×10^{-2}	2.3974×10^{-2}	2.8163×10^{-2}
10^{-1}	4.1013×10^{-3}	8.2767×10^{-3}	7.4102×10^{-3}	6.3852×10^{-3}	7.5959×10^{-3}
1	3.8256×10^{-3}	3.3176×10^{-3}	2.1203×10^{-3}	1.7420×10^{-3}	2.3851×10^{-3}

Table 4.6. 1% - TRIMMED SAMPLE MSNRE (e_{tot})

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	2.3249×10^{-2}	3.5643×10^{-2}	3.8492×10^{-2}	3.8735×10^{-2}	4.2865×10^{-2}
10^{-1}	1.7888×10^{-2}	1.4680×10^{-2}	1.3439×10^{-2}	1.2261×10^{-2}	1.3380×10^{-2}
1	6.9270×10^{-3}	4.4699×10^{-3}	3.1831×10^{-3}	2.7978×10^{-3}	3.4439×10^{-3}

Table 4.7. NORMALIZED SAMPLE MSNRE ($\triangleq 5000 \times e_{tot}$)
FOR $(A, K, \sigma^2) \in \Omega$

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	1.1624×10^{2}	1.7821×10^{2}	1.9246×10^{2}	1.9367×10^{2}	2.1433×10^{2}
10^{-1}	8.9438×10^{1}	7.3401×10^{1}	6.7194×10^{1}	6.1306×10^{1}	6.6898×10^{1}
1	3.4635×10^{1}	2.2350×10^{1}	1.5916×10^{1}	1.3989×10^{1}	1.7220×10^{1}

In summary then, we see that the proposed global decision-directed algorithm does in fact yield a global estimator of the parameters which performs very well for all parameter vectors in the parameter set of interest, even for moderate sample sizes. In view of this performance, this algorithm provides an effective recursive estimator of the Class A model parameters.

4.5. Conclusions

In this chapter, we have developed a global recursive estimator of the parameters of the strictly canonical Class A model which performs very well for all parameter vectors in the parameter set of interest. The starting point in the study was the development of the so-called BDD algorithm. This basic, decision-directed, adaptive scheme is physically motivated, easy to implement, and is a recursive version of a batch procedure which was seen in our earlier work to provide good estimates of the parameters. Unfortunately, examination of the performance of this algorithm via simulation reveals two inherent drawbacks of the scheme that adversely affect its performance even in a local setting. However, it is seen that by placing certain restrictions on the form of the initiation vector for the algorithm and by incorporating the appropriate modifications into its framework, the ensuing difficulties associated with the two basic shortcomings can be eliminated, and an improvement in the performance of the algorithm can be attained globally. Examination of the performance characteristics of the modified algorithm via simulation indicates that this algorithm does, in fact, yield an effective global recursive estimator of the Class A model parameters. Although this final algorithm is somewhat complex, the payoff for this complexity is excellent global performance.

5. EFFICIENT ESTIMATION FOR SMALL SAMPLE SIZES : THE EM ALGORITHM

5.1. Introduction

In the previous chapters, we have obtained several batch and recursive estimators of the Class *A* parameters that yield good estimates of the parameters for moderate sample sizes. In this chapter, we will focus our attention on the problem of obtaining a batch estimator of these parameters with good *small*-sample-size performance for all parameter vectors in the parameter set of interest.

One estimator that has the potential of providing estimates of the Class A parameters with the above-mentioned property is the EM algorithm. This algorithm, which is ideally suited for estimation problems where the observations can be viewed as "incomplete data," was popularized by Dempster, Laird, and Rubin in 1977 [17]. We begin this chapter with a description of this algorithm. We then examine the behavior of the EM algorithm within the Class A framework. In particular, for the single-parameter estimation problem, a closed-form expression for the estimator is obtained first. Several properties of the estimator are also derived. Using these properties, it is shown that the sequence of estimates obtained via the EM algorithm converges and a characterization of the nature of the point to which the sequence converges is given. An implementation of the estimator based on the execution of two EM algorithms in parallel is then described. Using this implementation, the small-sample-size performance of the proposed estimator is assessed via an extensive simulation study. The results of this study indicate that the proposed EM estimator yields excellent estimates of the parameter for small sample sizes.

The two-parameter estimation problem is then examined. For the two-parameter estimation problem, a description of the procedure through which estimates of the parameter are obtained is given first. Furthermore, using an implementation analogous to the one used for the single-parameter estimation problem, the small-sample-size performance of the proposed EM estimator is also assessed via an extensive simulation study. Again, as for the single-parameter estimation

problem, this study reveals that the EM algorithm yields an excellent estimator of the Class A parameters for small sample sizes.

5.2. The EM Algorithm

A commonly-used two-step iterative technique for estimating the parameters of a density when the observations can be viewed as "incomplete-data" is the EM algorithm [17]. Mixture densities, such as the Class A density, can be placed naturally in this framework. In particular, for such densities, the "incomplete-data" set is the set of observations, whereas each element of the "complete-data" set can be defined to be a two-component vector consisting of an observation and an indicator specifying which component of the mixture occurred during that observation. Instead of using the traditional incomplete-data density in the estimation process, the EM algorithm uses the properties of the complete-data density. In so doing, it can often make the estimation problem more tractable and also yield good estimates of the parameters for small sample sizes.

Let $\underline{\theta}$ denote the parameter vector to be estimated, $\underline{\theta}^{(p)}$ the estimate of $\underline{\theta}$ obtained at the p-th iteration of the algorithm, $\underline{z}$ the incomplete-data set (set of observations), $\underline{x}$ the complete-data set, and g the likelihood function associated with $\underline{x}$. The two steps, the *expectation* step (E-step) and *maximization* step (M-step), of the EM algorithm are then given as follows :

E-step: Evaluate $Q(\underline{\theta} \mid \underline{\theta}^{(p)}) \triangleq E[\log g(\underline{x} \mid \underline{\theta}) \mid \underline{z}, \underline{\theta}^{(p)}]$.
M-step: Determine $\underline{\theta} = \underline{\theta}^{(p+1)}$ to maximize $Q(\underline{\theta} \mid \underline{\theta}^{(p)})$.

The EM algorithm can be viewed as an alternative to maximizing the function g over $\underline{\theta}$. In particular, since g is unknown, we instead maximize its expectation given the available pertinent information, namely the observed data and the current estimate of the parameters.

Consider the following basic property of the EM algorithm : Let L denote the likelihood function associated with $\underline{z}$, i.e., the incomplete-data likelihood function. If the function g is positive almost everywhere on its domain, then it can be shown via a simple application of Jensen's Inequality that

$$L(\underline{\theta}^{(p+1)}) \geq L(\underline{\theta}^{(p)}), \qquad p = 0, 1, \ldots, \tag{5.1}$$

i.e., the likelihoods of *interest* are monotone increasing in $\underline{\theta}^{(p)}$. Given this desirable property of the EM algorithm and its estimation potential for small sample sizes, let us now consider the problem of estimating the parameters of the Class A model via the EM algorithm.

5.3. Estimation of Class A Parameters

Let $\underline{z} \triangleq \{z_1, \ldots, z_n\}$ denote the incomplete-data set consisting of n i.i.d. observations generated from the Class A envelope pdf w, with unknown parameter vector $\underline{\theta} \triangleq (A, K)^T$ to be estimated. With each observation z_i, we can associate an unobserved infinite-dimensional indicator vector $\mathbf{v}_i = (v_{ij}, j = 1, 2, \ldots)^T$, whose entries are all zero except for one equal to unity in the position corresponding to the unobserved component of the Class A envelope mixture density associated with z_i. Thus, let $\underline{x} \triangleq \{(z_i, \mathbf{v}_i); i = 1, \ldots, n\}$ denote the complete-data set and g denote the likelihood function associated with $\underline{x}$. Under the assumption that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are i.i.d. and that the z_i given $\mathbf{v}_i$ are conditionally independent and identically distributed, we have that

$$g(\underline{x}|\underline{\theta}) = \prod_{i=1}^{n} \prod_{j=1}^{\infty} \pi_j(A)^{v_{ij}} h_j(z_i; \underline{\theta})^{v_{ij}},$$

where

$$\pi_j(A) = \frac{e^{-A} A^{j-1}}{(j-1)!}$$

and

$$h_j(z_i, \underline{\theta}) = 2z_i \left\{ \frac{A+K}{j-1+K} \right\} e^{-z_i^2 \left\{ \frac{A+K}{j-1+K} \right\}}.$$

The corresponding logarithm is given by

$$L_0(\underline{\theta}) = \sum_{i=1}^{n} \langle v_i, V(\underline{\theta}) \rangle + \sum_{i=1}^{n} \langle v_i, U_i(\underline{\theta}) \rangle \tag{5.2}$$

where $V(\underline{\theta})$ and $U_i(\underline{\theta})$ are infinite-dimensional vectors with the j-th component $\ln \pi_j$ and

$\ln h_j(z_i;\underline{\theta})$, respectively, and $<\ldots>$ is the standard l_2-inner product. Let $\underline{\theta}^{(p)} \triangleq (A_p, K_p)^T$ $(A_p > 0, K_p \geq 0)$, $p = 0, 1, \ldots$, denote the estimate of $\underline{\theta}$ obtained at the p—th iteration of the EM algorithm. Then, using (5.2) and the definition of the Q function given in the E-step of the EM algorithm, we have that

$$Q(\underline{\theta} \mid \underline{\theta}^{(p)}) = \sum_{i=1}^{n} <a_i, V(\underline{\theta})> + \sum_{i=1}^{n} <a_i, U_i(\underline{\theta})> \quad . \tag{5.3}$$

where

$$a_i \triangleq E(\mathbf{v}_i \mid \underline{z}, \underline{\theta}^{(p)}) = E(\mathbf{v}_i \mid z_i, \underline{\theta}^{(p)}) \tag{5.4}$$

and the latter equality follows from the independence assumptions stated above. Let a_{ij} $(j = 1, 2, \ldots)$ denote the j—th component of a_i. Then it follows from (5.4) and the definition of $\mathbf{v}_i$ that

$$a_{ij} = \frac{\pi_j(A_p)\, h_j(z_i;\underline{\theta}^{(p)})}{\sum_{j=1}^{\infty} \pi_j(A_p)\, h_j(z_i;\underline{\theta}^{(p)})} \quad . \tag{5.5}$$

Now, the second step of the EM algorithm, the M-step, requires that we determine the argument $\underline{\theta} = (A, K)^T$ that maximizes $Q(\underline{\theta} \mid \underline{\theta}^{(p)})$. To make the dependence on A and K more explicit, we can use the definitions of V, U_i, π_j, and h_j to rewrite the expression given in (5.3) in terms of A and K as follows:

$$\begin{aligned} Q(\underline{\theta} \mid \underline{\theta}^{(p)}) = & \sum_{i=1}^{n} \sum_{j=1}^{\infty} a_{ij} \left[-A + (j-1)\ln A - \ln(j-1)! \right] \\ & + \sum_{i=1}^{n} \sum_{j=1}^{\infty} a_{ij} \left[\ln 2 z_i + \ln \frac{(A+K)}{(j-1+K)} - z_i^2 \frac{(A+K)}{(j-1+K)} \right] . \end{aligned} \tag{5.6}$$

Using the fact that $\sum_{j=1}^{\infty} a_{ij} = 1$ and eliminating all terms in (5.6) that are constants with respect to the maximization, we obtain the following objective function $Q_{obj}(\underline{\theta} \mid \underline{\theta}^{(p)})$:

$$Q_{obj}(\underline{\theta}\,|\,\underline{\theta}^{(p)}) = -n\,A + \alpha \ln A + n \ln(A+K) - \sum_{i=1}^{n}\sum_{j=1}^{\infty} a_{ij} \ln(j-1+K) - (A+K)\sum_{i=1}^{n}\sum_{j=1}^{\infty} z_i^{\,2}\,\frac{a_{ij}}{(j-1+K)}, \tag{5.7}$$

where $\alpha \triangleq \sum_{i=1}^{n}\sum_{j=1}^{\infty}(j-1)\,a_{ij}$. The M-step then yields as the estimate of the true parameter vector at the $p+1$-st iteration of the algorithm a maximizing argument of our objective function, i.e.,

$$\underline{\theta}^{(p+1)} = \arg\max_{\underline{\theta}} Q_{obj}(\underline{\theta}\,|\,\underline{\theta}^{(p)}). \tag{5.8}$$

With regard to this maximization, let us address the single-parameter estimation problem first.

5.4. Single-Parameter Estimation Problem

Consider the single-parameter estimation problem wherein the parameter A is unknown and the parameter K is known. Let Q_K denote the objective function to be maximized in the M-step of the EM algorithm for this single-parameter estimation problem. Then, upon fixing K_p to its known value K in (5.7), and thereafter eliminating all terms in (5.7) that are constants with respect to the one-dimensional maximization, we obtain the following expression for Q_K :

$$Q_K(A\,|\,A_p) = -n\,A + \beta \ln A + n \ln(A+K) - \phi A, \tag{5.9}$$

where $\beta \triangleq \sum_{i=1}^{n}\sum_{j=1}^{\infty}(j-1)\,b_{ij}$, $\phi \triangleq \sum_{i=1}^{n}\sum_{j=1}^{\infty} z_i^{\,2}(b_{ij}/(j-1+K))$, and $b_{ij} \triangleq a_{ij}\big|_{K_p = K}$. Now, recall from the definition of the parameter set Λ of interest, that A takes on values in the interval $\Lambda_A \triangleq [10^{-2}, 1]$. Define Λ_A^* to be an extension of this interval:

$$\Lambda_A^* \triangleq \{A \,|\, (10^{-2}/(1+\epsilon)) \leqslant A \leqslant 1+\epsilon\}, \qquad \epsilon > 0, \tag{5.10}$$

where ϵ is chosen arbitrarily[1], and let Λ_A^* be the compact set over which the maximization is performed in the M-step of the EM algorithm, i.e.,

[1]We consider an ϵ-extension of Λ_A instead of Λ_A itself since, for the boundary points 10^{-2} and 1 of Λ_A, it is only reasonable that we admit estimates of these parameter values that lie in a neighborhood of the true values (see (5.14)).

$$A_{p+1} = \arg \max_{A \in \Lambda_A^*} Q_K(A \mid A_p), \qquad A_0 \in \Lambda_A^* . \tag{5.11}$$

This maximization problem can be readily solved by noting the following property:

Basic Property :For each $p \in 0, 1, \ldots, Q_K(A \mid A_p)$ is concave in A on $(0, \infty)$ (for all $K > 0$).

