NOTICE

AVIS

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

If pages are missing, contact the university which granted the degree.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

Quickest Detection of a Change in a Random Sequence

with Application to Adaptive Identification of Fading Channels

by

Yong Liu

A thesis submitted to the Department of Electrical

Engineering in conformity with the requirements

for the degree of Doctor of Philosophy

Queen's University

Kingston, Ontario, Canada

June 1993

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Canada

Name __Yong Liu__

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

__Electronics and Electrical Engineering__
SUBJECT TERM

| 0 | 5 | 4 | 4 | **U·M·I**

SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523

Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
General 0679
Ancient 0289
Linguistics 0290
Modern 0291
Literature
General 0401
Classical 0294
Comparative 0295
Medieval 0297
Modern 0298
African 0316
American 0591
Asian 0305
Canadian (English) 0352
Canadian (French) 0355
English 0593
Germanic 0311
Latin American 0312
Middle Eastern 0315
Romance 0313
Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
General 0318
Biblical Studies 0321
Clergy 0319
History of 0320
Philosophy of 0322
Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
Archaeology 0324
Cultural 0326
Physical 0327
Business Administration
General 0310
Accounting 0272
Banking 0770
Management 0454
Marketing 0338
Canadian Studies 0385
Economics
General 0501
Agricultural 0503
Commerce-Business 0505
Finance 0508
History 0509
Labor 0510
Theory 0511
Folklore 0358
Geography 0366
Gerontology 0351
History
General 0578

Ancient 0579
Medieval 0581
Modern 0582
Black 0328
African 0331
Asia, Australia and Oceania 0332
Canadian 0334
European 0335
Latin American 0336
Middle Eastern 0333
United States 0337
History of Science 0585
Law 0398
Political Science
General 0615
International Law and
Relations 0616
Public Administration 0617
Recreation 0814
Social Work 0452
Sociology
General 0626
Criminology and Penology ... 0627
Demography 0938
Ethnic and Racial Studies 0631
Individual and Family
Studies 0628
Industrial and Labor
Relations 0629
Public and Social Welfare 0630
Social Structure and
Development 0700
Theory and Methods 0344
Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
General 0473
Agronomy 0285
Animal Culture and
Nutrition 0475
Animal Pathology 0476
Food Science and
Technology 0359
Forestry and Wildlife 0478
Plant Culture 0479
Plant Pathology 0480
Plant Physiology 0817
Range Management 0777
Wood Technology 0746
Biology
General 0306
Anatomy 0287
Biostatistics 0308
Botany 0309
Cell 0379
Ecology 0329
Entomology 0353
Genetics 0369
Limnology 0793
Microbiology 0410
Molecular 0307
Neuroscience 0317
Oceanography 0416
Physiology 0433
Radiation 0821
Veterinary Science 0778
Zoology 0472
Biophysics
General 0786
Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996

Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
General 0566
Audiology 0300
Chemotherapy 0992
Dentistry 0567
Education 0350
Hospital Management 0769
Human Development 0758
Immunology 0982
Medicine and Surgery 0564
Mental Health 0347
Nursing 0569
Nutrition 0570
Obstetrics and Gynecology ... 0380
Occupational Health and
Therapy 0354
Ophthalmology 0381
Pathology 0571
Pharmacology 0419
Pharmacy 0572
Physical Therapy 0382
Public Health 0573
Radiology 0574
Recreation 0575

Speech Pathology 0460
Toxicology 0383
Home Economics 0386

PHYSICAL SCIENCES
Pure Sciences
Chemistry
General 0485
Agricultural 0749
Analytical 0486
Biochemistry 0487
Inorganic 0488
Nuclear 0738
Organic 0490
Pharmaceutical 0491
Physical 0494
Polymer 0495
Radiation 0754
Mathematics 0405
Physics
General 0605
Acoustics 0986
Astronomy and
Astrophysics 0606
Atmospheric Science 0608
Atomic 0748
Electronics and Electricity 0607
Elementary Particles and
High Energy 0798
Fluid and Plasma 0759
Molecular 0609
Nuclear 0610
Optics 0752
Radiation 0756
Solid State 0611
Statistics 0463

Applied Sciences
Applied Mechanics 0346
Computer Science 0984

Engineering
General 0537
Aerospace 0538
Agricultural 0539
Automotive 0540
Biomedical 0541
Chemical 0542
Civil 0543
Electronics and Electrical 0544
Heat and Thermodynamics 0348
Hydraulic 0545
Industrial 0546
Marine 0547
Materials Science 0794
Mechanical 0548
Metallurgy 0743
Mining 0551
Nuclear 0552
Packaging 0549
Petroleum 0765
Sanitary and Municipal 0554
System Science 0790
Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

To my mother and in memory of my father.

Abstract

This thesis addresses the problem of quickest detection of an abrupt change in a random sequence and its application to adaptive identification of fading channels. This investigation is motivated by a wide variety of applications of quickest change-detection and the practical application of adaptive fading-channel identification to mobile communications. A new procedure, referred to as the minimum asymptotic risk (MAR) procedure, is proposed. It is shown that the MAR procedure converges almost surely to a theoretically optimum procedure under certain asymptotic conditions. It is also shown by using extensive simulations that the MAR procedure outperforms existing procedures in non-asymptotic situations, where a modified version of Shiryayev's criterion is used. False alarm probability bounds of change-detection procedures are derived. The MAR procedure is also extended to address change detection with unknown parameters, as well as nonstationary observations. Included is a proof of the optimality of the sequential probability ratio test for nonstationary observations using Bayesian analysis, as well as several useful properties of its optimum time-varying thresholds.

The quickest change-detection is then applied to adaptive identification of fading channels. An algorithm combining decision feedback and adaptive linear prediction (DFALP) is first proposed to adaptively track the phase and amplitude of fading channels. It is shown by both extensive simulations and theoretical studies that the DFALP algorithm with coherent phase shift keying (PSK) outperforms its differential PSK counterpart and an existing fading channel tracker. For fading channels experiencing abrupt changes of channel statistics, a change detector is used to monitor the nonstationarity of the channel statistics, including the vehicle speed. Improved bit-error-rate performance is achieved by incorporating the change detector with the DFALP channel tracker for nonstationary fading channels.

Acknowledgements

Contents

List of Tables

ix

List of Figures

List of Symbols

A_0^2 =Normal power level of fading channel

A_1^2 =Power level below which a fading channel enters a deep fade

AMH=Alternative MH approach

AR=Autoregressive

ARQ=Automatic retransmission request

ARMA=Autoregressive and moving average

AWGN=Additive white Gaussian noise

B =Decision threshold for CUSUM

B_0 =Decision threshold for CUSUM to detect a deep fade

B_1 =Decision threshold for CUSUM to detect a recovery from a deep fade

BER=bit-error-rate

CPSK=Coherent PSK

CSI=Channel state information

CUSUM=Cumulative sum

D =Set of digital signal constellation

D_f=Delay of the DFALP algorithm

$2D_f + 1$=Low pass filter order for DFALP and SADD

DFALP=Decision feedback and adaptive linear predictive (fading channel tracker)

DPSK=Differential PSK

$E_m()$ =Taking expectation under the hypothesis H_m

$F()$ =Cumulative distribution function

FAP=False alarm probability

FEC=Forward error correcting

FIR=Finite impulse response (filter)

FSS=Fixed sample size (procedure)

GLR=Generalized likelihood ratio approach

xviii

GRS=Girshick-Rubin-Shiryayev procedure

H_i =Hypothesis i

H_m =Hypothesis that a change occurs at the mth sample

$I_f = E_m(\ln l_m)$

K =Ratio of the line-of-sight power to the scattering power for Rician fading channels

K_t=Training period of the DFALP and SADD algorithms

LPF=Low pass filter

M =Random change time

MAR=Minimum asymptotic risk procedure

MH=Multiple hypothesis approach

MLE=Maximum likelihood estimation

MLR=Maximum likelihood ratio approach

N =Random stopping time

N=Linear predictor order

$\hat{N}$=Stopping variable of the MAR procedure

N_b=Theoretically optimum stopping variable

NSPRT=SPRT for nonstationary observations

$P()$ =Probability

P_e=Bit error rate

$P_m()$ =Probability under hypothesis H_m

PSK=Phase shift keying

$Q()$ =Tail function of the standard Gaussian distribution

QPSK=Quaternary PSK

$R()$ =Risk of a decision rule

R_c =Covariance matrix of the fading channel gain vector $\bar{c}$

$S_i^k = l_i...l_k$

SADD=symbol-aided plus decision-directed fading channel tracker

SNR=Signal to noise ratio

SPRT=sequential probability ratio test

T=Symbol period

T_n =CUSUM statistic

T =Transpose of a matrix or vector

TCM=Trellis-coded modulation

W=Number of simulations for estimating the performance of change-detection procedures