Proof: From (5.9), we have that

$$\frac{\partial^2}{\partial A^2} Q_K(A \mid A_p) = -\frac{\beta}{A^2} - \frac{n}{(A+K)^2} .$$

Since $A_p \in \Lambda_A^*$ for all p, it follows that $A_p > 0$ for all p, which implies from the definition of β that $\beta > 0$. Thus,

$$\frac{\partial^2}{\partial A^2} Q_K(A \mid A_p) < 0 \qquad \text{for all } p .$$

□

It follows from this basic property that, if Q_K attains its maximum at an interior point of $(0,\infty)$, then a necessary and sufficient condition for $A = A_{max}$ to be the maximizing root on this interval is that

$$\frac{\partial}{\partial A} Q_K(A \mid A_p)\Big|_{A = A_{max}} = 0 . \tag{5.12}$$

Evaluation of this gradient equation yields the following quadratic equation :

$$a_1 A^2 + a_2 A + a_3 \Big|_{A = A_{max}} = 0 ,$$

where $a_1 := -n - \phi$, $a_2 := a_1 K + \beta + n$, and $a_3 := \beta K$. Note that a_1, a_2, and a_3 are functions only of $\underline{z}$, A_p, and K. Now, it follows from the definition of ϕ that ϕ is nonnegative, which implies that $a_1 < 0$. Furthermore, since $\beta > 0$ (see proof of Basic Property) and $K > 0$, it also follows that $a_3 > 0$. Given that a_1 and a_3 are negative and positive, respectively, and the requirement that A_{max} be positive, one of the roots to the quadratic equation can be disregarded

immediately, and we obtain the following closed-form expression for $A_{\max}$ in terms of the observations, the known parameter K, and A_p:

$$A_{\max} = \frac{a_1(\underline{z}, A_p, K) + [a_2^2(\underline{z}, A_p, K) - 4a_1(\underline{z}, A_p, K)a_3(\underline{z}, A_p, K)]^{1/2}}{-2a_1(\underline{z}, A_p, K)} \quad . \tag{5.13}$$

Since $A_{\max}$ maximizes Q_K on $(0, \infty)$, it follows that if $A_{\max}$ lies in Λ_A^*, then $A_{\max}$ maximizes Q_K on Λ_A^*. However, if $A_{\max}$ lies outside of Λ_A^*, then it follows from the Basic Property of Q_K that the argument which maximizes Q_K on Λ_A^* is simply the boundary point of Λ_A^* nearest $A_{\max}$. In summary then, the solution to (5.11) is given as follows:

$$A_{p+1} = \begin{cases} \dfrac{10^{-2}}{1+\epsilon} & \text{if} \quad A_{\max} < \dfrac{10^{-2}}{1+\epsilon} \\ A_{\max} & \text{if} \quad A_{\max} \in \Lambda_A^* \\ 1+\epsilon & \text{if} \quad A_{\max} > 1+\epsilon \end{cases} \quad ; \tag{5.14}$$

$A_0 \in \Lambda_A^*$. The estimator described by recursion (5.14) will be referred to in the sequel as the *EM estimator of A*.

5.4.1. Properties of EM estimator of A

Let $(0, \infty)$ be the domain of Q_K $(K > 0)$. Note that $\frac{\partial}{\partial A} Q_K(A \mid A_p)$ and $\frac{\partial^2}{\partial A^2} Q_K(A \mid A_p)$ exist for all $A \in (0, \infty)$.

Property 1: For each $p \in \{0, 1, \ldots\}$, there exists $\lambda_p > 0$ such that

$$Q_K(A_{p+1} \mid A_p) - Q_K(A_p \mid A_p) \geq \lambda_p (A_{p+1} - A_p)^2 \quad .$$

Proof. The property clearly holds if $A_p = A_{p+1}$. Suppose $A_p \neq A_{p+1}$. Expand $Q_K(A \mid A_p)$ in a Taylor series about A_{p+1} to obtain

$$Q_K(A \mid A_p) = Q_K(A_{p+1} \mid A_p) + (A - A_{p+1}) \frac{\partial}{\partial A} Q_K(A_{p+1} \mid A_p) + (A - A_{p+1})^2 \frac{\partial^2}{\partial A^2} Q_K(A^*_{p+1} \mid A_p) \tag{5.15}$$

for some A^*_{p+1} satisfying $\min(A_p, A_{p+1}) < A^*_{p+1} < \max(A_p, A_{p+1})$. Evaluation of (5.15) at $A = A_p$ yields :

$$Q_K(A_p \mid A_p) = Q_K(A_{p+1} \mid A_p) + (A_p - A_{p+1}) \frac{\partial}{\partial A} Q_K(A_{p+1} \mid A_p) + (A_p - A_{p+1})^2 \frac{\partial^2}{\partial A^2} Q_K(A^*_{p+1} \mid A_p). \tag{5.16}$$

Now, A_{p+1} can take on one of three values : $(10^{-2}/(1+\epsilon))$, $A_{\max}$, or $1+\epsilon$ (see (5.14)).

(i) Suppose $A_{p+1} = A_{\max}$. Then $\frac{\partial}{\partial A} Q_K(A_{p+1} \mid A_p) = \frac{\partial}{\partial A} Q_K(A_{\max} \mid A_p) = 0$, where the latter inequality follows from the definition of $A_{\max}$. Consequently, $(A_p - A_{p+1}) \frac{\partial}{\partial A} Q_K(A_{p+1} \mid A_p) = 0$.

(ii) Suppose $A_{p+1} = (10^{-2}/(1+\epsilon))$. Then, $A_{p+1} \leqslant A_p$ (since, of course, $A_p \in \Lambda^*_A$ also). Also note that $A_{p+1} = (10^{-2}/(1+\epsilon))$ only when $A_{\max} \leqslant (10^{-2}/(1+\epsilon))$, i.e., $A_{p+1} \geqslant A_{\max}$. Since $A_{p+1} \geqslant A_{\max}$ and $\frac{\partial}{\partial A} Q_K(A_{\max} \mid A_p) = 0$, it follows from the Basic Property of Q_K that $\frac{\partial}{\partial A} Q_K(A_{p+1} \mid A_p) \leqslant 0$. Thus, we have that $A_p - A_{p+1} \geqslant 0$ and $\frac{\partial}{\partial A} Q_K(A_{p+1} \mid A_p) \leqslant 0$, which imply that $(A_p - A_{p+1}) \frac{\partial}{\partial A} Q_K(A_{p+1} \mid A_p) \leqslant 0$.

(iii) Suppose $A_{p+1} = 1+\epsilon$. Then $A_{p+1} \geqslant A_p$. Also note that $A_{p+1} = 1+\epsilon$ only when $A_{\max} \geqslant 1+\epsilon$, i.e., $A_{p+1} \leqslant A_{\max}$. Since $A_{p+1} \leqslant A_{\max}$ and $\frac{\partial}{\partial A} Q_K(A_{\max} \mid A_p) = 0$, it again follows from the Basic Property of Q_K that $\frac{\partial}{\partial A} Q_K(A_{p+1} \mid A_p) \geqslant 0$. Thus, we have that $A_p - A_{p+1} \leqslant 0$ and $\frac{\partial}{\partial A} Q_K(A_{p+1} \mid A_p) \geqslant 0$, which imply that $(A_p - A_{p+1}) \frac{\partial}{\partial A} Q_K(A_{p+1} \mid A_p) \leqslant 0$.

In all three cases, $(A_p - A_{p+1}) \frac{\partial}{\partial A} Q_K(A_{p+1} | A_p) \leq 0$, which together with (5.15) implies that

$$Q_K(A_{p+1} | A_p) - Q_K(A_p | A_p) \geq -(A_{p+1} - A_p)^2 \frac{\partial^2}{\partial A^2} Q_K(A^*_{p+1} | A_p).$$

Moreover, from the concavity of Q_K, we have that $\frac{\partial^2}{\partial A^2} Q_K(A^*_{p+1} | A_p) < 0$. Thus, by taking

$$\lambda_p = -\frac{\partial^2}{\partial A^2} Q_K(A^*_{p+1} | A_p),$$

the statement of the property follows. □

The following property follows straightforwardly from Property 1:

Property 2: For each $K > 0$, there exists a scalar $\lambda > 0$ such that

$$Q_K(A_{p+1} | A_p) - Q_K(A_p | A_p) \geq \lambda (A_{p+1} - A_p)^2 \quad \text{for all } p \in \{0, 1, \ldots\}.$$

Proof We have shown in Property 1 that for each $p \in \{0, 1, \ldots\}$, there exists $\lambda_p > 0$ such that

$$Q_K(A_{p+1} | A_p) - Q_K(A_p | A_p) \geq \lambda_p (A_{p+1} - A_p)^2 \quad . \tag{5.17}$$

where

$$\lambda_p = -\frac{\partial^2}{\partial A^2} Q_K(A^*_{p+1} | A_p) \tag{5.18}$$

when $A_p \neq A_{p+1}$ and $\min(A_p, A_{p+1}) < A^*_{p+1} < \max(A_p, A_{p+1})$. Equations (5.9) and (5.18) yield the following expression for λ_p :

$$\lambda_p = \frac{\beta_p}{(A^*_{p+1})^2} + \frac{n}{(A^*_{p+1} + K)^2} \quad ,$$

where the dependence of β on p is explicitly shown here. Since $\beta_p > 0$ for each p, it follows that

$$\lambda_p > \frac{n}{(A^*_{p+1} + K)^2} \quad .$$

Furthermore, since $\min(A_p, A_{p+1}) < A^*_{p+1} < \max(A_p, A_{p+1})$ and $A_p, A_{p+1} \in \Lambda^*_A$, we have that $A^*_{p+1} \in \Lambda^*_A$, i.e., $A^*_{p+1} \leqslant 1 + \epsilon$. Thus,

$$\lambda_p > \frac{n}{(1 + \epsilon + K)^2} \quad \text{for all } p \quad .$$

which implies that

$$\lambda \triangleq \inf_p \lambda_p \geqslant \frac{n}{(1 + \epsilon + K)^2} > 0 \quad . \tag{5.19}$$

It then readily follows from (5.17) and (5.19) that

$$Q_K(A_{p+1} | A_p) - Q_K(A_p | A_p) \geqslant \lambda (A_{p+1} - A_p)^2 \quad \text{for all } p \quad .$$

Thus, we have the desired result. □

We are now in a position to prove the following convergence theorem. Let L_K denote the incomplete-data log-likelihood function for known K $(K > 0)$, i.e., the traditional log-likelihood function for the single-parameter estimation problem :

$$\begin{aligned} L_K(A) &= \sum_{i=1}^{n} \ln \sum_{j=1}^{\infty} \pi_j(A)\, h_j(z_i; A, K) \qquad , A \in \Lambda^*_A \quad . \\ &= \sum_{i=1}^{n} \ln \sum_{j=1}^{\infty} \frac{2 e^{-A} A^{j-1}}{(j-1)!} \left(\frac{A+K}{j-1+K} \right) z_i \, e^{-z_i^2 \left| \frac{A+K}{j-1+K} \right|} , \quad A \in \Lambda^*_A \quad . \end{aligned} \tag{5.20}$$

Note that L_K is continuous on Λ^*_A and differentiable in the interior of Λ^*_A. Since L_K is a continuous function on the compact set Λ^*_A, it follows that L_K is bounded. Furthermore, from (5.1), we have that $L_K(A_{p+1}) \geqslant L_K(A_p)$ for all $p \in \{0, 1, \ldots\}$. Thus, since $\{L_K(A_p)\}$ is a monotone bounded sequence, $\{L_K(A_p)\}$ converges monotonically to some L^*. Now, let

$\underline{\underline{L}}(\beta) \triangleq \{A \in \Lambda_A^* : L_K(A) = \beta\}$. The statement of the theorem is then given as follows:

Convergence Theorem : Let L_K and L^* be defined as above. Then $A_p \to A^* \in \Lambda_A^*$ and $L_K(A_p) \to L^* = L_K(A^*)$. Furthermore, if A^* is an interior point of Λ_A^*, then $\left.\frac{\partial L_K}{\partial A}\right|_{A=A^*} = 0$, i.e., A^* is a stationary point of L_K, the likelihood function of interest.

Proof: It is evident that if $A_p \to A^*$, then $L^* = L_K(A^*)$. Now, in order to show that $A_p \to A^*$, it suffices to show that

(i) $|A_{p+1} - A_p| \to 0$ as $p \to \infty$, and

(ii) $\underline{\underline{L}}(L^*)$ is discrete.

That these two conditions are sufficient for the convergence of $\{A_p\}$ follows from Theorem 6 of Wu [18].

Consider condition (i). We have shown in Property 2 that there exists a scalar $\lambda > 0$ such that

$$Q_K(A_{p+1}|A_p) - Q_K(A_p|A_p) \geq \lambda (A_{p+1} - A_p)^2 \quad \text{for all } p \in \{0, 1, \ldots\}. \tag{5.21}$$

From the definitions of L_K and Q_K, it can also be readily shown via a simple application of Jensen's Inequality that

$$L_K(A_{p+1}) - L_K(A_p) \geq Q_K(A_{p+1}|A_p) - Q_K(A_p|A_p) \quad \text{for all } p \in \{0, 1, \ldots\} \tag{5.22}$$

(see proof of Theorem 1 of Dempster, Laird, and Rubin [17]). Thus, it follows from (5.21) and (5.22) that

$$L_K(A_{p+1}) - L_K(A_p) \geq \lambda (A_{p+1} - A_p)^2 \quad \text{for all } p \in \{0, 1, \ldots\}. \tag{5.23}$$

Since $\{L_K(A_p)\}$ converges, we have that $|L_K(A_{p+1}) - L_K(A_p)| = L_K(A_{p+1}) - L_K(A_p) \to 0$ as $p \to \infty$, where the first equality follows from the monotinicity property of $\{L_K(A_p)\}$. It then follows from (5.23) that $|A_{p+1} - A_p| \to 0$ as $p \to \infty$ and the verification of condition (i) is complete.

Consider condition (ii). We need to show that $\{A \in \Lambda_A^* : L_K(A) - L^* = 0\}$ is discrete. It can be readily shown that $L_K(A) - L^*$ is analytic in Λ_A^*. Since it is well known that the zeroes of an analytic function (which is not identically to zero) are isolated, the result follows.

The verification of conditions (i) and (ii) is now complete. Since $A_p \to A^*$ in the closure of Λ_A^* and Λ_A^* is a closed set, it follows that $A^* \in \Lambda_A^*$. We must now show that if A^* is an interior point of Λ_A^*, then A^* is a stationary point of L_K. Again, from Theorem 6 of Wu [18], it suffices to show that

(iii) $\dfrac{\partial}{\partial A} Q_K(A_{p+1} \mid A_p) = 0$ for all p sufficiently large, and

(iv) $\dfrac{\partial}{\partial A} Q_K(A \mid A_p)$ is continuous in (A, A_p) on $\Lambda_A^* \times \Lambda_A^*$.

Consider condition (iii). If A^* is an interior point of Λ_A^*, then there exists some p' such that A_{p+1} is an interior point of Λ_A^* for all $p \geq p'$. Thus, it follows from (5.14) that $A_{p+1} = A_{\max}$ for all $p \geq p'$. (Of course, $A_{\max}$ varies with p.) Since $\dfrac{\partial}{\partial A} Q_K(A_{\max} \mid A_p) = 0$ by definition of $A_{\max}$, we have that $\dfrac{\partial}{\partial A} Q_K(A_{p+1} \mid A_p) = 0$ for all $p \geq p'$. The verification of the condition is complete.

Consider condition (iv). Note that the infinite series in the expression for $Q_K(A \mid A_p)$ given in (5.9) are uniformly convergent series on $\Lambda_A^* \times \Lambda_A^*$. Furthermore, each term of a given series is a continuous function on $\Lambda_A^* \times \Lambda_A^*$. It follows then that each series is a continuous function on this domain. Thus, the expression for $\dfrac{\partial}{\partial A} Q_K(A \mid A_p)$ consists of sums and ratios of continuous functions on $\Lambda_A^* \times \Lambda_A^*$. The result follows.

Since conditions (iii) and (iv) hold, it follows that if A^* is an interior point of Λ_A^*, then A^* is a stationary point of L_K. The proof of the theorem is now complete. □

The particular stationary point to which $\{A_p\}$ converges is dependent upon the initiating point A_0. Thus, given a "sufficiently rich" set of initiating points, then upon executing several EM estimators of A in parallel, each initiated with a different point from this set, all of the stationary points corresponding to relative maxima of the (traditional) likelihood function can be located, and hence, the point(s) corresponding to the absolute maximum. With this in mind, consider the following implementation of the EM estimator of A.

5.4.2. Implementation of EM estimator of A

First, choose $\epsilon = 0.1$ in the definition of Λ_A^* (given in (5.10)). Secondly, let the set of initiating points, Λ_{init}^A, consist of the left boundary point of Λ_A^* and the (logarithmic) mean of the interval defined by Λ_A^*, i.e., $\Lambda_{init}^A = \{9.\overline{09} \times 10^{-3}, 10^{-1}\}$. (That the two points in Λ_{init}^A form a "sufficiently rich" set for the determination of the point(s) corresponding to the absolute maximum of the (traditional) likelihood function was verified experimentally.) For each true parameter A, two EM estimators of A were executed in parallel, one initiated with $9.\overline{09} \times 10^{-3}$, the other with 10^{-1}. Let $\{A_p^1\}$ and $\{A_p^2\}$, $p = 0, 1, \ldots$, denote the sequences of estimates obtained via these two EM estimators ($A_0^1 = 9.\overline{09} \times 10^{-3}$ and $A_0^2 = 10^{-1}$). Both EM estimators "converged" in the sense that, for each estimator, the magnitude of the relative difference in the values of the estimates on successive iterations decreased as the number of iterations increased. For each of the two sequences of estimates $\{A_p^i\}$, $i = 1, 2$, the so-called convergence iteration value was then taken to be the minimum iteration value l ($l \geq 1$) for which $|((A_l^i - A_{l-1}^i)/A_l^i)| < 10^{-7}$. Let p^* and p^{**} denote the convergence iteration values for $\{A_p^1\}$ and $\{A_p^2\}$, respectively. Then, of the two "limits" $A_{p^*}^1$ and $A_{p^{**}}^2$, the estimate $\hat{A}$ of the true parameter A was chosen to be that limit which maximized the traditional incomplete-data likelihood function, i.e.,

$$\hat{A} \triangleq \arg \max_{A \in \{A_{p^*}^1, A_{p^{**}}^2\}} L_X(A).$$

Let us now examine the performance of this estimator.

5.4.3. Simulation results

An extensive simulation study of the proposed estimator was performed for a wide range of parameter vectors $(A, K)^T$ in Λ, in particular, for all $(A, K)^T \in \Omega^*$, where $\Omega^* \triangleq \{(A, K)^T \in \Lambda \mid \log A \in Z \text{ and } \log K \in Z\}$. For each (A, K) pair, K was fixed to its true value and estimates of the paramter A were obtained using the implementation of the EM estimator of A described above. Using 100 data sets, each containing 100 samples generated from the Class A envelope pdf, the sample mean-square relative error was first computed. Let ρ_A denote this quantity. Then

$$\rho_A \triangleq \frac{1}{100} \sum_{j=1}^{100} \left(\frac{\hat{A}_j - A}{A} \right)^2 ,$$

where $\hat{A}_j$ denotes the estimate of A obtained using the j—th data set. The values for ρ_A are tabulated in Table 5.1. Note from this table that the values for the sample mean-square relative error are extremely low for all (A, K) pairs under consideration (on the order of 10^{-3} to 10^{-2}). The values of the normalized sample mean-square error (NSMSE), which is defined to be the product of the sample size (100) and the sample mean-square relative error, were then computed (see Table 5.2) and compared to the Cramer-Rao Lower Bound (CRLB). (The values for the CRLB are given in Table 5.3.) Note that the values for the NSMSE are very close to the CRLB for all (A, K) pairs. The reason that the values for the NSMSE are sometimes lower than the CRLB is due to the fact that the estimator is slightly biased, as is evident from the values of the sample relative bias, η_A,

$$\eta_A \triangleq \frac{1}{100} \sum_{j=1}^{100} \frac{\hat{A}_j - A}{A} ,$$

tabulated in Table 5.4. Now, it was noted in the implementation described above that the estimation process involved executing two EM estimators of A and each "converged" within a certain number of iterations, namely p_j^* and p_j^{**}. (The association with the j—th data set is made explicit

Table 5.1. SAMPLE MEAN-SQUARE RELATIVE ERROR (ρ_A)
(100 SAMPLES, 100 RUNS)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	3.2532×10^{-2}	1.0449×10^{-2}	8.7841×10^{-3}	8.6326×10^{-3}	8.6176×10^{-3}
10^{-1}	1.3591×10^{-2}	1.0324×10^{-2}	9.4479×10^{-3}	9.2596×10^{-3}	9.2451×10^{-3}
1	1.7950×10^{-2}	6.4732×10^{-3}	6.8095×10^{-3}	6.5755×10^{-3}	6.6118×10^{-3}

Table 5.2. NORMALIZED SAMPLE MEAN-SQUARE ERROR
(100 SAMPLES, 100 RUNS)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	3.2532	1.0449	8.7841×10^{-1}	8.6326×10^{-1}	8.6176×10^{-1}
10^{-1}	1.3591	1.0324	9.4479×10^{-1}	9.2596×10^{-1}	9.2451×10^{-1}
1	1.7950	6.4732×10^{-1}	6.8095×10^{-1}	6.5755×10^{-1}	6.6118×10^{-1}

Table 5.3. CRAMER-RAO LOWER BOUND FOR ESTIMATE OF A

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	4.0291	1.2041	1.0112	9.9246×10^{-1}	9.9055×10^{-1}
10^{-1}	1.3054	9.7417×10^{-1}	9.3217×10^{-1}	9.2600×10^{-1}	9.2509×10^{-1}
1	1.3227	9.6827×10^{-1}	9.0303×10^{-1}	8.9045×10^{-1}	8.8821×10^{-1}

Table 5.4. SAMPLE RELATIVE BIAS (η_A)
(100 SAMPLES, 100 RUNS)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	1.3683×10^{-1}	8.1403×10^{-2}	7.4757×10^{-2}	7.4103×10^{-2}	7.4038×10^{-2}
10^{-1}	9.1323×10^{-2}	8.0240×10^{-2}	7.6436×10^{-2}	7.5664×10^{-2}	7.5585×10^{-2}
1	8.4126×10^{-2}	6.8618×10^{-2}	7.1043×10^{-2}	6.8923×10^{-2}	6.9436×10^{-2}

Table 5.5. WORST-CASE CONVERGENCE ITERATION VALUE (p_{max})
(100 SAMPLES, 100 RUNS)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	7.88	6.59	5.32	4.85	4.41
10^{-1}	11.18	6.61	5.47	5.24	5.21
1	27.99	20.02	18.75	17.45	16.90

here.) Let $p_{\max}$ denote the sample worst-case convergence iteration value, i.e.,

$$p_{\max} \triangleq \frac{1}{100} \sum_{j=1}^{100} \max (p_j^* , p_j^{**}).$$

The values for $p_{\max}$ are tabulated in Table 5.5. Note that the values for this sample worst-case convergence iteration value are extremely low for all parameters under consideration (on the order of 10).

In summary, we see then that for a sample size of 100, the proposed EM estimator of A performs very well in a mean-square error sense, in terms of attaining the CRLB, and in terms of rate of convergence. Consequently, for the single-parameter estimation problem, it yields an excellent estimator for small sample sizes.

5.5. Two-Parameter Estimation Problem

The maximization problem described by (5.8) will now be considered. As for the single-parameter case, define $\Lambda_{\underline{\theta}}^*$ to be the following extension of the parameter set Λ:

$$\Lambda_{\underline{\theta}}^* \triangleq \{ (A , K)^T : \; 9.\overline{09} \times 10^{-3} \leqslant A \leqslant 1.1 \text{ and } 9.\overline{09} \times 10^{-7} \leqslant K \leqslant 1.1 \times 10^{-2} \} \; ,$$

and let $\Lambda_{\underline{\theta}}^*$ be the compact set over which the maximization will be performed, i.e.,

$$\underline{\theta}^{(p+1)} = \arg \max_{\underline{\theta} \in \Lambda_{\underline{\theta}}^*} Q_{obj} (\underline{\theta} \,|\, \underline{\theta}^{(p)}) \; , \qquad \underline{\theta}^{(0)} \in \Lambda_{\underline{\theta}}^* \; . \tag{5.24}$$

Consider first the gradient equations

$$\frac{\partial}{\partial A} Q_{obj} (\underline{\theta} \,|\, \underline{\theta}^{(p)}) = 0 \tag{5.25a}$$

and

$$\frac{\partial}{\partial K} Q_{obj} (\underline{\theta} \,|\, \underline{\theta}^{(p)}) = 0 \; . \tag{5.25b}$$

Upon computing the partial derivatives on the left-hand sides of (5.25a) and (5.25b), the gradient equations become

$$-n + \frac{\alpha}{A} + \frac{n}{A+K} - \sum_{i=1}^{n}\sum_{j=1}^{\infty} z_i^2 \frac{a_{ij}}{(j-1+K)} = 0 . \tag{5.26a}$$

and

$$\frac{n}{A+K} - \sum_{i=1}^{n}(1+z_i^2)\sum_{j=1}^{\infty}\frac{a_{ij}}{(j-1+K)} + (A+K)\sum_{i=1}^{n}\sum_{j=1}^{\infty} z_i^2 \frac{a_{ij}}{(j-1+K)^2} = 0 . \tag{5.26b}$$

Now (5.26a) simplifies straightforwardly to a quadratic in A. Using an argument analogous to the one given in Section 5.4, it can be shown that one of the roots to this quadratic can be disregarded, and the following closed-form expression for the parameter A can be obtained in terms of the parameter K:

$$A = \frac{nK + \xi_K K - \alpha - n - [(nK + \xi_K K - \alpha - n)^2 + 4(n+\xi_K)\alpha K]^{1/2}}{-2(n+\xi_K)}, \tag{5.27}$$

where $\xi_K \triangleq \sum_{i=1}^{n}\sum_{j=1}^{\infty} z_i^2 \frac{a_{ij}}{(j-1+K)}$. Upon substituting the expression for A given in (5.27) into (5.26b), we obtain an equation in K only. Thus, the two-variable maximization problem has been reduced to a maximization over the single variable K. Unfortunately, the resulting equation in K is highly nonlinear in this parameter. However, since the maximization is performed over the set $\Lambda_{\underline{\theta}}^{*}$, K takes on values in the interval $[9.\overline{09} \times 10^{-7}, 1.1 \times 10^{-2}]$. Thus $j-1+K \approx j-1$ for $j \geq 2$. Using this approximation, the highly nonlinear equation in K can be greatly simplified. Moreover, the square of the simplified equation can be reduced to the following fourth-order polynomial equation in K :

$$c_1 K^4 + c_2 K^3 + c_3 K^2 + c_4 K + c_5 \approx 0 , \tag{5.28}$$

where the coefficients c_i $(1 \leq i \leq 5)$ are functions only of the observations and the current estimate $\underline{\theta}^{(p)}$ of the parameters. (The expressions for these coefficients are given in Appendix B.) The roots of (5.28) can be readily determined. Consider the set consisting of the positive, real roots of (5.28) which lie in the open interval $(8 \times 10^{-7}, 1.2 \times 10^{-2})$ and which satisfy the simplified, unsquared equation in K. These roots can be refined using Newton's root-finding method on the original

equation initiated with these roots. For those refined roots which take on values in the interval $(9.\overline{09} \times 10^{-7}, 1.1 \times 10^{-2})$,[2] the corresponding value of A is then evaluated using (5.27). If the value of A lies in $(9.\overline{09} \times 10^{-3}, 1.1)$, then the resultant (A, K) pair yields a stationary point in the interior of $\Lambda_{\underline{\theta}}^{*}$. If the value of A does not lie in this interval, then the resultant (A, K) pair yields a stationary point that does not lie in the interior of $\Lambda_{\underline{\theta}}^{*}$, and hence, is eliminated from future consideration.