X_i =A scalar random variable

$X_i^k = (X_i, ..., X_k)^T$

$a = E\{c_k\}$

arg()=Taking the phase of a complex number

$\bar{b}(k)$-Vector of linear predictor coefficients at time k

b_i=Linear predictor coefficient

c_0 =Design parameter for MAR

c_k =Complex fading channel gain at time k

$c_{i,k}$ =Complex fading channel gain of the (equivalent) ith beam at time k

$\bar{c}_k = (c_{k-1}, ..., c_{k-N})^T$

$\hat{c}_k$=Predicted fading channel gain

$\check{c}_k$=Corrected fading channel gain estimate

$\bar{c}_{k-D_f}$=Final output of DFALP tracker

f=Wave frequency

$f()$ =Probability density function

f_m=Fading channel bandwidth

$f_m T$=Normalized fading channel bandwidth

f_s=Symbol rate

h_i=Low pass filter coefficients

xx

i.i.d.=independent and identically distributed

l=Number of the equivalent beams of a frequency selective fading channel

l_i =The ith likelihood ratio

m =Actual change time

n =A realization of a random stopping time N

n_b=A realization of a random stopping time N_b

r_n =Expected risk function of an optimum decision rule in Chapters 3 and 4

$r_n = E\{(c_k - a)(c_{k-n} - a)^*\}$ in Chapters 5 and 6

$\hat{r}_n$ =MAR statistic

$\bar{r}_n$ =Asymptotic convergent value of r_n and $\hat{r}_n$

v=Vehicle speed

v_l=Speed of light

w= Window size of moving-window FSS procedure

$x_i^k = (x_i, ...x_k)^T$

y_k =Received signal plus noise over a fading channel

α=False alarm probability

γ=SNR per bit

γ_s=SNR per symbol

λ=Wavelength

μ=Step-size of adaptive linear predictor

π=Detection probability

σ_n^2=Power of the AWGN

Chapter 1

INTRODUCTION

In this chapter, the general background of signal change detection is addressed, together with a brief introduction of its applications to communications and other fields. The organization of the thesis is presented. The major contributions of this thesis are summarized.

1.1 Overview

Binary and M-ary signal detection problems have been studied since the 1930s [38, 63, 97, 117, 119]. The signal detection techniques used either fixed sample size (FSS) tests [63, 97, 117], or sequential procedures [119, 38]. The FSS tests are usually designed according to different criteria: Bayesian, Minimax, and Neyman-Pearson [97]. For certain problems, non-parametric and robust detection techniques have been developed [47, 56]. A well-known sequential procedure for solving the binary hypothesis testing problems is the sequential probability ratio test (SPRT) [38, 119].

Another important statistical signal detection problem, which differs from classical binary and M-ary hypothesis testing, is to quickly detect a change in a random sequence. It is assumed that the distributions before and after the change are (at least partially) known, while the *change-time* is unknown. The objective is to detect the change and

1

raise an alarm as soon as a change has occurred, subject to a *false alarm* constraint.
This problem has a variety of applications. A short list is as follows.

- Monitoring communication channels [90, 100].

- Radar signal detection [36, 48].

- Digital signal processing [8, 9].

- On-line image processing [6, 16, 18].

- Speech processing [4, 35].

- Adaptive parameter estimation and system identification [12, 75, 93].

- Failure detection [118, 128].

- Industrial, chemical and mechanical engineering [41, 42, 50, 102].

This problem was first studied in quality control applications [29, 109] in the 1930s. Since then there have been many well-documented studies of the change-detection problem, some of which are surveyed in [9, 42, 111]. However, there are still many open problems in this area. This thesis is intended to solve certain key open problems and to apply the results to adaptive identification of fading channels encountered in mobile communications.

A number of change-detection algorithms have been proposed. Among those, three procedures have attracted much attention. The first one is a *moving-window fixed sample size (FSS) procedure* [92]. The well-known $\bar{X}$-*chart* is a special case of the FSS approach for detecting shift of mean in the Gaussian distribution. The second procedure due to Page [89] is a repeated sequential probability ratio test *(SPRT)* with the lower decision threshold set to 1, and the upper threshold greater than 1, which is known as *CUSUM.* The third procedure, first proposed by Girshick and Rubin [39] in 1952 and later independently proposed by Shiryayev [110, 111], is based on an a priori distribution

2

of the unknown random change-time. We refer to the third procedure as the *GRS procedure.*

Different criteria have been used by researchers. As a summary, we list three criteria used frequently in published literature. The first one, formalized by Shiryayev [110, 111], is to minimize the *expected delay* subject to a given *probability of false alarm,* i.e., raising an alarm before the actual change-time, where the unknown change-time is assumed to be a random variable. We refer to this criterion as *Shiryayev's criterion.* The second criterion, widely used in quality control engineering [102], is to minimize the expected delay subject to a given *false alarm average run length (ARL),* i.e., the average run length under the condition that no change ever occurs, where the change-time is assumed to be an unknown non-random parameter. This criterion is referred to as the *QC criterion* in this thesis. The third criterion, first proposed by Lorden [77] and very popular among theoretical statisticians, is a *minimax criterion,* which minimizes an *upper bound of expected delay* among all change-times subject to a given false alarm average run length under the condition that no change ever occurs. This last criterion is referred to as *Lorden's criterion.*

It is interesting that the above three criteria differ from one another, and performance results depend upon the choice of criterion. It is noted that the false alarm probability constraint in Shiryayev's criterion is meaningful only if the change-time is finite. This property may restrict its applications in certain situations. For example, in search radar it is possible that a target may never appear in a given area, i.e., that the change-time may be infinite. In this case, the false alarm probability constraint is not well defined since it will be always equal to one for almost all existing procedures. A similar situation occurs in failure detection [128] or in quality control engineering [42]. On the other hand, both the Lorden and QC criteria do not consider the false alarm probability at all, and appear unreasonable for a number of important applications. For example, in speech segmentation [35], it is clear that a change in statistics across phonemes will occur after

3

a finite number of samples with probability one. In edge detection in digital images [6], the change-time is bounded by a finite number of pixels. In mobile communications in a fading environment, power drops and recoveries (deep fades) will occur almost surely if the vehicle is moving. In these cases, repeated change detections over moderate-size sequences are of interest. Therefore, in situations where it is of interest to quickly detect changes that are known to occur after a finite number of samples, Shiryayev's criterion may be preferred.

Using the Shiryayev criterion, it has been shown that the GRS procedure is optimal if the unknown random change-time is distributed geometrically. For non-geometric change-time distributions, the relative performance of existing procedures appears unknown thus far in terms of the Shiryayev criterion. Using the QC criterion, extensive simulations have been conducted to compare the performance of existing procedures, including the GRS, CUSUM and FSS procedures [102, 31, 41, 42]. It is found that none of these procedures are optimal, and one procedure may outperform others in certain situations and may be worse in other cases. Using Lorden's minimax criterion, it has been shown that the CUSUM procedure is optimal both asymptotically and nonasymptotically [77, 85, 101].

Modified versions of the well-known CUSUM and the FSS procedures have been considered by some researchers [17, 51, 52, 88, 92]. A number of upper and lower bounds of the CUSUM procedure were derived in [19].

Interesting applications of change-detection in communications have been reported in [90, 100]. In this thesis we will apply change-detection techniques to adaptive fading-channel identification, an important problem encountered in mobile communications. There exist a number of algorithms to extract channel information for reliable digital transmission over fading channels. A very practical and efficient approach appears to be Irvine and McLane's *symbol aided plus decision-directed (SADD)* algorithm [57, 58], which is an important extension of the symbol aided tracker proposed by Moher and

4

Lodge [82]. The SADD algorithm tracks the fading channel gain using linear interpolation of the channel estimates derived from training symbols. A decision directed approach is then used by the SADD to improve the initial channel estimates. Using trellis coded modulation (TCM), interleaving and soft Viterbi decoding, the SADD with coherent phase shift keying (CPSK) has achieved 3 dB gain over its differential PSK (DPSK) counterpart [57]. However, this gain cannot be always achieved if no coding and interleaving are used. In addition, the large delay of the SADD is a problem in a number of applications. Other channel tracking algorithms include the conventional adaptive equalizer, phase lock loop, and Kalman filtering. The adaptive equalizer and phase lock loop suffer slow convergence and false-lock, and cannot track with acceptable performance satellite channels and typical cellular telephone channels involving fast moving vehicles. The difficulties of the Kalman filtering approach include its complexity, instability, and poor adaptability to varying channel statistics (such as vehicle speed).