The above procedure yields those stationary points that lie in the interior of $\Lambda_{\underline{\theta}}^{*}$. Let us now focus our attention on the problem of determining those stationary points that lie in the interior of each of the four intervals which define the boundary of $\Lambda_{\underline{\theta}}^{*}$. First, consider the interval defined by $K = 9.\overline{09} \times 10^{-7}$ and $A \in [9.\overline{09} \times 10^{-3}, 1.1]$. Clearly, for fixed K, any stationary point must satisfy (5.27) evaluated at this fixed value of K. Now, upon evaluating (5.27) with $K = 9.\overline{09} \times 10^{-7}$, if the resulting value of A lies in $(9.\overline{09} \times 10^{-3}, 1.1)$, then the stationary point defined by this value of A and $K = 9.\overline{09} \times 10^{-7}$ is retained for future consideration; otherwise, it is disregarded. Now, consider the interval defined by $K = 1.1 \times 10^{-2}$ and $A \in [9.\overline{09} \times 10^{-3}, 1.1]$. Again, upon evaluating (5.27) with $K = 1.1 \times 10^{-2}$, if the resulting value of A lies in $(9.\overline{09} \times 10^{-3}, 1.1)$, then the stationary point defined by this value of A and $K = 1.1 \times 10^{-2}$ is retained for future consideration; otherwise, it is disregarded. Let us now determine the stationary points in the interior of the interval defined by $A = 9.\overline{09} \times 10^{-3}$ and $K \in [9.\overline{09} \times 10^{-7}, 1.1 \times 10^{-2}]$. Now, any stationary point in the interior of this interval must satisfy (5.26b) evaluated at $A = 9.\overline{09} \times 10^{-3}$. Since (5.26b) is highly nonlinear in K, the determination of the roots of (5.26b) $(A = 9.\overline{09} \times 10^{-3})$ can be simplified by using once again the approximation that $j-1+K \approx j-1$ for $j \geq 2$. Using this approximation, (5.26b) can be reduced to the following fourth-order polynomial equation in K :

---

[2]It was verified via simulation that the interval $(8\times10^{-7}, 1.2\times10^{-2})$ was sufficiently large to contain those roots whose refinement yields, at least, all roots of the original equation in $(9.\overline{09}\times10^{-7}, 1.1\times10^{-2})$.

$$d_1 K^4 + d_2 K^3 + d_3 K^2 + d_4 K + d_5 \approx 0 \ , \tag{5.29}$$

where $d_1 = \beta_4$, $d_2 = 2A\beta_4 - \beta_2$, $d_3 = \beta_4 A^2 - \beta_2 A + \beta_3 - \beta_1 + n$, $d_4 = 2\beta_3 A - \beta_1 A$, $d_5 = \beta_3 A^2$, and where the β_i's ($1 \leq i \leq 4$) are functions only of the observations and the current estimate of the parameters. (The expressions for these β_i's are given in Appendix B.) Consider the positive, real roots of (5.29) that lie in the interval $(8 \times 10^{-7}, 1.2 \times 10^{-2})$. These roots can be refined using Newton's root-finding method on (5.26b) ($A = 9.\overline{09} \times 10^{-3}$) initiated with these roots. If the value of the refined root lies in the interval $(9.\overline{09} \times 10^{-7}, 1.1 \times 10^{-2})$, then the stationary point defined by $A = 9.\overline{09} \times 10^{-3}$ and K equal to the value of the refined root is retained for future consideration.[3] Otherwise, it is disregarded. The stationary points in the interior of the interval defined by $A = 1.1 \times 10^{-2}$ and $K \in [9.\overline{09} \times 10^{-7}, 1.1 \times 10^{-2}]$ are determined in a completely analogous manner.

The function Q_{obj} is now evaluated at all stationary points in the interior of $\Lambda^*_{\underline{\theta}}$, at all stationary points in the interior of each of the four intervals which define the boundary of $\Lambda^*_{\underline{\theta}}$ and at the four corner points of $\Lambda^*_{\underline{\theta}}$: $(9.\overline{09} \times 10^{-3}, 1.1 \times 10^{-2})^T$, $(9.\overline{09} \times 10^{-3}, 9.\overline{09} \times 10^{-7})^T$, $(1.1 , 1.1 \times 10^{-2})^T$, and $(1.1 , 9.\overline{09} \times 10^{-7})^T$. Of the points at which Q_{obj} is evaluated, the solution to (5.24) is taken to be that point which maximizes Q_{obj}. For a given $\underline{\theta}^{(0)}$, we will refer to this solution as the *EM estimator of* $\underline{\theta}$.

5.5.1. Implementation of EM estimator of $\underline{\theta}$

As for the single-parameter estimation problem, two EM estimators of $\underline{\theta}$ were executed in parallel, one initiated with $(9.\overline{09} \times 10^{-3}, 1.1 \times 10^{-2})^T$ (the vector in $\Lambda^*_{\underline{\theta}}$ for which the ratio of the second coordinate to the first is largest), the other initiated with $(10^{-1}, 10^{-4})^T$ (an interior vector of $\Lambda^*_{\underline{\theta}}$ for which the ratio of the second coordinate to the first is the logarithmic mean over all

[3] Again, it was verified via simulation that the interval $(8\times10^{-7}, 1.2\times10^{-2})$ was sufficiently large to contain those roots whose refinement yields, at least, all roots of the original equation in $(9.\overline{09}\times10^{-7}, 1.1\times10^{-2})$.

possible values that this ratio can take on for parameter vectors in $\Lambda_{\underline{\theta}}^{*}$).[4] Let $\{\underline{\theta}_1^{(p)}\} \triangleq \{(A_p^1, K_p^1)^T\}$ and $\{\underline{\theta}_2^{(p)}\} \triangleq \{(A_p^2, K_p^2)^T\}$, $p = 0, 1, \ldots$, denote the sequences of estimates obtained via these two EM estimators ($\underline{\theta}_1^{(0)} = (9.\overline{09} \times 10^{-3}, 1.1 \times 10^{-2})^T$ and $\underline{\theta}_2^{(0)} = (10^{-1}, 10^{-4})^T$). Both EM estimators "converged" in the sense that, for each estimator, the magnitude of the relative difference in the values of the estimate (of each parameter) on successive iterations decreased as the number of iterations increased. For each of the two sequences of estimates $\{\underline{\theta}_i^{(p)}\}$, $i = 1, 2$, the "convergence iteration value" was then taken to be the minimum iteration value l ($l \geq 1$) for which $|((A_l^i - A_{l-1}^i)/A_l^i)| < 10^{-7}$ and $|((K_l^i - K_{l-1}^i)/K_l^i)| < 10^{-7}$. Let p' and p'' ($p', p'' \geq 1$) denote the convergence iteration values for $\{\underline{\theta}_1^{(p)}\}$ and $\{\underline{\theta}_2^{(p)}\}$, respectively. Then of the two "limits" $\underline{\theta}_1^{(p')}$ and $\underline{\theta}_2^{(p'')}$, the estimate $\hat{\underline{\theta}}$ of the true parameter vector was chosen to be that limit which maximized the traditional incomplete-data likelihood function, i.e.,

$$\hat{\underline{\theta}} = \arg \max_{\underline{\theta} \in \{\underline{\theta}_1^{(p')}, \underline{\theta}_2^{(p'')}\}} L(\underline{\theta}) \ .$$

where

$$L(\underline{\theta}) = \sum_{i=1}^{n} \ln \sum_{j=1}^{\infty} \pi_j(A) f_j(z_i; A, K) \ .$$

Let us examine the performance of this estimator.

5.5.2. Simulation results

As for the single-parameter estimation problem, an extensive simulation study of the proposed estimator was performed. Using 100 data sets, each containing 100 samples generated from the Class A envelope pdf, the sample mean-square norm relative error (MSNRE) was first

[4] It was observed via simulation that the convergence of the EM estimator to a vector that closely approximated the true parameter vector was highly dependent on the value of this ratio. In particular, if the ratio of the second coordinate of the initiating vector to the first exceeded (but not considerably) the corresponding ratio for the true parameter vector, then the EM estimator was guaranteed to converge to a vector that closely approximated the true parameter vector. Thus, given the two initiating vectors, at least one of the EM estimators was guaranteed to converge to a vector with this property. This restriction on the above-defined ratio was also seen to be a necessary condition for convergence of the BDD algorithm formulated in Section 4.2.

computed for each $\underline{\theta} \in \Omega^*$. (The definition of Ω^* was given in Section 5.4.3.) Let $\rho_{\underline{\theta}}$ denote this quantity. Then,

$$\rho_{\underline{\theta}} = \frac{1}{100} \sum_{j=1}^{100} \left[\frac{\hat{A}_j - A}{A} \right]^2 + \left[\frac{\hat{K}_j - K}{K} \right]^2 ,$$

where $\hat{\underline{\theta}}_j \triangleq (\hat{A}_j, \hat{K}_j)^T$ $(1 \leq j \leq 100)$ denotes the estimate of $\underline{\theta}$ obtained using the j−th data set. The values for $\rho_{\underline{\theta}}$ are tabulated in Table 5.6. Note from this table that the values of this relative error are quite low for all parameter vectors under consideration (on the order of 10^{-2} to 10^{-1}). The values of the normalized sample mean-square norm relative error ($\triangleq n \times$ MSNRE) were then computed (see Table 5.7) and compared to the CRLB. (The values of the CRLB were given in Table 3.8.) Note that the values of the normalized MSNRE are either close to the CRLB or lower than the CRLB. The reason that the values for the normalized MSNRE are sometimes lower than the CRLB is due to the fact that the estimator is somewhat biased, as is evident from the values of the sample relative bias, $\eta_{\underline{\theta}}$,

$$\eta_{\underline{\theta}} = \frac{1}{100} \sum_{j=1}^{100} \left[\frac{\hat{A}_j - A}{A} \right] + \left[\frac{\hat{K}_j - K}{K} \right],$$

tabulated in Table 5.8. Now, it was noted earlier that the estimation of $\underline{\theta}$ involved the execution of two EM estimators of $\underline{\theta}$ in parallel and each "converged" within a certain number of iterations, namely p_j' and p_j''. (The association with the j−th data set is made explicit here.) Let $p_{\underline{\theta}}^{\max}$ denote the sample worst-case convergence iteration value, i.e.,

$$p_{\underline{\theta}}^{\max} \triangleq \frac{1}{100} \sum_{j=1}^{100} \max(p_j', p_j'').$$

The values for $p_{\underline{\theta}}^{\max}$ are tabulated in Table 5.9. As was the case for the single-parameter estimation problem, note that the values for this sample worst-case convergence iteration value here are also extremely low for all parameter vectors under consideration (on the order of 10).

In summary then, we see that for a sample size of 100, the proposed EM estimator of $\underline{\theta}$ performs again very well in a mean-square error sense, in terms of attaining the CRLB, and in terms

Table 5.6. SAMPLE MEAN-SQUARE NORM RELATIVE ERROR ($\rho_{\underline{\theta}}$)
(100 SAMPLES, 100 RUNS)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	5.6312×10^{-2}	7.7317×10^{-1}	7.0539×10^{-1}	6.9907×10^{-1}	6.8783×10^{-1}
10^{-1}	8.5002×10^{-2}	1.3443×10^{-1}	1.2567×10^{-1}	1.2522×10^{-1}	7.484×10^{-2}
1	3.4455×10^{-2}	4.9856×10^{-2}	4.8199×10^{-2}	4.6672×10^{-2}	3.0990×10^{-2}

Table 5.7. NORMALIZED MEAN-SQUARE NORM RELATIVE ERROR
(100 SAMPLES, 100 RUNS)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	5.6312	77.317	70.539	69.907	68.783
10^{-1}	8.5002	13.443	12.567	12.522	7.484
1	3.4455	4.9856	4.8199	4.6672	3.0990

Table 5.8. SAMPLE RELATIVE BIAS ($\eta_{\underline{\theta}}$)
(100 SAMPLES, 100 RUNS)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	2.6277×10^{-1}	6.7566×10^{-1}	6.5280×10^{-1}	6.5049×10^{-1}	6.0676×10^{-1}
10^{-1}	3.2896×10^{-1}	4.1521×10^{-1}	4.0252×10^{-1}	4.0185×10^{-1}	2.7972×10^{-1}
1	2.0539×10^{-1}	2.3343×10^{-1}	2.3146×10^{-1}	2.3049×10^{-1}	1.8362×10^{-1}

Table 5.9. WORST-CASE CONVERGENCE ITERATION VALUE ($p_{\max}^{\underline{\theta}}$)
(100 SAMPLES, 100 RUNS)

A \ K	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	4.98	14.80	6.78	4.22	4.25
10^{-1}	22.24	10.70	8.70	8.53	7.82
1	27.98	23.86	22.73	22.62	20.38

of rate of convergence. Consequently, it yields an excellent estimator for small sample sizes for the two-parameter estimation problem as well.

5.6. Conclusions

In summary, we have seen in this chapter that the EM algorithm uses the properties of the incomplete-date likelihood function in the estimation process, and so is ideally suited for estimating the parameters of mixture densities such as the Class A density. It has many desirable features, e.g., its monotonicity property for the likelihood sequence and its estimation potential for small sample sizes. Moreover, within the context of estimating the Class A parameters, we have shown that, for the single-parameter estimation problem, the sequence of EM estimates converges. Furthermore, we have shown that if the limit point to which the sequence converges is an interior point of the compact set over which the maxmization is performed, then it must necessarily be a *stationary point of the likelihood function of interest.* Using an implementation based on the execution of two EM algorithms in parallel, an extensive simulation study for both the single-parameter and two-parameter estimation problems was also performed. The results of these studies indicate that the proposed EM estimator does, in fact, yield an excellent estimator of the Class A parameters for small sample sizes for all parameter vectors in the parameter set of interest.

6. SUMMARY AND CONCLUSIONS

In this work, we have developed and examined the performance of various optimal and near-optimal identification procedures for the Class A interference model.

We considered first the problem of basic batch estimation of the Class A parameters from an independent sequence of Class A envelope samples. From this study, it was seen that the method-of-moments estimator based on the fourth and sixth moments yielded strongly consistent and asymptotically normal estimates of the parameters, but was highly inefficient due to the insensitivity of the moments to changes in the parameter K. However, via an examination of the Cramer-Rao Lower Bound, it was seen that a tremendous improvement in performance over the method-of-moments estimator was possible if an asymptotically efficient estimator could be found. Unfortunately, maximum likelihood estimation proved to be computationally unwieldy due largely to the multiplicity of roots in the likelihood equation. However, by initiating Newton's root-finding method on the likelihood equation with the method-of-moments estimator, a procedure was obtained that combined the consistency and efficiency of the two approaches. Despite its asymptotic optimality, this Moment/Likelihood procedure did not perform well for moderate sample sizes because of the extremely high inefficiency of the moments estimator. However, by initiating likelihood search with the more efficient, physically-motivated Threshold-Comparison estimator, a batch estimator of the Class A parameters was obtained that yields good estimates of the parameters for moderate sample sizes.

The problem of recursive identification of the Class A parameters was then addressed. The starting point in the development of a global, recursive estimator of the parameters was the BDD algorithm. This algorithm is physically motivated, easy to implement, and is a recursive version of the Threshold-Comparison estimator, which was seen in the batch estimation problem to yield accurate estimates of the parameters. In particular, this basic decision-directed algorithm is based on an adaptive, Bayesian classification of each of a sequence of Class A envelope samples as an impulsive sample or as a background sample. As each sample is so classified, recursive updates of

the estimates of the second moment of the background component of the interference envelope density, the second moment of the impulsive component of the interference envelope density, and the probability with which the impulsive component occurs, are readily obtained. From these estimates, estimates of the parameters of the Class A model follow straightforwardly, since closed-form expressions exist in terms of these quantities. Examination of the performance characteristics of the algorithm revealed two inherent drawbacks of the algorithm, which adversely affect its performance even locally. However, by imposing the necessary restrictions on the form of the initiation vector for the algorithm and incorporating the appropriate modifications into its framework, the ensuing difficulties associated with these two drawbacks can be eliminated. The result was a global, recursive estimator of the Class A parameters that yields excellent estimates of the parameters for all parameter vectors in the parameter set of interest.

The problem of efficient estimation of the Class A parameters for small sample sizes was then considered. The proposed estimator was based on the EM algorithm, a two-step iterative estimation technique which was ideally suited for the Class A estimation problem since the observations could be readily treated as "incomplete data." For the single-parameter estimation problem (A unknown, K known), a closed-form expression for the estimator was obtained. The convergence properties of the EM algorithm as they pertain to the Class A estimation problem were also examined. Again, for the single-parameter estimation problem, (A unknown, K known), it was shown that the sequence of estimates obtained via the EM algorithm converges. Moreover, it was shown that if the limit point to which the sequence converges is an interior point of the set over which the optimization is performed, then it must necessarily be a stationary point of the traditional likelihood function. The small-sample-size performance of the EM algorithm was also examined via simulation (for both the single-parameter and two-parameter estimation problems). For each true parameter vector, two EM algorithms were executed in parallel, each initiated with a different initiating vector, the pair of initiating vectors being fixed for each true parameter vector. For each initiating vector, the EM algorithm converged to a limit vector in the parameter set of interest, and the estimate vector was then taken to be the limit vector that maximized the incomplete-data likelihood

function. For each initiating vector and true parameter vector, the EM algorithm converged to the limit vector within relatively few iterations (on the order of 10). Moreover, via an extensive simulation study, it was seen that this likelihood-based scheme yields excellent estimates of the parameters of the Class A model (in terms of attaining the Cramer-Rao Lower Bound) for small sample sizes (on the order of 10^2).

This study has been devoted to the problem of obtaining optimal and near-optimal identification procedures for the parameters of the Class A interference model. A thorough investigation of this problem has been made, with the objective of obtaining theoretically optimal and practically efficient estimation procedures for these parameters. The problem of efficient estimation in both the batch and recursive frameworks has been addressed. It is anticipated that the results of this study will find widespread application in the areas of digital communications, sonar, and radar due to the common occurrence of impulsive channels in such systems.

APPENDIX A. DERIVATION OF RESTRICTIONS ON BDD ALGORITHM

Let us now determine the source of the two flaws of the BDD algorithm given in Section 4.2.2. Note that, in a given iteration of the algorithm, the only error that can be incurred is in the decision-making process, and the decision-making process is based on a LRT that uses an estimate of the true parameter vector in the LR to classify the given sample. Thus, by examining the effects on the error probability in the classification process due to the mismatch introduced from a lack of knowledge of the true parameter vector value, some insight can be gained into the source of the two flaws of the BDD algorithm. In particular, for a comprehensive range of values of $(A\ \ K, \sigma^2) \in \Lambda'$ and $(\underline{A}, \underline{K}, \underline{\sigma}^2) \in \Lambda'$, we can compute the probability of making an incorrect classification when $(\underline{A}, \underline{K}, \underline{\sigma}^2)$ is used in the LRT to classify the given observation given that (A, K, σ^2) is the true parameter vector. The expression for this probability of error can be easily derived by noting from the definition of $\tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)}$ given in Section 4.2.1 that this is simply the probability of making an incorrect classification when the decision regions $(0, \tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)})$, $(\tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)}, \infty)$ are used, given that (A, K, σ^2) is the true parameter vector. Let us denote this probability of error by $P_e((\underline{A}, \underline{K}, \underline{\sigma}^2); (A, K, \sigma^2))$. (In the sequel, the parenthetical arguments will be included only if necessary for clarity.) We then have that

$$
\begin{aligned}
&P_e((\underline{A}, \underline{K}, \underline{\sigma}^2); (A, K, \sigma^2)) \\
&= e^{-A} \int_{\tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)}}^{\infty} p_0(z; A, K, \sigma^2)\, dz + (1 - e^{-A}) \int_0^{\tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)}} p_1(z; A, K, \sigma^2)\, dz \\
&= e^{-A} \int_{\tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)}}^{\infty} \frac{2z}{\sigma^2} \left| \frac{A+K}{K} \right| e^{-\frac{z^2}{\sigma^2}\left|\frac{A+K}{K}\right|} dz \\
&+ (1 - e^{-A}) \int_0^{\tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)}} \left| \frac{2e^{-A}}{\sigma^2 (1 - e^{-A})} \sum_{m=1}^{\infty} \frac{A^m}{m!} \left| \frac{A+K}{m+K} \right| z\, e^{-\frac{z^2}{\sigma^2}\left|\frac{A+K}{m+K}\right|} \right| dz
\end{aligned}
\tag{A.1}
$$

After some manipulation, (A.1) yields the following expression for $P_e((\underline{A}, \underline{K}, \underline{\sigma}^2); (A, K, \sigma^2))$:

$$P_e((\underline{A}, \underline{K}, \underline{\sigma}^2); (A, K, \sigma^2)) = e^{-A} e^{-\left|\frac{(\tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)})^2}{\sigma^2}\right|\left|\frac{A+K}{K}\right|} + (1 - e^{-A}) - \sum_{m=1}^{\infty} \frac{e^{-A} A^m}{m!} e^{-\left|\frac{(\tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)})^2}{\sigma^2}\right|\left|\frac{A+K}{m+K}\right|} . \quad \text{(A.2)}$$

Let $\Omega \triangleq \{(\alpha, \beta, \xi) : (\alpha, \beta, \xi) \in \Lambda \text{ and } \log \alpha \in \mathbf{Z}, \log \beta \in \mathbf{Z}, \log \xi = 0\}$. Given the above expression for P_e, we have computed, for each $(A, K, \sigma^2) \in \Omega$, $P_e((\underline{A}, \underline{K}, \underline{\sigma}^2); (A, K, \sigma^2))$ for all $(\underline{A}, \underline{K}, \underline{\sigma}^2) \in \Omega$. It is evident from (A.2) that the evaluation of these error probabilities requires that $\tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)}$ first be evaluated for all $(\underline{A}, \underline{K}, \underline{\sigma}^2) \in \Omega$. Since $f(z; \underline{A}, \underline{K}, \underline{\sigma}^2)$ is a strictly monotone increasing function of z for each $(\underline{A}, \underline{K}, \underline{\sigma}^2) \in \Lambda'$, this was easily done using a numerical search procedure. The results are tabulated in Table A.1. (All tables referenced in this appendix can be found at the end of the appendix.) Using the tabulated values for $\tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)}$ and (A.2), $P_e((\underline{A}, \underline{K}, \underline{\sigma}^2); (A, K, \sigma^2))$ was then computed for the aforementioned values of (A, K, σ^2) and $(\underline{A}, \underline{K}, \underline{\sigma}^2)$. The computed error probabilities are tabulated in Tables (A.2)-(A.6), (A.7)-(A.11), and (A.12)-(A.16). We will now use these error probabilities to explain the observed difficulties of the BDD algorithm cited above.

Explanation of Observed Difficulties

We will now justify the observed difficulties of the BDD algorithm, and, in so doing, obtain restrictions on the form of the initiation vector $(\tilde{A}_0, \tilde{K}_0, \tilde{\sigma}_0^2)$ for the algorithm. As stated in Section 4.2.2, the observed difficulties of the algorithm are cited for the case when $\tilde{\sigma}_n^2$ is fixed so that $\tilde{\sigma}_n^2 = \sigma^2 = 1$ for all $n \geq 0$ and, thus, will be justified for this case.

Consider the tables for P_e. Since $\underline{\Gamma} \triangleq \underline{K}/\underline{A}$ (see Eq. (2.6)), we see that for a given table, the lines of constant $\underline{\Gamma}$ correspond to diagonals along that table, with $\underline{\Gamma}$ attaining a maximum value of 1 when $(\underline{A}, \underline{K}) = (10^{-2}, 10^{-2})$ and a minimum value of 10^{-6} when $(\underline{A}, \underline{K}) = (1, 10^{-6})$. Note from tables (A.2)-(A.6) and (A.7)-(A.11) that for fixed $\underline{A}$, $P_e((\underline{A}, \underline{K}, \underline{\sigma}^2); (A, K, \sigma^2))$ becomes very large very rapidly with decreasing $\underline{K}$ for parameter vectors $(\underline{A}, \underline{K}, \underline{\sigma}^2)$ for which $\underline{\Gamma}$ is less

than Γ. Moreover, for values of the parameter vectors $(\underline{A}, \underline{K}, \underline{\sigma}^2)$ for which $\underline{\Gamma} \geq \Gamma$, P_e for these parameter vectors is relatively small. Thus, we have a so-called "diagonal effect," wherein for values of the parameter vectors $(\underline{A}, \underline{K}, \underline{\sigma}^2)$ for which $\underline{\Gamma} < \Gamma$, the probability of error incurred in the decision-making process by using these vectors in the LRT when (A, K, σ^2) is the true parameter vector is relatively large, whereas for values of the parameter vectors $(\underline{A}, \underline{K}, \underline{\sigma}^2)$ for which $\underline{\Gamma} \geq \Gamma$, the probability of error incurred is relatively small. Thus, for each (A, K, σ^2), the "diagonal effect" divides the plane of Λ' consisting of parameter vectors for which the third coordinate has unity value into an upper-diagonal and a lower-diagonal region as defined by the line corresponding to constant Γ. To parameter vectors in the upper-diagonal region, we can associate values of P_e that are predominantly small, and to those in the lower-diagonal region, we can associate values of P_e that are predominantly large.