In addition, the above tracking algorithms assume stationarity of channel statistics, which depend on physical parameters such as vehicle speed. This assumption is not realistic in most applications. It is therefore necessary to study the fading channel tracking problem while considering the nonstationarity of channel statistics.

1.2 Organization of the Thesis

This thesis is organized as follows.

- In Chapter 1, general background concerning signal change detection and fading channel identification is provided. The organization of the thesis is then presented. Following this, the major contributions of the thesis are summarized.

- In Chapter 2, a formal statement of the problem of change detection is presented. The history and development of the research on this problem and its applications

5

are surveyed. Three criteria for measuring performance of change detection procedures are introduced and compared. The properties of three existing well-known procedures, Page's CUSUM, the Girshick-Rubin-Shiryayev (GRS) procedure, and a moving-window fixed sample size (FSS) procedure, are briefly discussed. The key open problems are summarized. A number of applications of change-detection techniques are presented, including an application to adaptive error-control for time-varying channels.

- In Chapter 3, a new procedure for quickest change detection is introduced, which minimizes an asymptotic Bayes risk function and is referred to as the *minimum asymptotic risk (MAR) procedure*. The MAR procedure has low computational complexity and does not require any prior knowledge of the unknown change time. It is shown that the MAR procedure approaches almost surely an optimum procedure under certain asymptotic conditions. The non-asymptotic properties of the MAR procedure are studied via extensive simulations. It is shown that the new MAR procedure outperforms both Page's CUSUM and an optimized FSS procedure. Simulation results also reveal that MAR performs favorably compared with GRS in most cases, if the actual unknown change-time is not geometrically distributed. The simulation results also lead to several practical observations concerning choosing the design parameters of the MAR procedure. The false alarm probability bounds of change detection procedures are studied in the final section of the chapter.

- In Chapter 4, the change detection procedures are generalized to problems with non-i.i.d observations and situations where the distributions contain unknown parameters. For non-i.i.d. observations, an approximated MAR procedure and a modified CUSUM procedure are proposed. The SPRT for non-i.i.d. observations is investigated since it has application to the change detection problem for non-i.i.d. observations. For change detection with unknown distribution parameters, two

6

multiple-hypothesis (MH) approaches and a moving-window maximum-likelihood-estimation (MLE) approach are combined with the MAR procedure derived in this thesis, which are referred to as MH MAR, AMH MAR, and MLE MAR procedures, respectively. The performance of the new procedures are compared along with an existing moving window maximum-likehood-ratio (MLR) procedure.

- In Chapter 5, general background of communications over fading channels is presented. An optimum signal detection rule for fading channels proposed by Lodge and Moher [76] is studied. It is shown that the optimum rule for constant envelope modulation over fading channels is approximately equivalent to a coherent detector combined with a minimum mean square error (MMSE) channel estimator. This result provides a motivation to finding a good fading channel estimator.

- In Chapter 6, the important problem of adaptive identification of nonstationary fading channels is addressed. Using decision feedback and adaptive linear prediction, a new algorithm (DFALP) is proposed to adaptively track the phase and amplitude of fading channels. Combined with coherent phase shift keying (CPSK) modulation/demodulation, the new algorithm is shown to outperform Irvine-McLane's symbol-aided plus decision-directed (SADD) algorithm and conventional differential PSK (DPSK). The performance of the DFALP algorithm is also studied analytically. When the fading channel experiences change of its statistics, such as vehicle speed, it is demonstrated that the change detection techniques described in Chapter 3 can be applied to adjust the parameters of the DFALP algorithm to improve its performance.

1.3 Summary of Contributions

The major contributions of the thesis may be categorized into two classes: (i) theoretical contributions to change detection; and (ii) practical contributions to adaptive

7

identification of fading channels.

1.3.1 Contributions to Change Detection

1. Proposed a new procedure (MAR procedure) for change detection.

2. Proved the asymptotic convergence of the MAR procedure to a theoretically optimum procedure.

3. Demonstrated the nonasymptotic superiority of the MAR procedure over existing procedures using simulations.

4. Computed false alarm bounds of existing change detection procedures.

5. Proved optimality of the sequential probability ratio test for nonstationary observations (NSPRT) using Bayesian analysis. Derived the properties of optimum thresholds for the NSPRT.

6. Extended the MAR procedure to address change detection with unknown distribution parameters and non-i.i.d. observations.

1.3.2 Contributions to Adaptive Identification of Fading Channels

1. Proposed a new algorithm (DFALP algorithm) to adaptively track the phase and amplitude of fading channels.

2. Demonstrated the superiority of the DFALP algorithm over the SADD algorithm and DPSK.

3. Studied the problem of choosing design parameters of the DFALP algorithm.

4. Studied the performance of the DFALP algorithm analytically.

8

5. Applied change detection techniques to monitoring changes of fading channel statistics, particularly the vehicle speed. The performance of the DFALP algorithm is improved by adjusting its parameters according to changes in channel statistics.

Chapter 2

CHANGE DETECTION PROCEDURES AND THEIR APPLICATIONS

In this chapter, a formal statement of the problem of quickest change detection is presented. Existing procedures are surveyed. Three criteria for measuring performance of change detection procedures are introduced and compared to each other. Properties of three existing well-known procedures are briefly discussed. The key open problems are summarized. Finally a number of applications are discussed with a detailed discussion of an application to adaptive error control over time-varying channels.

2.1 Problem Statement and Criteria

2.1.1 Problem Statement

Suppose $\{ X_i | i = 1, 2, ... \}$ are independent random variables observed sequentially and there exists some positive, non-random integer m such that the distribution function of X_i is $F_0(x_i)$ with density $f_0(x_i)$ for $i = 1, ... m - 1$, and $F_1(x_i)$ for $i = m, m + 1, ...$ with density $f_1(x_i)$, where F_0 and F_1 are known but m is unknown. Our objective is to stop and raise alarm after m occurs, as quickly as possible, subject to a given false alarm constraint.

10

Let $N = n$ denote a stopping time. For convenience let $X_i^n \triangleq (X_i, X_{i+1}, ..., X_n)^t$ and $x_i^n \triangleq (x_i, x_{i+1}, ..., x_n)^t$ for $i \leq n$. When $i > n$, X_i^n or x_i^n means no observations have been taken. Let $dx_i^n \triangleq dx_i dx_{i+1}...dx_n$ for $i \leq n$. For any positive, non-random integer m, we define joint distributions P_m as

$$P_m\{X_i^n \leq x_i^n\} = \begin{cases} \int_{-\infty}^{x_i^n} f_0(x_i)f_0(x_{i+1})...f_0(x_n)dx_i^n & \text{if } n < m \\ \int_{-\infty}^{x_i^n} f_0(x_i)...f_0(x_{m-1})f_1(x_m)...f_1(x_n)dx_i^n & \text{if } i < m \leq n \\ \int_{-\infty}^{x_i^n} f_1(x_i)f_1(x_{i+1})...f_1(x_n)dx_i^n & \text{if } m \leq i \leq n \end{cases} \quad (2.1.1)$$

and expectation E_m under P_m. Technically, the situation where no change ever occurs is P_∞. Often in the literature, the alternative notation $P_0 \triangleq P_\infty$ and E_0 are used. Let H_m denote the hypothesis that P_m is true. Let a_m denote the event that one accepts H_m. Define the likelihood ratio of the ith sample,

$$l_i \triangleq \frac{f_1(x_i)}{f_0(x_i)} \quad (2.1.2)$$

and the product form

$$S_k^n \triangleq \begin{cases} \prod_{i=k}^n l_i & \text{if } n \geq k \\ 1 & \text{if } n < k. \end{cases} \quad (2.1.3)$$

For any stopping time $N = n$, define

$$\Theta_n \triangleq \{X_n \sim f_1\} \qquad \overline{\Theta}_n \triangleq \{X_n \sim f_0\}. \quad (2.1.4)$$

This definition of Θ_n implies a *correct detection*, i.e., that one stops at or after the real change-time m $(n \geq m)$, $\overline{\Theta}_n$ implies a *false alarm*, i.e., that one stops before m $(n < m)$. Let

$$\pi \triangleq P(\Theta_N) \quad (2.1.5)$$

represent the probability that Θ_N is true for a stopping rule N, where $0 < \pi < 1$. We remark that π is the *correct detection probability*, i.e., the overall probability that a change detection procedure terminates at or after the change-time m, and

$$\alpha \triangleq 1 - \pi \quad (2.1.6)$$

11

is the overall *false alarm probability*. It should be noted that π and α depend on the decision rule and the unknown, non-random change time m.

Given a random sequence $\{X_1, X_2, ..., X_n, ...\}$, a *stopping variable* N is a positive integer-valued random variable defined by some stopping rule $\delta(n)$ with

$$P\{N = n\} = \psi_n \qquad (2.1.7)$$

where

$$\sum_{n=1}^{+\infty} \psi_n = 1. \qquad (2.1.8)$$

A *stopping rule* $\delta(n)$ is

$$\delta(n) \begin{cases} \in (0,1] & \text{if } X_1^n = x_1^n \in D_n \\ = 0 & \text{if } X_1^n = x_1^n \in D_n^c \end{cases} \qquad (2.1.9)$$

where D_n is a decision region, $D_n^c = R^n - D_n$, and $n = 1, 2, ...$ For a non-randomized stopping rule

$$\delta(n) \begin{cases} = 1 & \text{if } X_1^n = x_1^n \in D_n \\ = 0 & \text{if } X_1^n = x_1^n \in D_n^c. \end{cases} \qquad (2.1.10)$$

Notice that

$$\psi_n = P(N = n) = P(X_1^n \in D_n)\delta(n). \qquad (2.1.11)$$

Therefore we sometimes use ψ_n to represent a stopping rule, or simply call a stopping rule N.

If the stopping time $N = n$ is only a function of X_1^n for all $n = 1, 2, ...,$ then the stopping variable N and the stopping rule $\delta(N)$ are said to be *Markovian* or *causal*. If the stopping time $N = n$ is not only a function of X_1^n for some $n = 1, 2, ...,$ then the stopping variable N and the stopping rule $\delta(N)$ are said to be *non-Markovian* or *non-causal*. Alternatively, a causal stopping time $N = n$ is determined by X_1^n, while a non-causal stopping time $N = n$ is determined by both X_1^n and some X_k, where $k > n$ and possibly $k = +\infty$. In other words, a non-causal stopping rule depends on the future, while a causal stopping rule does not depend on the future. Most real-time detection

12

problems require a causal stopping rule, while in some off-line signal processing problems the future information may be utilized.

2.1.2 Three Existing Criteria

To compare the performance of procedures, Page [89] defined the average run length (ARL), i.e., that

$$N_d \triangleq E\{N|N \geq m\} = E\{N|\Theta_N\}, \tag{2.1.12}$$

which is the detection ARL, and

$$N_{fm} \triangleq E\{N|N < m\} = E\{N|\overline{\Theta}_N\}, \tag{2.1.13}$$

which is the false alarm ARL. The expected delay for any given change-time m is then

$$D_m \triangleq N_d - m + 1 = E\{N - m + 1|\Theta_N\}. \tag{2.1.14}$$

The definition of D_m was first given by Moustakides [85].

We now present three existing criteria for performance comparison.

- **Shiryayev Criterion:** To prove the optimality of the GRS procedure, the change time is assumed to be a random integer, which we denote by M. A realization of the unknown random change time M is $M = m$, where m is an unknown non-random change time. For a random change time M, Shiryayev ([111], Eqs.(4.130)(4.131), p.198) introduced a criterion which minimizes the average delay

$$D_M = E\{N - M + 1|N \geq M\} \tag{2.1.15}$$

subject to a given false alarm probability

$$\alpha = P(N < M) \leq \gamma, \tag{2.1.16}$$

where γ is a given constant. The above criterion is referred to as the Shiryayev criterion.

13

A slightly different criterion can be obtained if the above optimality is conditioned by $M = m$ for any $m > 0$. In this case one can simply substitute m for M into the above criterion. It should be noted that if m instead of M is used, the Shiryayev optimality is stronger since the optimality conditioned by $M = m$ implies the optimality for M, and the converse is not true. The Shiryayev criterion minimizes the expected delay subject to a given false alarm probability. It should be noted that the false alarm ARL N_{fm} is not considered in the Shiryayev criterion.

- **Lorden Criterion:** Lorden's [77] criterion is a minimax criterion, which minimizes

$$\max_{m \geq 1} D_m \qquad (2.1.17)$$

subject to

$$N_{fm}|_{m=\infty} = N_{f\infty} \geq t. \qquad (2.1.18)$$

Lorden's criterion minimizes an upper bound of the expected delay subject to a given false alarm ARL when no change ever occurs. It is noted that the false alarm probability $\alpha = P(N < m)$ is not considered in this criterion.

- **QC Criterion:** The QC (quality control) criterion used by Roberts [102] minimizes

$$D_m \qquad (2.1.19)$$

subject to

$$N_{fm}|_{m=\infty} = N_{f\infty} \geq t. \qquad (2.1.20)$$

The QC criterion minimizes an actual value of the expected delay subject to a given false alarm ARL when no change ever occurs. Different from Lorden's criterion, the QC criterion is not a minimax criterion. It should be noted that the QC criterion does not consider the false alarm probability $\alpha = P(N < m)$, and differs from the Shiryayev criterion.

14

The advantages of each criterion presented above are now briefly discussed.

First, both the Lorden and QC criteria do not consider the overall false alarm probability α. For a finite change-time m, it appears unreasonable to ignore α. This observation is motivated by the fact that a large false alarm ARL $N_{f\infty}$ alone cannot guarantee a small overall false alarm probability α for an actual finite change-time m. In a number of applications, such as speech processing, edge detection in digital images, and monitoring mobile communication channels in a congested urban environment, a change will certainly occur in a finite number of observations and the false alarm probability is an important index for measuring the performance of procedures.

Second, Lorden's criterion is a minimax criterion. A minimax procedure may not necessarily minimize the actual value of the expected delay for a given $N_{f\infty}$ and an actual change-time m, and therefore may be conservative. This was first pointed out by Moustakides [85].

Third, the Shiryayev criterion may not be well-defined if $m = \infty$, i.e. no change ever occurs. In this case, $\alpha = 1$ for almost any procedure imaginable; for example, it is easily shown that this is true for the CUSUM, GRS, and the FSS procedures if the decision threshold is finite (see section 3.6 of this thesis).

Based on the above observations, it appears unreasonable to accept one criterion (e.g. Lorden's minimax criterion) exclusively and reject others. It may instead be worth studying procedures under different criteria and choosing suitable results according to different applications.

Considering the disadvantages of the existing criteria, we will introduce a criterion which concerns both the false alarm probability and false alarm average run length. This criterion will be presented in Section 3.1.

In addition to the above three frequently used criteria, a less known criterion proposed by Bojdecki [20] is worth mentioning. This criterion maximizes the probability that one stops within a neighborhood of the real change-time, and is referred to as the probability

15

maximization criterion (PM criterion). Formally, the PM criterion maximizes

$$P\{|N - M| \leq w\} \tag{2.1.21}$$

where N is the stopping time, M is the possibly random unknown change-time, and w is a given positive integer. This problem appears nontrivial and will not be studied in this thesis.

Using a general cost function, Pelkowitz [91] studied the change-detection problem and treated the above criteria as special cases of the cost function introduced in [91]. However, practical solutions are not available for this general framework.

2.2 Existing Procedures for Change Detection

Several procedures have been proposed in published literature, including Page's CUSUM procedure [89], the moving window FSS procedure [42] and the Girshick-Rubin-Shiryayev (GRS) procedure (see [39] and [111] section 4.3). In the following subsections the three important procedures are reviewed. Their properties are briefly discussed.