This phenomenon justifies the initial portion of the observation cited in (ii) of Section 4.2.2, namely, that for initiation vectors $(\tilde{A}_0, \tilde{K}_0, \tilde{\sigma}_0^2)$ for which $\tilde{\Gamma}_0 \geq \Gamma$, the frequency with which the algorithm converges to the true parameter vector is relatively high, whereas for initiation vectors $(\tilde{A}_0, \tilde{K}_0, \tilde{\sigma}_0^2)$ for which $\tilde{\Gamma}_0 < \Gamma$, the frequency with which the algorithm converges to the *wrong* parameter vector is relatively high. We see then that the diagonal effect can be used advantageously by always initiating the BDD algorithm with an initiation vector $(\tilde{A}_0, \tilde{K}_0, \tilde{\sigma}_0^2)$ for which

(1) $\tilde{\Gamma}_0$ either provides an accurate estimate of Γ, or, $\tilde{\Gamma}_0$ provides an estimate of Γ for which $\tilde{\Gamma}_0 > \Gamma$.

Note that for each true parameter vector (A, K, σ^2) the upper-diagonal and lower-diagonal regions defined by the line of constant Γ have the aforementioned properties of $P_e((\underline{A}, \underline{K}, \underline{\sigma}^2); (A, K, \sigma^2))$ associated with them if $\underline{\sigma}^2 = \sigma^2$ and $\sigma^2 = 1$. The constraint given by the first equality demands that the restriction given in (1) be accompanied by the additional restriction that

(2) $\tilde{\sigma}_0^2$ must provide an accurate estimate of σ^2.

Given that $\tilde{\sigma}_0^2$ accurately estimates σ^2, we will, in addition, replace (4.8c) with an estimator of σ^2 consisting of an update equation for $\tilde{\sigma}_0^2$. In so doing, we will obtain an estimate equation for σ^2 which yields an accurate estimate of σ^2 at each iteration of the algorithm.

The second constraint, namely $\sigma^2 = 1$, does not impose any additional restrictions for the following reason: From the approximation given in Step (2') of the MBDD algorithm (Section 4.3.1), it can be seen that, via some minor approximations, the dependence of $(\tau_{opt}^{(\underline{A}, \underline{K}, \underline{\sigma}^2)})^2$ on $\underline{\sigma}^2$ is essentially linear for all $(\underline{A}, \underline{K}, \underline{\sigma}^2) \in \Lambda$. From (A.2), it then follows that $P_e((\underline{A}, \underline{K}, \underline{\sigma}^2); (A, K, \sigma^2))$ depends only on the ratio of $\underline{\sigma}^2$ to σ^2, i.e., given that $\underline{\sigma}^2 = \sigma^2$, the values for $P_e((\underline{A}, \underline{K}, \underline{\sigma}^2); (A, K, \sigma^2))$ are essentially invariant to changes in σ^2 for each true parameter vector (A, K, σ^2). Consequently, since restriction (2) takes into account the fact that the properties for P_e are valid only when $\underline{\sigma}^2 = \sigma^2$, no additional restriction on σ^2 need be imposed.

Note, in addition, from Tables (A.7)-(A.11) that for each true parameter vector (A, K, σ^2) for which $A = 1$, there exists a set of vectors $(\underline{A}, \underline{K}, \underline{\sigma}^2)$ for which $\underline{\Gamma} > \Gamma$ but for which P_e once again becomes relatively large, namely, the set of vectors for which $\underline{\Gamma}$ is approximately on the order of 10^{-1} or larger. The values for P_e tabulated for these $\underline{\Gamma}$ support the observation, cited in (ii) of Section 4.2.2, that for values of A close to 1, the frequency with which the BDD algorithm converges to the wrong parameter vector becomes relatively high for values of $\tilde{\Gamma}_0$ approximately on the order of 10^{-1} or larger. Instead of translating this observation into a restriction on $\tilde{\Gamma}_0$ as was done with the diagonal effect, we will see in the sequel that this observation more readily translates into a restriction on $\tilde{A}_0$.

Consider the observations cited in (i) of Section 4.2.2. Suppose that we have an iteration n'' for which the associated proportion of samples classified as impulsive exceeds the expected percentage (and for which all samples at iterations $n \leq n''$ have been correctly classified). (In the sequel, this will be referred to as a situation wherein there is a "disproportionate excess of impulses.") Let

us see how this can cause the algorithm to behave poorly: Since the proportion of samples classified as impulsive at iteration n'' exceeds the expected percentage, it follows from (4.8a) that the estimate of A at that iteration will exceed the true A. Now, if the estimate of K at iteration n'' is such that the estimate of Γ (which is defined to be the estimate of K divided by the estimate of A) at that iteration either exceeds or equals the true Γ, then the estimate vector at iteration n'' will lie in the upper-diagonal region corresponding to the true parameter vector. Consequently, by the "diagonal effect," the probability of error incurred in the decision-making process at the next iteration of the algorithm will tend to be relatively low, and thus, with relatively high probability, the sample at iteration $n''+1$ will be classified correctly. However, if the estimate of K at iteration n'' is such that the estimate of Γ is less than the true Γ, then the estimate vector at iteration n'' will lie in the lower-diagonal region corresponding to the true parameter vector and, by the "diagonal effect," the probability of error incurred at the next iteration of the algorithm will tend to be relatively high. Now, from Table A.1, we see that if the estimate vector is such that the estimate of A is greater than the true A and the estimate of Γ is less than the true Γ, then the optimum threshold corresponding to the estimate vector will be less than the optimum threshold corresponding to the true parameter vector. Given this and the relatively high probability of an incorrect classification associated with the estimate vector lying in the lower-diagonal region, we have a situation wherein a background sample can be incorrectly classified as an impulsive sample at the next iteration of the algorithm. Since the estimate of A is directly proportional to the proportion of samples that have been classified as impulsive (of course, within the boundary restrictions imposed by Λ), an incorrect classification of a background sample as impulsive at iteration $n''+1$ will raise the estimate of A even further. Thus, via these successively increasing values for the estimate of A, the estimate vector can potentially be "forced away" from the true parameter vector. We see then that a disproportionate excess of impulses at a given iteration of the algorithm can have a detrimental effect on its performance.

Note that the likelihood with which the performance of the algorithm is adversely affected increases with increasing values obtained for the estimate of A, given that the estimate of A

exceeds the true value. This is accounted for as follows: Consider Tables (A.4)-(A.6) and (A.7)-(A.11). In each table, note that for a fixed but arbitrary value of $\underline{A}$, the portion of the lower-diagonal region consisting of the parameter vectors for which the first coordinate is $\underline{A}$ is an interval, and the length of this interval increases with increasing $\underline{A}$. Thus, as the estimate of A increases, the length of the interval associated with this estimate increases. Since this interval defines the set of parameter vectors under consideration for which the first coordinate is the estimate of A and for which the ratio of the second coordinate to the first is less than the true Γ, it follows that as the estimate of A increases, the likelihood that the estimate of Γ will be less than the true Γ increases. Given the ensuing detrimental effect on the behavior of the algorithm (described in the previous paragraph) associated with the estimate of A being greater than the true A and the estimate of Γ being less than the true Γ, it follows that the increased likelihood of the estimate of Γ being less than the true Γ in turn increases the likelihood that the performance of the algorithm will be adversely affected.

It was noted in (i) of Section 4.2.2 that the convergence of the BDD algorithm is sensitive to the distribution of impulses in the *initial* stages of the algorithm. That a disproportionate excess of impulses at a given iteration of the algorithm is more likely to affect its convergence if the iteration occurs in the initial stages of the algorithm follows from the fact that the estimates of the update parameters and, hence, the estimates of the model parameters, are more sensitive to changes in the number of samples classified as impulsive in the initial stages of the algorithm. For example, suppose that the true value of the parameter A is 10^{-2}. Since the corresponding true value for π_1 is 9.95×10^{-3}, suppose, in addition, that $\tilde{\pi}_1(100) = 9.95\times10^{-3}$. Now, if the sample at the 101-st iteration is classified as impulsive, then $\tilde{\pi}_1(101)$ will be 1.98×10^{-2}. However, if $\tilde{\pi}_1(2000) = 9.95\times10^{-3}$ and the sample at the 2001-st iteration is classified as impulsive, then $\tilde{\pi}_1(2001)$ will be 1.04×10^{-2}. Thus, even though the proportions of samples classified as impulsive at the 100-th and 2000-th iterations are the same (and are equal to the true value), an increase in the number of samples classified as impulsive by one (resulting in a disproportionate excess of impulses) yields a larger value for the estimate of π_1 at the 101-st iteration, and hence, a larger value for the estimate of A.

than at the 2001-st iteration ($\tilde{A}_{101} = 2.00\times10^{-2}$ whereas $\tilde{A}_{2001} = 1.05\times10^{-2}$). Since the estimate of A associated with the 101-st iteration is larger than that corresponding to the 2001-st iteration (and both exceed the true value), it follows from the observation cited in the first statement of the previous paragraph that the performance of the algorithm is more likely to be adversely affected by the estimate of A associated with the 101-st iteration. Via this example, we see then that a disproportionate excess of impulses at a given iteration of the algorithm is more likely to affect its convergence if the iteration occurs in the initial stages of the algorithm.

It was also noted in (i) of Section 4.2.2 that a disproportionate excess of impulses will have a significant effect on the convergence of the algorithm for values of A close to 10^{-2}. That a disproportionate excess of impulses is more likely to affect the convergence of the algorithm for smaller values of A is best explained with an example: Suppose that the true value of the parameter A is 10^{-2} and suppose that the first impulsive sample occurs at the second iteration of the algorithm. Then, given that the samples have been classified correctly, there will be a disproportionate excess of impulses at this second iteration. Furthermore, the minimum possible iteration value for which there will not be a disproportionate excess of impulses is 101 (such would be the case if there are no impulses for iteration values $3 \leqslant n \leqslant 101$ and all samples are classified correctly). Now if the true value of the parameter A is 10^{-1} and the first impulsive sample occurs at the second iteration of the algorithm, then again there will be a disproportionate excess of impulses at this second iteration but the minimum possible iteration value thereafter for which there will not be a disproportionate excess of impulses is now 11. Thus, an iteration for which there is a disproportionate excess of impulses has resulted in a greater number of ensuing iterations for which there is a disproportionate excess of impulses for the smaller value of A. Given the possible detrimental effect on the performance of the algorithm associated with an iteration for which there is a disproportionate excess of impulses (discussed above), it follows that an increased number of iterations for which there is such an excess increases the likelihood that the algorithm will behave poorly. We see then that a disproportionate excess of impulses at a given iteration of the algorithm is more likely to affect its convergence for smaller values of A.

Finally, let us justify the last observation cited in (i) of Section 4.2.2, namely, that for a fixed value of the true parameter A, the likelihood that a disproportionate excess of impulses will affect the convergence of the algorithm increases with increasing K: Consider Tables (A.2)-(A.6) and (A.7)-(A.11). For each true parameter vector, let the "truncated lower-diagonal region" and "truncated upper-diagonal region" denote the set of parameter vectors in the lower-diagonal region and upper-diagonal region, respectively, for which the first coordinate exceeds the true A. Note that for fixed value of the true parameter A and increasing value of the true parameter K, there is a corresponding increase in the size of the truncated lower-diagonal region associated with the true parameter vector and a decrease in the size of the truncated upper-diagonal region. Now, suppose that we have an iteration for which there is a disproportionate excess of impulses. Then the estimate of A at that iteration will exceed the true A and, consequently, the estimate vector at that iteration will either lie in the truncated lower-diagonal region or in the truncated upper-diagonal region corresponding to the true parameter vector. Since the size of the truncated lower-diagonal region increases with increasing K, the probability that the estimate vector will lie in the truncated lower-diagonal region increases with increasing K. Moreover, the estimate vector is more likely to lie in the truncated lower-diagonal region associated with the true parameter vector than in the truncated upper-diagonal region since the size of the first region relative to the second increases with increasing K. Given the detrimental effect on the performance of the algorithm (discussed earlier) that results from the estimate of A being greater than the true A and the estimate of Γ being less than the true Γ, it follows from the increased likelihood of the estimate vector lying in the truncated lower-diagonal region (and decreased likelihood of it lying in the truncated upper-diagonal region) that there will be an increased likelihood of the algorithm performing poorly.

For fixed A, the most pronounced effect on the algorithm's convergence is for $K = 10^{-2}$ since the truncated lower-diagonal region associated with the true parameter vector is largest for this value of K. Moreover, the truncated upper-diagonal region is an empty set. This suggests a means for improving the performance of the algorithm somewhat for larger values of K, namely by extending the upper-diagonal region (and, hence, the truncated upper-diagonal region) associated

with the parameter vectors corresponding to these larger values of K. This can be done by considering a modified parameter set consisting of the following parallelepiped: $\{(A, K, \sigma^2): 10^{-2} \leq A \leq 1, 10^{-6} \leq \Gamma \leq 1, \sigma^2 > 0\}$. However, given that Λ' is the parameter set of practical interest for the estimation problem at hand, another means of improving the performance of the algorithm suggests itself by recalling that (a) a disproportionate excess of impulses forces the estimate vector away from the true parameter vector through successively increasing values in the estimate of A, and that (b) the performance of the algorithm is more likely to be affected if the disproportionate excess of impulses occurs in the initial stages of the algorithm. From these earlier observations, we see then that an improvement in performance is possible if, in the initial stages of the algorithm, the estimate of A is fixed to its initial value $\tilde{A}_0$ and only the estimates of K and σ^2 are updated. In this manner, the divergence of the estimate of A from its true value can be prevented, and this, in turn, should increase the frequency with which the algorithm converges to the true parameter vector. To accommodate this modification of the BDD algorithm, the following restriction on $\tilde{A}_0$ must be imposed: Namely,

(3) $\tilde{A}_0$ must either provide an accurate estimate of A, or, $\tilde{A}_0$ must provide an estimate of A for which $\tilde{A}_0 < A$.

Suppose this was not the case, i.e., suppose that $\tilde{A}_0$ were significantly larger than the true A. Now, restrictions (1) and (2) on $\tilde{\Gamma}_0$ and $\tilde{\sigma}_0^2$, respectively, guarantee that the initiation vector $(\tilde{A}_0, \tilde{K}_0, \tilde{\sigma}_0^2)$ for the algorithm will lie in the upper-diagonal region corresponding to the true parameter vector. Thus, the probability of error incurred in the decision-making process will be relatively small initially. Consequently, with relatively high probability, the impulsive and background samples will be correctly classified and, eventually, this results in an estimate of K which accurately approximates the true value. (The estimate of σ^2 will always accurately approximate the true value for the reasons discussed above.) Eventually, then, we have a situation where the estimate of A is fixed at its initial value $\tilde{A}_0$ and the estimates of K and σ^2 accurately describe the true values, i.e., the estimate vector of the true parameter vector lies in the lower-diagonal region

corresponding to the true parameter vector. Thus, the probability of error associated with this estimate vector will be relatively high and this may adversely affect the estimate of K and/or the estimate of A (if the algorithm is at the stage where the estimate of A is once again being updated). Given these ensuing adverse effects on the estimates of the parameters resulting from $\tilde{A}_0$ being larger than the true A, we necessarily impose the restriction given in (3).

Recall our earlier observation that for parameter vectors (A, K, σ^2) for which $A = 1$ $(\sigma^2 = 1)$, $P_e((\underline{A}, \underline{K}, \underline{\sigma}^2) \cdot (A, K, \sigma^2))$ becomes relatively large for parameter vectors $(\underline{A}, \underline{K}, \underline{\sigma}^2)$ $(\underline{\sigma}^2 = 1)$ for which $\underline{\Gamma}$ is approximately on the order of 10^{-1} or larger. In particular, for the true parameter vector $(A, K, \sigma^2) = (1, 10^{-2}, 1)$, the vectors $(\underline{A}, \underline{K}, \underline{\sigma}^2)$ for which $\underline{A} \leq 10^{-1}$, $\underline{K} = 10^{-2}$, and $\underline{\sigma}^2 = 1$ fall in this category. Note that for these vectors the value of $\underline{A}$ is less than A by an order of magnitude or more. Thus, as this example indicates, if the initiation vector for the algorithm has a value for $\tilde{A}_0$ which is less than the true A by an order of magnitude or more, then it is possible that the probability of error incurred in the decision-making process will be relatively large in the initial stages of the algorithm. Consequently, we replace condition (3) with the more stringent condition that

(3′) $\tilde{A}_0$ must either provide an accurate estimate of A, or, $\tilde{A}_0$ must provide an estimate of A for which $\tilde{A}_0 < A$ and not less by an order of magnitude or more.

Note that the region, defined jointly by conditions (1) and (3′), from which $(\tilde{A}_0, \tilde{\Gamma}_0)$ can be chosen consists of a trapezoid. Now, the BDD algorithm directly provides initial estimates for the parameters A and K, not A and Γ. Thus, it is necessary to translate restrictions (1) and (3′) on $\tilde{\Gamma}_0$ and $\tilde{A}_0$, respectively, into restrictions on $\tilde{K}_0$ and $\tilde{A}_0$. Given that $\tilde{\Gamma}_0$ is the ratio of $\tilde{K}_0$ to $\tilde{A}_0$, restrictions (1) and (3′) do not yield a restriction for $\tilde{K}_0$ in terms of a simple inequality, as they did for $\tilde{\Gamma}_0$ and $\tilde{A}_0$. Consequently, instead of attempting to obtain estimates $(\tilde{A}_0, \tilde{K}_0)$ such that the associated estimates $(\tilde{A}_0, \tilde{\Gamma}_0)$ lie in the aforementioned trapezoid, we will instead attempt to obtain estimates $(\tilde{A}_0, \tilde{K}_0)$ for which $\tilde{A}_0$ satisfies (3′) and for which $\tilde{K}_0$ satisfies the following restriction:

(1′) $\tilde{K}_0$ must either provide an accurate estimate of K, or, $\tilde{K}_0$ must provide an estimate of K for which $\tilde{K}_0 > K$,

i.e., the set of vectors from which the estimates $(\tilde{A}_0, \tilde{K}_0)$ can be chosen is simply the largest rectangle contained in the aforementioned trapezoid. Note that restriction (1′) on $\tilde{K}_0$ is satisfied if we set $\tilde{K}_0$ to the maximum allowable value for the parameter K. Note, in addition, from the tables given for P_e, that for a fixed value of $\underline{A} \leqslant A$, $P_e((\underline{A}, \underline{K}, \underline{\sigma^2}); (A, K, \sigma^2))$ decreases with decreasing values of $\underline{K} > K$. Thus, to use this property of P_e more advantageously, we will modify restriction (1′) as follows :

(1″) $\tilde{K}_0$ must either provide an accurate estimate of K, or, $\tilde{K}_0$ must provide an estimate of K for which $\tilde{K}_0 > K$ and not greater by two orders of magnitude or more.

In summary then, we see from (1″), (2), and (3′) and the above discussion that the following restrictions (on $(\tilde{A}_0, \tilde{K}_0, \tilde{\sigma}_0^2)$) and modifications must be imposed:

(R1) $\tilde{A}_0$ must either provide an accurate estimate of A, or, $\tilde{A}_0$ must provide an estimate of A for which $\tilde{A}_0 < A$ and not less by an order of magnitude or more.

(R2) $\tilde{K}_0$ must either provide an accurate estimate of K, or, $\tilde{K}_0$ must provide an estimate of K for which $\tilde{K}_0 > K$ and not greater by two orders of magnitude or more.

(R3) $\tilde{\sigma}_0^2$ must provide an accurate estimate of σ^2.

(M1) The estimator of σ^2 given by Eq. (4.8c) must be replaced by an estimator of σ^2 consisting of an update equation for $\tilde{\sigma}_0^2$.

(M2) The estimate of A must be fixed to its initial value $\tilde{A}_0$ in the initial stages of the algorithm, with only the estimates of K and σ^2 being updated.

Table A.1. $\tau_{opt}^{(\underline{A},\underline{K},\underline{\sigma}^2)}$ FOR $(\underline{A},\underline{K},\underline{\sigma}^2) \in \Omega$

$\underline{A}$ \ $\underline{K}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	2.1575	1.0235	0.3698	0.1269	4.2914×10^{-2}
10^{-1}	0.7955	0.3017	0.1071	3.7134×10^{-2}	1.2686×10^{-2}
1	0.2081	8.1435×10^{-2}	2.9890×10^{-2}	1.0600×10^{-2}	3.6796×10^{-3}

Table A.2. $P_e((\underline{A}, \underline{K}, \underline{\sigma^2}); (A, K, \sigma^2))$ FOR $(\underline{A}, \underline{K}, \underline{\sigma^2}) \in \Omega$ AND $(A, K, \sigma^2) = (10^{-2}, 10^{-2}, 1)$

$\underline{A}$ \ $\underline{K}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	9.6367×10^{-4}	0.1220	0.7531	0.9587	0.9864
10^{-1}	0.2794	0.8253	0.9676	0.9873	0.9897
1	0.9079	0.9770	0.9883	0.9898	0.9900

Table A.3. $P_e((\underline{A}, \underline{K}, \underline{\sigma^2}); (A, K, \sigma^2))$ FOR $(\underline{A}, \underline{K}, \underline{\sigma^2}) \in \Omega$ AND $(A, K, \sigma^2) = (10^{-2}, 10^{-3}, 1)$

$\underline{A}$ \ $\underline{K}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	4.9496×10^{-4}	1.2340×10^{-4}	0.2199	0.8294	0.9702
10^{-1}	1.0081×10^{-3}	0.3637	0.8726	0.9751	0.9883
1	0.6148	0.9204	0.9804	0.9888	0.9899

Table A.4. $P_e((\underline{A}, \underline{K}, \underline{\sigma^2}); (A, K, \sigma^2))$ FOR $(\underline{A}, \underline{K}, \underline{\sigma^2}) \in \Omega$ AND $(A, K, \sigma^2) = (10^{-2}, 10^{-4}, 1)$

$\underline{A}$ \ $\underline{K}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	4.5580×10^{-4}	1.0445×10^{-4}	1.4692×10^{-5}	0.1947	0.8220
10^{-1}	6.3223×10^{-5}	1.0964×10^{-4}	0.3106	0.8613	0.9741
1	1.2462×10^{-2}	0.5067	0.9046	0.9789	0.9887

Table A.5. $P_e((\underline{A}, \underline{K}, \underline{\sigma}^2); (A, K, \sigma^2))$ FOR $(\underline{A}, \underline{K}, \underline{\sigma}^2) \in \Omega$ AND $(A, K, \sigma^2) = (10^{-2}, 10^{-5}, 1)$

$\underline{A}$ \ $\underline{K}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	4.5188×10^{-4}	1.0353×10^{-4}	1.3580×10^{-5}	1.6986×10^{-6}	0.1567
10^{-1}	6.2667×10^{-5}	9.0409×10^{-6}	1.1281×10^{-5}	0.2490	0.8427
1	4.3031×10^{-6}	1.2968×10^{-3}	0.4048	0.8847	0.9767

Table A.6. $P_e((\underline{A}, \underline{K}, \underline{\sigma}^2); (A, K, \sigma^2))$ FOR $(\underline{A}, \underline{K}, \underline{\sigma}^2) \in \Omega$ AND $(A, K, \sigma^2) = (10^{-2}, 10^{-6}, 1)$

$\underline{A}$ \ $\underline{K}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	4.5148×10^{-4}	1.0344×10^{-4}	1.3568×10^{-5}	1.5980×10^{-6}	1.9273×10^{-7}
10^{-1}	6.2612×10^{-5}	9.0329×10^{-6}	1.1392×10^{-6}	1.1521×10^{-6}	0.1980
1	4.2992×10^{-6}	6.5826×10^{-7}	1.3045×10^{-4}	0.3218	0.8647

Table A.7. $P_e((\underline{A}, \underline{K}, \underline{\sigma^2}):(A, K, \sigma^2))$ FOR $(\underline{A}, \underline{K}, \underline{\sigma^2}) \in \Omega$ AND $(A, K, \sigma^2) = (10^{-1}, 10^{-2}, 1)$

$\underline{A}$ \ $\underline{K}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	3.7025×10^{-2}	1.0023×10^{-2}	0.2024	0.7581	0.8867
10^{-1}	7.0416×10^{-3}	0.3333	0.7976	0.8912	0.9032
1	0.5623	0.8412	0.8960	0.9037	0.9047

Table A.8. $P_e((\underline{A}, \underline{K}, \underline{\sigma^2}):(A, K, \sigma^2))$ FOR $(\underline{A}, \underline{K}, \underline{\sigma^2}) \in \Omega$ AND $(A, K, \sigma^2) = (10^{-1}, 10^{-3}, 1)$

$\underline{A}$ \ $\underline{K}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	3.4883×10^{-2}	9.3140×10^{-3}	1.2729×10^{-3}	0.1781	0.7513
10^{-1}	5.7418×10^{-3}	9.4046×10^{-4}	0.2840	0.7872	0.8903
1	1.1790×10^{-2}	0.4632	0.8268	0.8946	0.9036

Table A.9. $P_e((\underline{A}, \underline{K}, \underline{\sigma^2}):(A, K, \sigma^2))$ FOR $(\underline{A}, \underline{K}, \underline{\sigma^2}) \in \Omega$ AND $(A, K, \sigma^2) = (10^{-1}, 10^{-4}, 1)$

$\underline{A}$ \ $\underline{K}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	3.4662×10^{-2}	9.2430×10^{-3}	1.2618×10^{-3}	1.4951×10^{-4}	0.1432
10^{-1}	5.6972×10^{-3}	8.4180×10^{-4}	1.1581×10^{-4}	0.2276	0.7702
1	4.0151×10^{-4}	1.2462×10^{-3}	0.3700	0.8086	0.8927

Table A.10. $P_e((\underline{A}, \underline{K}, \underline{\sigma}^2); (A, K, \sigma^2))$ FOR $(\underline{A}, \underline{K}, \underline{\sigma}^2) \in \Omega$ AND $(A, K, \sigma^2) = (10^{-1}, 10^{-5}, 1)$

$\underline{A}$ \ $\underline{K}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	3.4640×10^{-2}	9.2359×10^{-3}	1.2608×10^{-3}	1.4929×10^{-4}	1.7099×10^{-5}
10^{-1}	5.6927×10^{-3}	8.4112×10^{-4}	1.0646×10^{-4}	1.3724×10^{-5}	0.1810
1	4.0118×10^{-4}	6.1526×10^{-5}	1.2744×10^{-4}	0.2941	0.7902

Table A.11. $P_e((\underline{A}, \underline{K}, \underline{\sigma}^2); (A, K, \sigma^2))$ FOR $(\underline{A}, \underline{K}, \underline{\sigma}^2) \in \Omega$ AND $(A, K, \sigma^2) = (10^{-1}, 10^{-6}, 1)$