2.2.1 Page's CUSUM Procedure

Page's cumulative sum (CUSUM) procedure is based on the idea of maximum likelihood ratio test. The CUSUM statistic is defined as

$$T_n \triangleq \max_{1 \leq i \leq n} S_i^n, \tag{2.2.1}$$

which can be written in a recursive form,

$$
\begin{aligned}
T_n &= \max_{1 \leq i \leq n} S_i^n \\
&= \max\{S_1^n, S_2^n, ..., S_{n-1}^n, S_n^n\} \\
&= \max\{S_1^{n-1}, S_2^{n-1}, ..., S_{n-1}^{n-1}, 1\} l_n \\
&= \max\{T_{n-1}, 1\} l_n, \tag{2.2.2}
\end{aligned}
$$

16

$$N = N_k + k - 1$$
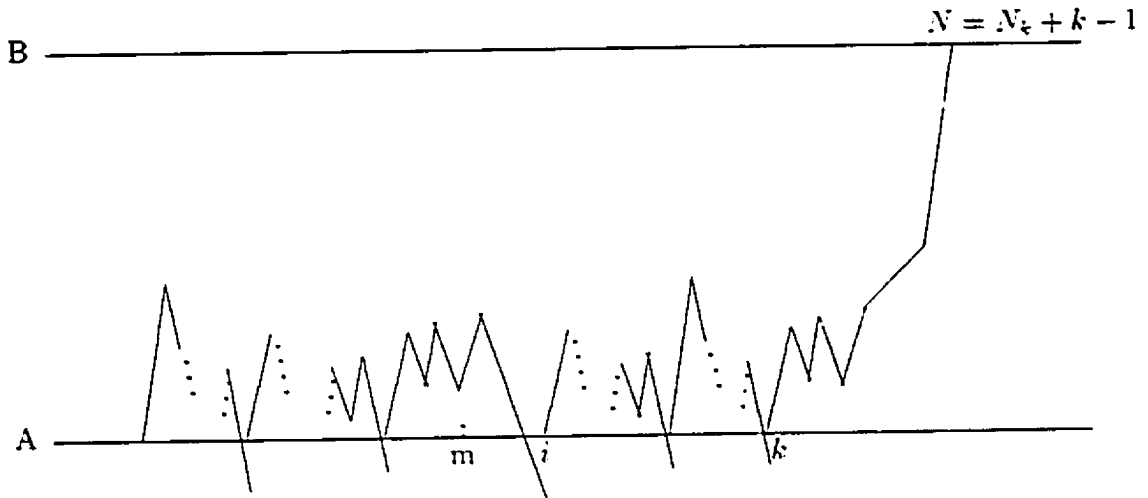
Figure 2.1: *The repeated SPRT model for Page's CUSUM procedure.*

where $n \geq 1$, and $T_0 = 1$.

The decision rule for the CUSUM procedure is to stop and decide that a change has occurred as soon as

$$T_n \geq B \tag{2.2.3}$$

for a chosen decision threshold $B(\geq 1)$.

From Eqs.(2.2.2) and (2.2.3), it is clear that the computational complexity of the CUSUM procedure is minimal.

It is observed from (2.2.2) and (2.2.3) that the CUSUM procedure is a repeated SPRT with the lower threshold set to $A = 1$ and the upper threshold B greater than 1. This model is shown in Figure 2.1. From the figure one can see that the overall false alarm probability α is the probability that $N < m$, which is much different from the false alarm probability for one of these repeated binary SPRTs.

Using Lorden's minimax criterion, it has been shown that the CUSUM procedure is optimal [77, 85, 101]. Using the QC criterion, simulation results reported in [102, 31, 41, 42] reveal that the CUSUM may be better or worse than others, depending on

17

the situations. It appears to be difficult to prove theoretical properties of the CUSUM procedure in terms of the Shiryayev criterion and the QC criterion.

As is shown in Figure 2.1, let N_k denote the sample size of the last SPRT. Define the *one-sided SPRT* over $\{X_k, X_{k+1}, ...\}$ as an SPRT which never accepts H_0, i.e., that

$$P\{\text{accept} H_0 | X_k, X_{k+1}, ...\} = 0. \tag{2.2.4}$$

It is obvious that N_k is the sample size of a one-sided SPRT. It is clear from Figure 2.1 that the total sample size for the CUSUM procedure is

$$N = N_k + k - 1. \tag{2.2.5}$$

An interesting property of the CUSUM procedure is described in the following

LEMMA 2.1 *The stopping rule of Page's CUSUM procedure*

$$N = \min_{n=1,2,...} (n|S_1^n \geq B) = \min_{n=1,2,...} (n| \max_{1 \leq j \leq n}(\prod_{i=j}^{n} \frac{f_1(x_i)}{f_0(x_i)}) \geq B), \tag{2.2.6}$$

is equivalent to the one-sided SPRT

$$\min_{j=1,2,...} (j + N_j - 1| \prod_{i=j}^{j+N_j-1} \frac{f_1(x_i)}{f_0(x_i)} \geq B), \tag{2.2.7}$$

where N_j is the sample size of a one-sided SPRT to test $f_1(x_i)$ v.s. $f_0(x_i)$ over $X_j, X_{j+1}, ...$ for $j \geq 1$. For all m, the latter procedure is bounded according to

$$D_m \leq E_m(N_m). \tag{2.2.8}$$

$\square$

PROOF: According to the definition, the total sample size of the CUSUM procedure is

$$
\begin{aligned}
N &= \min_{n=1,2,...} (n| \max_{1 \leq j \leq n}(\prod_{i=j}^{n} \frac{f_1(x_i)}{f_0(x_i)}) \geq B) \\
&= \min_{n=1,2,...} (n|\exists j \in \{1,2,...,n\} \ni \prod_{i=j}^{n} \frac{f_1(x_i)}{f_0(x_i)} \geq B)
\end{aligned}
$$

18

$$= \min_{j=1,2,...} \left(n \mid n \in \{j, j+1, ...\}, \prod_{i=j}^{n} \frac{f_1(x_i)}{f_0(x_i)} \geq B \right)$$

$$= \min_{j=1,2,...} \left(j + N_j - 1 \mid \prod_{i=j}^{j+N_j-1} \frac{f_1(x_i)}{f_0(x_i)} \geq B \right)$$

$$\triangleq N_{k'} + k' - 1.$$

where $\exists$ denotes 'there exists' and $\ni$ denotes 'such that'. It is obvious that k' equals k defined in (2.2.5) since both k and k' correspond to the starting point of the last one-sided SPRT. In particular, the above result implies that regardless of the location of the change, m,

$$N = \min_{j=1,2,...} \left(j + N_j - 1 \mid \prod_{i=j}^{j+N_j-1} \frac{f_1(x_i)}{f_0(x_i)} \geq B \right) \leq N_m + m - 1,$$

i.e.,

$$D_m = E_m(N - m + 1) \leq E_m(N_m).$$

proving the upper bound.

QED.

The above proof of Lemma 2.1 was given by Blostein and Liu [19]. Lemma 2.1 is a generalization of (11) in [77]. Lemma 2.1 implies that the expected delay of the CUSUM procedure would not exceed the expected sample size of a single one-sided SPRT with lower threshold 1 and upper threshold B, with all samples distributed according to f_1. However, since the overall false alarm probability α of the CUSUM procedure is much different from that of a single binary SPRT, the optimality of the SPRT [120] does not imply the optimality of the CUSUM procedure.

The distributions of the ARL of the CUSUM procedure have been studied by some researchers using integral equation techniques, Markov chains and Brownian motion approximations, which are surveyed in [42].

19

2.2.2 Moving Window Fixed Sample Size Procedure

A moving window fixed sample size (FSS) procedure was used for quality control applications as early as the 1920s [29, 109], where an $\overline{X}$-chart method was used. This FSS procedure has been used for a long time due to its simplicity and good performance in certain situations. The FSS procedure is described as follows.

Let w be the fixed sample size of a single binary test of H_0 v.s. H_1. Beginning from $n = 1$, $X_1, ..., X_w$ are taken and a binary hypothesis test of H_0 v.s. H_1 is performed. If H_1 is accepted, then the test is terminated. If H_0 is accepted, one more sample X_{w+1} is taken and a binary hypothesis test of H_0 v.s. H_2 is performed over $X_2, ..., X_{w+1}$. Again, if H_2 is accepted, the test is terminated. If H_0 is accepted, one more sample X_{w+2} is taken and a binary hypothesis test of H_0 v.s. H_3 is performed over the most recent w samples. This procedure is repeated until H_i ($i > 0$) is first accepted. Let the total sample size be N_{FSS}. The decision threshold for the FSS test is denoted by d. The decision rule is

$$S_n^{n+w-1} \begin{cases} \geq d & a_n \\ < d & a_0 \end{cases} \tag{2.2.9}$$

for all $n \geq 1$, where a_n stands for accepting H_n.

A comparative study of the performance between the FSS procedure (a $\overline{X}$-chart method) and Page's CUSUM procedure has been conducted by Ewan [31] and Goel [41, 42] via simulations, where the QC criterion was used. It is found that neither the CUSUM nor the FSS is optimal, and that the CUSUM may be better or worse than the FSS depending upon the situation.

2.2.3 The Girshick-Rubin-Shiryayev Procedure

In 1952, Girshick and Rubin [39] proposed a procedure which is an interesting competitor of the CUSUM procedure. The performance of this procedure was studied by Roberts [102] using simulations.

Independently, Shiryayev (see [110] and [111], Section 4.3, pp.193-200) used Martingale and Bayesian analysis to approach this problem. It is interesting that Girshick-Rubin's procedure is almost identical to Shiryayev's procedure except that Girshick and Rubin [39] assumed that the probability (π_0) that the change occurs at the beginning is zero. We call the latter slightly special case the Girshick-Rubin-Shiryayev procedure, or simply the GRS procedure. When π_0 is not zero, the procedure is called the Shiryayev procedure.

Following [111], the GRS and Shiryayev procedures are presented. The relationship between the Girshick-Rubin procedure and the Shiryayev procedure is explicitly emphasized.

It is assumed that the change time m is a random variable M. Assuming that the prior distribution of the unknown change time $M = m$ is geometric:

$$P\{M = m\} = \begin{cases} \pi_0 & \text{if } m = 0 \\ (1 - \pi_0)(1 - p)^{m-1}p & \text{if } m \geq 1. \end{cases} \qquad (2.2.10)$$

Then the detection probability is

$$
\begin{aligned}
\pi &= P\{\Theta_N\} \\
&= P\{N \geq M\} \\
&= \sum_{n=1}^{+\infty} P\{M \leq n\}P\{N = n\} \\
&= \sum_{n=1}^{+\infty} \sum_{m=0}^{n} P\{M = m\}P\{N = n\} \\
&= \sum_{n=1}^{+\infty} [\pi_0 + \sum_{m=1}^{n} (1 - \pi_0)(1 - p)^{m-1}p]P\{N = n\} \\
&= 1 - (1 - \pi_0)\sum_{n=1}^{+\infty}(1 - p)^n P\{N = n\} \\
&\geq 1 - (1 - \pi_0)\sum_{n=1}^{+\infty} P\{N = n\} \\
&= \pi_0. \qquad (2.2.11)
\end{aligned}
$$

From the above equation it is clear that π_0 is not equal to the detection probability in general.

21

Following [111], a posterior probability is defined as

$$\pi_n = P\{M \leq n | X_1^n = x_1^n\}. \tag{2.2.12}$$

It is interesting that the above definition of π_n is similar to a posterior detection probability $P\{\Theta_N | X_1^N = x_1^n\}$ according to eq.(2.1.5) in this thesis.

Based on the assumption (2.2.10), it can be shown using Bayes' formula that π_n can be written in a recursive form [see [111] Eq.(4.120)], as is stated in the following lemma:

LEMMA 2.2 *Let f_1 and f_0 denote the densities of i.i.d. observations X_i for $i \geq M$ and $i < M$ respectively, and l_n the likelihood ratio for nth observation, then*

$$
\begin{aligned}
\pi_{n+1} &= \frac{\pi_n f_1(x_{n+1}) + (1 - \pi_n)p f_1(x_{n+1})}{\pi_n f_1(x_{n+1}) + (1 - \pi_n)p f_1(x_{n+1}) + (1 - \pi_n)(1 - p)f_0(x_{n+1})} \\
&= \frac{[\pi_n + (1 - \pi_n)p]l_{n+1}}{[\pi_n + (1 - \pi_n)p]l_{n+1} + (1 - \pi_n)(1 - p)}
\end{aligned} \tag{2.2.13}
$$

for $n \geq 0$.

□

PROOF: For convenience we define densities of a random vector X_1^n conditioned by an event A as

$$p_x(x_1^n | A) \triangleq \frac{\partial^n P(X_1^n \leq x_1^n | A)}{\partial x_1 ... \partial x_n} \tag{2.2.14}$$

and a joint density as

$$p_x(x_1^n, A) \triangleq \frac{\partial^n P(X_1^n \leq x_1^n, A)}{\partial x_1 ... \partial x_n}. \tag{2.2.15}$$

From the definition of π_n we have

$$
\begin{aligned}
\pi_{n+1} &= P(M \leq n + 1 | X_1^{n+1} = x_1^{n+1}) \\
&= \frac{p_x(x_1^{n+1}, M \leq n + 1)}{p_x(x_1^{n+1})} \\
&= \frac{p_x(x_1^{n+1}, M \leq n + 1)}{p_x(x_1^{n+1}, M \leq n + 1) + p_x(x_1^{n+1}, M > n + 1)}
\end{aligned} \tag{2.2.16}
$$

22

The numerator of the above equation is

$$
\begin{aligned}
p_x(x_1^{n+1}, M \le n+1) &= p_x(x_1^{n+1}|M \le n+1)P(M \le n+1) \\
&= p_x(x_1^n|M \le n+1)P(M \le n+1)f_1(x_{n+1}) \\
&= p_x(x_1^n, M \le n+1)f_1(x_{n+1}) \\
&= [p_x(x_1^n, M \le n) + p_x(x_1^n, M = n+1)]f_1(x_{n+1}) \\
&= [P(M \le n|X_1^n = x_1^n)p_x(x_1^n) + P(M = n+1|X_1^n = x_1^n)p_x(x_1^n)]f_1(x_{n+1}) \\
&= [\pi_n + P(M = n+1|X_1^n = x_1^n)]p_x(x_1^n)f_1(x_{n+1}) \quad\quad (2.2.17)
\end{aligned}
$$

Using the geometric-distribution assumption (2.2.10) and the Markovian property of M and X_n yields

$$
(1 - \pi_n)(1 - p)^0 p = (1 - \pi_n)p = P(M = n+1|X_1^n = x_1^n) \quad\quad (2.2.18)
$$

Substituting the above equation into (2.2.17) yields

$$
p_x(x_1^{n+1}, M \le n+1) = [\pi_n + p(1 - \pi_n)]p_x(x_1^n)f_1(x_{n+1}) \qu\quad (2.2.19)
$$

Similarly it can be shown that the denominator of (2.2.16) is

$$
\begin{aligned}
p_x(x_1^{n+1}, M &\le n+1) + p_x(x_1^{n+1}, M > n+1) \\
&= [\pi_n + p(1 - \pi_n)]p_x(x_1^n)f_1(x_{n+1}) + (1 - \pi_n)(1 - p)p_x(x_1^n)f_0(x_{n+1})
\end{aligned}
$$

$$
(2.2.20)
$$

Substituting the above two equations into (2.2.16) yields (2.2.13).

QED

From the above derivation, it is clear that the geometric-distribution assumption (2.2.10) is essential for the recursive algorithm stated in Lemma 2.2.

The Shiryayev procedure is to stop and decide that a change has occurred as soon as

$$
\pi_n \ge A' \quad\quad (2.2.21)
$$

23

for some constant $A' > 0$.

Based on the assumption (2.2.10), Shiryayev [111] has shown that the above procedure is optimal in a certain sense. This result is presented as follows.

Shiryayev's Theorem ([111], p.200, Theorem 8) *Let $0 \leq \pi_0 < 1$ and $p > 0$. Let N_s be the stopping variable for Shiryayev's procedure and N for any alternative procedure. Suppose that the prior distribution of the random unknown change time M is given by (2.2.10) and that the false alarm probability*

$$\alpha_s \triangleq P\{N_s < M\} \tag{2.2.22}$$

for Shiryayev's procedure is a continuous function of the decision threshold A' for any $\alpha_s \in (0,1)$. If for any alternative procedure its false alarm probability satisfies

$$\alpha \triangleq P\{N < M\} \leq \alpha_s. \tag{2.2.23}$$

then

$$E\{N_s - M | N_s \geq M\} \leq E\{N - M | N \geq M\}. \tag{2.2.24}$$

□

Several remarks have been made in [111], notably the continuity problem of α_s, i.e., that for any $\alpha_s \in (0.1)$ there may not necessarily exist a procedure (2.2.14) which is able to achieve this false alarm probability α_s. Actually this continuity problem of error probabilities also exists for the optimality of the SPRT [125]. i.e., that for any given error probabilities, there may not necessarily exist an SPRT which is able to achieve these error probabilities. In this sense, the optimality of the SPRT is less strong than the Neyman-Pearson Lemma since the Neyman-Pearson Lemma states that for any given false alarm probability there always exists a (possibly randomized) likelihood ratio test which can achieve this false alarm probability [63, 97]. A few authors have tried to introduce a randomized (extended) SPRT [103, 125]. In [103], a randomized SPRT

24

for Bernoulli distribution is studied. However, a general and practical solution to the randomized SPRT seems unavailable thus far.