$\underline{A}$ \ $\underline{K}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	3.4637×10^{-2}	9.2352×10^{-3}	1.2607×10^{-3}	1.4928×10^{-4}	1.7088×10^{-5}
10^{-1}	5.6922×10^{-3}	8.4106×10^{-4}	1.0645×10^{-4}	1.2795×10^{-5}	1.5862×10^{-6}
1	4.0115×10^{-4}	6.1521×10^{-5}	8.2904×10^{-6}	1.2966×10^{-5}	0.2336

Table A.12. $P_e((\underline{A}, \underline{K}, \underline{\sigma^2}); (A, K, \sigma^2))$ FOR $(\underline{A}, \underline{K}, \underline{\sigma^2}) \in \Omega$ AND $(A, K, \sigma^2) = (1, 10^{-2}, 1)$

$\underline{A}$ \ $\underline{K}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	0.5918	0.3365	6.2612×10^{-2}	8.0124×10^{-2}	0.3063
10^{-1}	0.2370	4.2532×10^{-2}	0.1210	0.3207	0.3620
1	2.5268×10^{-2}	0.1915	0.3366	0.3638	0.3674

Table A.13. $P_e((\underline{A}, \underline{K}, \underline{\sigma^2}); (A, K, \sigma^2))$ FOR $(\underline{A}, \underline{K}, \underline{\sigma^2}) \in \Omega$ AND $(A, K, \sigma^2) = (1, 10^{-3}, 1)$

$\underline{A}$ \ $\underline{K}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	0.5915	0.3361	6.2539×10^{-2}	7.7522×10^{-3}	5.9115×10^{-2}
10^{-1}	0.2367	4.2445×10^{-2}	5.5414×10^{-3}	9.3193×10^{-2}	0.3132
1	2.0615×10^{-2}	3.6880×10^{-3}	0.1509	0.3288	0.3629

Table A.14. $P_e((\underline{A}, \underline{K}, \underline{\sigma^2}); (A, K, \sigma^2))$ FOR $(\underline{A}, \underline{K}, \underline{\sigma^2}) \in \Omega$ AND $(A, K, \sigma^2) = (1, 10^{-4}, 1)$

$\underline{A}$ \ $\underline{K}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	0.5914	0.3361	6.2531×10^{-2}	7.7513×10^{-3}	8.9218×10^{-4}
10^{-1}	0.2367	4.2440×10^{-2}	5.5370×10^{-3}	6.6852×10^{-4}	7.3654×10^{-2}
1	2.0613×10^{-2}	3.2060×10^{-3}	4.8144×10^{-4}	0.1196	0.3213

Table A.15. $P_e((\underline{A}, \underline{K}, \underline{\sigma}^2); (A, K, \sigma^2))$ FOR $(\underline{A}, \underline{K}, \underline{\sigma}^2) \in \Omega$ AND $(A, K, \sigma^2) = (1, 10^{-5}, 1)$

$\underline{A}$ \ $\underline{K}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	0.5914	0.3361	6.2531×10^{-2}	7.7512×10^{-3}	8.9216×10^{-4}
10^{-1}	0.2367	4.2440×10^{-2}	5.5369×10^{-3}	6.6813×10^{-4}	7.8054×10^{-5}
1	2.0613×10^{-2}	3.2060×10^{-3}	4.3299×10^{-4}	5.9325×10^{-5}	9.4999×10^{-2}

Table A.16. $P_e((\underline{A}, \underline{K}, \underline{\sigma}^2); (A, K, \sigma^2))$ FOR $(\underline{A}, \underline{K}, \underline{\sigma}^2) \in \Omega$ AND $(A, K, \sigma^2) = (1, 10^{-6}, 1)$

$\underline{A}$ \ $\underline{K}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
10^{-2}	0.5914	0.3361	6.2531×10^{-2}	7.7512×10^{-3}	8.9216×10^{-4}
10^{-1}	0.2367	4.2440×10^{-2}	5.5369×10^{-3}	6.6813×10^{-4}	7.8017×10^{-5}
1	2.0613×10^{-2}	3.2060×10^{-3}	4.3299×10^{-4}	5.4477×10^{-5}	7.0491×10^{-6}

APPENDIX B. EXPRESSIONS FOR THE COEFFICIENTS OF EQ. (5.28)

Let α and a_{ij} $(i = 1,\ldots, n\,; j = 1, 2,\ldots)$ be defined as in Section 5.3. Futhermore, let

$$\beta_1 \triangleq \sum_{i=1}^{n} (1 + z_i^2)\, a_{i1} \ .$$

$$\beta_2 \triangleq \sum_{i=1}^{n} (1 + z_i^2) \sum_{j=2}^{\infty} \frac{a_{ij}}{j-1} \ .$$

$$\beta_3 \triangleq \sum_{i=1}^{n} z_i^2 a_{i1} \ .$$

$$\beta_4 \triangleq \sum_{i=1}^{n} \sum_{j=2}^{\infty} z_i^2 \frac{a_{ij}}{(j-1)^2} \ .$$

and

$$\beta_5 \triangleq \sum_{i=1}^{n} \sum_{j=2}^{\infty} z_i^2 \frac{a_{ij}}{(j-1)} \ .$$

Then,

$$c_1 = \beta_4^2 n^2 - \beta_2 \beta_4 n (\beta_5 + n) + \beta_4 n (\beta_5 + n)^2 \ .$$

$$\begin{aligned} c_2 = {} & (-\beta_2 \beta_4 n^2 - \beta_2 \beta_3 \beta_4 n - \beta_2 \beta_4 \alpha n) \\ & + (2\beta_3 \beta_4 n - \beta_1 \beta_4 n + \beta_2^2 n + 2\beta_4 \alpha n)(\beta_5 + n) \\ & - (\beta_2 n)(\beta_5 + n)^2 \ . \end{aligned}$$

$$\begin{aligned} c_3 = {} & (\beta_4 n^3 + 2\beta_3 \beta_4 n^2 - \beta_1 \beta_4 n^2 + 2\beta_4 \alpha n^2) \\ & + (\beta_3^2 \beta_4 n + \beta_2^2 \beta_3 n - \beta_1 \beta_3 \beta_4 n - \beta_1 \beta_4 \alpha n + 2\beta_3 \beta_4 \alpha n + \beta_4 \alpha^2 n) \\ & + (-\beta_2 n^2 - 3\beta_2 \beta_3 n + 2\beta_1 \beta_2 n - \beta_2 \alpha n)(\beta_5 + n) \\ & + (n^2 + \beta_3 n - \beta_1 n)(\beta_5 + n)^2 \end{aligned}$$

$$\begin{aligned} c_4 = {} & (-2\beta_2 \beta_3 n^2 - 2\beta_2 \beta_3^2 n + 2\beta_1 \beta_2 \beta_3 n - 2\beta_2 \beta_3 \alpha n) \\ & + (2\beta_3 n^2 - \beta_1 n^2 - 3\beta_1 \beta_3 n + 2\beta_3^2 n + \beta_1^2 n - \beta_1 \alpha n + 2\beta_3 \alpha n)(\beta_5 + n) \end{aligned}$$

and

$$\begin{aligned} c_5 = {} & (\beta_3 n^3 - 2\beta_1 \beta_3 n^2 + 2\beta_3^2 n^2 + 2\beta_3 \alpha n^2) \\ & + (\beta_1^2 \beta_3 n - 2\beta_1 \beta_3^2 n + \beta_3^3 n - 2\beta_1 \beta_3 \alpha n + 2\beta_3^2 \alpha n + \beta_3 \alpha^2 n) \ . \end{aligned}$$

REFERENCES

[1] D. Middleton, "Canonical non-Gaussian noise models : their implications for measurement and for prediction of receiver performance," *IEEE Transactions on Electromagnetic Compatibility*, vol. EMC-21, pp. 209-220, August 1979.

[2] A. D. Spaulding and D. Middleton, "Optimum reception in an impulsive interference environment," Office of Telecommunications Report OT-76-67, June 1975 (NTIS Order No. COM 75-11097 /AS).

[3] A. D. Spaulding and D. Middleton, "Optimum reception in an impulsive interference environment, part I : coherent detection, part II : incoherent reception, *IEEE Transactions on Communications*, vol. COM-25, pp. 910-934, Sept. 1977.

[4] D. Middleton, "Threshold detection in non-Gaussian interference environments: exposition and interpretation of new results for EMC applications," *IEEE Transactions on Electromagnetic Compatibility*, vol. EMC-26, no. 1, pp. 19-28, February 1984.

[5] D. Middleton and A. D. Spaulding, "A tutorial review of elements of weak signal detection in non-Gaussian EMI environments," National Telecommunications and Information Administration (NTIA), U.S. Department of Commerce, NTIA Report 86-194, May 1986.

[6] D. Middleton, "Statistical-physical models of urban radio-noise environments - part I : foundations," *IEEE Transactions on Electromagnetic Compatibility*, vol. EMC-14, no. 2, pp. 38-56, 1972.

[7] D. Middleton, "Statistical-physical models of electromagnetic interference," *IEEE Transactions on Electromagnetic Compatibility*, vol. EMC-19, no. 3, pp. 106-127, August 1977.

[8] D. Middleton, "Canonical and quasi-canonical probability models of class A interference," *IEEE Transactions on Electromagnetic Compatibility*, vol. EMC-25, no. 2, pp. 76-106, May 1983.

[9] E. N. Skomal, *Man-Made Radio Noise*. Van Nostrand Reinhold: New York, NY, 1978.

[10] D. Middleton, "Statistical-physical models of man-made and natural radio noise, part IV: determination of the first-order parameters of class A and class B interference," National Telecommunications and Information Administration (NTIA), U.S. Department of Commerce, Technical Report NTIA-CR-2, Sept. 1978.

[11] D. Middleton, "Procedures for determining the parameters of the first-order canonical models of class A and class B electromagnetic interference," *IEEE Transactions on Electromagnetic Compatibility*, vol. EMC-21, no. 3, pp. 190-208, August 1979.

[12] D. R. Powell and G. R. Wilson, "Class A modeling of ocean acoustic noise processes," Technical Report ARL-TP-84-29, Applied Research Laboratories, University of Texas at Austin, October 1984.

[13] S. M. Zabin, "Parameter estimation for the class A Middleton model," Technical Report R-1048, UILU-ENG 85-2223, Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, July 1985.

[14] E. L. Lehmann, *Theory of Point Estimation*. Wiley: New York, 1983.

[15] C. R. Rao, *Linear Statistical Interference and Its Applications*, 2nd Ed. Wiley: New York, 1973.

[16] J. H. Miller and J. B. Thomas, "The detection of signals in impulsive noise modeled as a mixture process," *IEEE Transactions on Communications*, vol. COM-24, no. 5, pp. 559-563, May 1976.

[17] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, pp. 1-38, 1977.

[18] C. F. Jeff Wu, "On the convergence properties of the EM algorithm," *The Annals of Statistics*, vol. 11, pp. 95-103, 1983.

VITA

Serena M. Zabin was born in Chicago, IL on September 2, 1962. She received the B.S. (with honors) and M.S. degrees in Electrical Engineering from the University of Illinois at Urbana-Champaign in 1983 and 1985, respectively. She has held summer positions with the Electrical Engineering Research Laboratory at the University of Illinois at Urbana-Champaign, in 1980, and with IBM, in 1982. In 1983, she was a Teaching Assistant in the Department of Electrical and Computer Engineering at the University of Illinois and has been a research assistant in the Coordinated Science Laboratory at the University of Illinois since 1983.

Miss Zabin has been designated a James Scholar at the University of Illinois at Urbana-Champaign. She is a recipient of the Joseph Blazek Math and Science Scholarship, and in 1985 she was awarded a departmental fellowship in Electrical and Computer Engineering at the University of Illinois. She has served on various committees, including the 1981 Engineering Executives Forum Committee, sponsored by Tau Beta Pi. She is a member of Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi Honor Societies, and has been selected for membership in the Golden Key Honor Society.

PUBLICATIONS

"Parameter Estimation for Middleton Class A Interference Processes," To appear in *IEEE Transactions on Communications*, vol. COM-37, 1989.

"Recursive Algorithms for Identification of Impulsive Noise Channels," To appear in *IEEE Transactions on Information Theory*.

"Identification of Impulsive Electromagnetic Interference Channels," *Industrial Affiliates Program in Communications, Antennas, and Propagation, Eighth Annual Workshop*, University of Illinois at Urbana-Champaign, April 1986, p. 18.

"Parameter Estimation for the Class A Middleton Model," *Abstracts of Papers: 1986 IEEE International Symposium on Information Theory*, Ann Arbor, MI, October 1986, p. 28.

"Recursive Estimation of the Class A Noise Model Parameters," *Proceedings of the 22nd Annual Conference on Information Sciences and Systems*, Princeton University, Princeton, NJ, March 1988, p. 178.

"New Algorithms for the Identification of Impulsive Noise," *Abstracts of Papers: 1988 IEEE International Symposium on Information Theory*, Kobe, Japan, June 1988, p. 31.

"Estimation of Impulsive-Channel Parameters via the EM Algorithm," *Proceedings of the 23rd Annual Conference on Information Sciences and Systems*, The Johns Hopkins University, Baltimore, MD, March 1989, p. 301.

"Optimum Identification of Impulsive Interference Channels," *Industrial Affiliates Program in Communications, Electromagnetics, Propagation, and Signal Processing, Eleventh Annual Workshop*, University of Illinois at Urbana-Champaign, April 1989, p. 27.

"Parameter Estimation for the Class A Middleton Model," M.S. Thesis, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign. Also Technical Report No. R-1048, UILU-ENG85-2223, July 1985.