In order for the GRS procedure to be optimal, the unknown change-time M has to be distributed geometrically, i.e., the assumption (2.2.10) must hold. However, a geometric assumption of the unknown change-time appears to be restrictive for a number of applications.

In addition, Shiryayev's optimality criterion does not consider the false alarm average run length (false alarm ARL) at all, which seems a little unreasonable for certain situations.

An interesting observation about Shiryayev's procedure is that it is equivalent to a generalized likelihood ratio procedure. This observation is presented as the following lemma.

LEMMA 2.3 *Shiryayev's procedure defined by* (2.2.21) *is equivalent to a generalized likelihood ratio procedure defined by the following decision rule. The decision statistic is*

$$R_n \triangleq \frac{\pi_n}{p(1 - \pi_n)}$$

$$= \frac{p(1 - \pi_0) + \pi_0}{p(1 - \pi_0)} \prod_{i=1}^{n} \frac{l_i}{1 - p} + \sum_{k=2}^{n} \prod_{i=k}^{n} \frac{l_i}{1 - p} \tag{2.2.25}$$

which can be written recursively as

$$R_{n+1} = (R_n + 1) \frac{l_{n+1}}{1 - p} \qquad n \geq 0 \tag{2.2.26a}$$

with

$$R_0 = \frac{\pi_0}{p(1 - \pi_0)}. \tag{2.2.26b}$$

One stops and decides a change has occurred as soon as

$$R_n \geq A \tag{2.2.27}$$

is first satisfied for some $n \geq 1$, where

$$A = \frac{A'}{p(1 - A')} \tag{2.2.28}$$

25

□

PROOF: From Lemma 2.2 it is seen that

$$
\begin{aligned}
\pi_{n+1} &= \frac{[\pi_n + (1 - \pi_n)p]l_{n+1}}{[\pi_n + (1 - \pi_n)p]l_{n+1} + (1 - \pi_n)(1 - p)} \\
&= \frac{(R_n + 1)l_{n+1}}{(R_n + 1)l_{n+1} + \frac{1-p}{p}}
\end{aligned}
\tag{2.2.29}
$$

Using the above equation and the definition of R_{n+1} yields

$$
\begin{aligned}
R_{n+1} &= \frac{\pi_{n+1}}{(1 - \pi_{n+1})p} \\
&= (R_n + 1)\frac{l_{n+1}}{1 - p}.
\end{aligned}
\tag{2.2.30}
$$

The other equations in the Lemma are straightforward from the above recursive equation and the definition of R_n.

QED

It is noted that when $\pi_0 = 0$,

$$
R_{pn} \triangleq R_n|_{\pi_0=0} = \sum_{k=1}^{n}\prod_{i=k}^{n}\frac{l_i}{1 - p} = (R_{pn-1} + 1)\frac{l_n}{1 - p}
\tag{2.2.31}
$$

with

$$
R_{p0} = 0.
\tag{2.2.32}
$$

As $\pi_0 = 0$ and $p \to 0$

$$
R_{0n} \triangleq R_n|_{\pi_0=0, p\to 0} = \sum_{k=1}^{n}\prod_{i=k}^{n}l_i = (R_{0n-1} + 1)l_n
\tag{2.2.33}
$$

with

$$
R_{00} = 0.
\tag{2.2.34}
$$

Following Lorden's minimax criterion, some properties of the latter special case were studied in [95]. However, it should be emphasized that Lorden's minimax criterion is conservative, as is pointed out in [85] and Subsection 2.1.2 of this thesis.

In [94], it is proposed that the special case R_{0n} might be comparable to other procedures. In this thesis we will conduct this comparison in Section 2.6.

26

It should also be noted that the major properties of the GRS procedure do not require the samples to be identically distributed before and after the change-time. This is an appealing property of the GRS procedure.

It appears difficult to design the decision threshold A analytically. In [111], Shiryayev proposed an upper bound of the false alarm probability for the GRS procedure. However, it turns out to be loose for a number of applications and may not reflect some true properties of the procedure. We now present the Shiryayev upper bound.

Shiryayev's Upper Bound: *If the unknown change-time is a random variable with geometric distribution, the false alarm probability α of the GRS procedure satisfies the following inequality*

$$\alpha \leq 1 - \frac{pA}{1 + pA} = 1 - A' \tag{2.2.35}$$

PROOF: If the unknown change-time is a random variable with geometric distribution, from (2.2.21) it follows that

$$\alpha = 1 - \pi = 1 - E\{\pi_n\} \leq 1 - A' \tag{2.2.36}$$

Combining (2.2.28) and the above equation yields the upper bound (2.2.35).
QED

The upper bound (2.2.35) holds only for geometrically distributed unknown change-time. For non-geometrically distributed change-time or non-random change-time, it is not clear if the bound still holds. In addition, the upper bound is obviously loose for small p and small change-time m. Particularly, for $p = 0$, the upper bound equals one and becomes trivial. For $m = 0$, it is always true that $\alpha = 0$ for any finite A and any stopping rules with stopping probability 1. However, the Shiryayev upper bound is always greater than zero for any finite A and $m = 0$. This fact suggests that the Shiryayev upper bound is loose for small m. Since the α is averaged over all possible $M = m$ which is geometrically distributed, the Shiryayev upper bound does not vary

27

with the real change-time m. In fact, it has been shown, theoretically and experimentally, that for every different realization of the possibly random change-time $M = m$, the false alarm probability (FAP) of the GRS procedure varies significantly with m (see section 3.5 of this thesis for experimental results and Section 3.6 for theoretical justification). Therefore, it is necessary to compute an FAP upper bound of the GRS procedure for a non-random change-time m. This problem will be considered in Section 3.6.

2.3 Generalization of Change Detection Procedures

The procedures presented in the last section are designed for independent and identically distributed (i.i.d.) observations except for the GRS procedure whose properties still hold for independent and non-identically distributed observations. In many applications, the i.i.d. assumption is not realistic. Therefore, generalization of the existing procedures to non-i.i.d. situations has been studied by a number of researchers. In this section we present some of these existing results.

2.3.1 Generalization to Time-Varying Situations

For time-varying and independent observations, the properties of the GRS procedure still hold, including its computational efficiency and the optimality under the condition that the unknown change-time is geometrically distributed. This point can be easily observed from the derivation of the GRS procedure.

However, the CUSUM procedure has a number of difficulties for the time-varying situation. To overcome this difficulty, Blostein [17] has proposed to approximate the CUSUM statistic. The approximated CUSUM (A-CUSUM) statistic is defined as follows:

$$\hat{T}_n = \max\{\hat{T}_{n-1}, 1\} l_{jn} \qquad (2.3.1)$$

28

for $0 \leq j \leq n$ and $n \geq 0$, where $\hat{T}_0 \triangleq 1$,

$$l_{jn} \triangleq \frac{f_{n-j}(x_n)}{f_0(x_n)}. \qquad (2.3.2)$$

and

$$j \triangleq \sup_{1 \leq i \leq n} \{i | \hat{T}_{i-1} \leq 1\}. \qquad (2.3.3)$$

It can be seen that j is an assumed change-time, i.e., the most recent time when $\hat{T}$ crosses over 1. Then the approximated CUSUM (A-CUSUM) procedure is a repeated SPRT with nonstationary observations, which we refer to as NSPRT. The properties of the NSPRT will be studied in Section 4.2 of this thesis. It turns out that the optimum threshold for the NSPRT and A-CUSUM is time-varying for time-varying signals. The effectiveness of the A-CUSUM procedure has been demonstrated in [17].

It should be noted that the A-CUSUM statistic is interesting since the new MAR procedure proposed in this thesis is a function of the CUSUM statistic. This problem will be discussed Section 4.2.

2.3.2 Generalization to Dependent Observations

When the observations are dependent, the CUSUM and GRS statistics cannot be computed recursively, and their properties cannot be easily analyzed. As a result, the existing procedures for i.i.d. observations cannot be directly implemented. A widely used approach is to preprocess the dependent observations. The preprocessor may be a Kalman filter for a state-space model, an autoregressive-moving-average (ARMA) model [88, 81, 7], or a linear predictor[108, 8]. The idea is to whiten the dependent observations. Using this error-signal approach (also known as innovation-based approach), the change-detection problem for dependent observations is transformed to a change-detection problem for independent observations.

To detect a change in the error sequence, intuitive likelihood-ratio-type procedures different from CUSUM and GRS are used in [7, 8, 88]. However, it appears that neither

29

theoretical nor experimental evidence is available to justify why these intuitive procedures should be used instead of the CUSUM, GRS or other optimum procedures.

2.3.3 Generalization to Unknown Distribution Parameters

A problem of practical interest is change-detection with unknown distribution parameters. Since in many applications enough information may be available to estimate the parameters before the change, many researchers have been concerned with the change-detection with unknown distribution parameters after the change and a completely known distribution before the change.

One approach to this problem is to ignore the information about the density f_1 after the change [108]. In [108], it is actually assumed that

$$f_1(x_n|x_1^{n-1}) = \varepsilon \qquad (2.3.4)$$

for all $n \geq 1$ with ε to be a constant. Therefore the likelihood ratio is

$$l_n = \frac{f_1(x_n|x_1^{n-1})}{f_0(x_n|x_1^{n-1})} = \frac{\varepsilon}{f_0(x_n|x_1^{n-1})} \qquad (2.3.5)$$

or equivalently the log-likelihood ratio is

$$z_n = -\ln f_0(x_n|x_1^{n-1}) \qquad (2.3.6)$$

for all $n \geq 1$, where the constant $\ln \varepsilon$ is ignored and f_0 is the density before the change and is assumed completely known. After the above simplification, the change-detection problem becomes a typical change-detection problem with completely known distributions and existing techniques can be used.

It has been noted that much information is lost by assuming $f_1(x_n|x_1^{n-1}) = 1$ [8]. Therefore the density f_1 should be used and the unknown parameter θ after the change is replaced by a maximum-likelihood estimator $\hat{\theta}$ [127, 7]. This approach is sometimes called the *generalized likelihood ratio (GLR)* approach. If an adaptive estimation algorithm is used, the computational complexity of this approach may not be very high

30

and the task of adaptive identification of the changed parameters is also accomplished simultaneously.

A multiple-model approach is also used in [126, 7], where it is assumed that the unknown parameters may take one of a bank of assumed parameters. Then a parallel bank of change-detection procedures are used. Whenever one of these procedures reaches its stopping threshold, it is decided that a change has occurred. This approach may react to a change faster than the parameter-estimation approach. However, the complexity of this approach may be too high in a number of applications, particularly when the observations are dependent and whitening preprocessing is required. In addition, our experiments show that the estimate of the changed parameter is often not close to the real parameter if the number of assumed parameters is not very large. Adaptive estimation of the unknown parameter is impossible for this approach.

We will study the multiple-model approach and parameter-estimation approach in section 4.1.

2.4 Applications of Change Detection

Change detection has been applied to a number of fields. We will first give a list of these applications. Then an important application to adaptive error control over slowly time-varying channels [100] will be presented.

2.4.1 A List of Applications

- Application to Communications:

  In [100], change-detection was applied to adaptive error control over slowly time-varying channels. Using a good change-detection technique, significant improvement over nonadaptive error control techniques has been reported in [100]. A detailed discussion of this application will be presented in the next subsection. In

31

Chapter 6 of this thesis, change detection will be applied to adaptive identification of nonstationary fading channels.

- **Application to Speech Processing:**

  In [4, 35], change detection techniques have been applied to speech segmentation. The speech signal is modeled as an autoregressive (AR) process. Then a change from one segment to another can be viewed as a change of the AR process parameters. Therefore the segmentation problem becomes a problem of detecting changes in AR process parameters.

- **Application to Search Radar:**

  Change detection has been applied to search radar [36, 18]. When a target does not exist, the received signal would be noise only. The observations from the matched filter and sampler may be modeled as a Gaussian distributed sequence with zero-mean. When a target suddenly appears, the mean of the observations will change abruptly. Therefore the problem is to detect a change in Gaussian mean.

- **Application to Adaptive System Identification:**

  This problem has been studied by a number of researchers [115, 116, 93, 10]. The techniques used mainly involve two steps: detection statistics formulation and change-detection-procedure application. In addition, this problem is closely related to change-detection for dependent observations, and in a number of applications they are almost identical [8, 10, 108]. It should be noted that most research on this problem has focused on linear state space models and ARMA models. The changes involved are typically additive disturbances and system-state jumps.

- **Application to Image Processing:**

  This application has been concentrating on edge detection of digital images [6]. In [6], the images are modeled as a number of locally stationary areas separated by

32

edges. The edges are areas where abrupt changes occur. Edge detection involves two steps: first, detection of elements, i.e., isolating points of possible edges; second, edge following, i.e., linking the detected points together. The first step of the edge detection is essentially to detect an abrupt change in the mean of the Gaussian distribution, as is addressed in [6].

Change detection also has been applied to failure detection [128], quality control engineering [42, 102], geophysical signal processing [9], biomedical signal processing [104], chemical engineering [50], among others. A survey on the applications of change detection is found in [7, 9].

2.4.2 An Application to Adaptive Error Control Over Time-Varying Channels

In this subsection, we report a research on application of change-detection to adaptive error control for packet data transmission over time-varying channels. This work was conducted by M. Rice during his recent PhD thesis research at Georgia Tech [100], and has not been published.

Automatic retransmission request (ARQ), hybrid-ARQ and forward error correction (FEC) are popular techniques to control the bit-error-rate of data transmission over communications networks. It is well-known [65] that the ARQ is best suited for a clear channel (a channel with low bit error rate), the FEC offers the highest throughput over a noisy channel, while hybrid-ARQ strategies work best for a channel between the two extremes. Therefore a natural idea to maximize the throughput subject to a given error rate is to use the ARQ for a clear channel and switch to a hybrid-ARQ strategy when channel is not clear, and vice versa.

In ARQ and type-I hybrid-ARQ strategies, decoding results can be described by the following two events: (i) the packet is accepted; and (ii) the packet is rejected. A random

33

variable X is defined as follows:

$$X = \begin{cases} 0 & \text{if the packet is accepted} \\ 1 & \text{if the packet is rejected.} \end{cases} \tag{2.4.1}$$

Let R_0 and R_1 denote the probability of retransmission (i.e. the probability that the packet is rejected) before and after a channel change, respectively. Then the density of X is

$$f_0(x) = (\frac{R_0}{1 - R_0})^x (1 - R_0) \tag{2.4.2}$$

before the change and

$$f_1(x) = (\frac{R_1}{1 - R_1})^x (1 - R_1) \tag{2.4.3}$$

after the change.

The one-sample log-likelihood ratio is

$$z_n = \ln \frac{f_1(x_n)}{f_0(x_n)} = x_n \ln \frac{R_1(1 - R_0)}{R_0(1 - R_1)} \frac{(1 - R_1)}{(1 - R_0)} \tag{2.4.4}$$

and the (log) CUSUM statistic is

$$T'_n = \max\{z_n, 0\} + T'_{n-1}. \tag{2.4.5}$$

The decision rule of the CUSUM procedure is

$$T'_n \geq b \tag{2.4.6}$$

for some constant b.

A two-sided test is used for this problem by assuming $R_1 = R_+ > R_0$ and $R_1 = R_- < R_0$. Two symmetric CUSUM procedures are implemented simultaneously and certain ARQ, hybrid-ARQ strategies are then chosen according to the change-detection results.

Using the above scheme, significant improvement (as much as 16%) of the throughput and error-rate performance over non-adaptive strategies is reported in [100].

Chapter 3

MAR: A New Procedure for Quickest Change Detection

In this chapter, a new procedure, referred to as the minimum-asymptotic-risk (MAR) procedure for quickest detection of an abrupt change in a random sequence is introduced. It is assumed that the observations are i.i.d. before and after the change and that the distributions are completely known. It is shown that the MAR procedure converges almost surely to a theoretically optimum procedure. The superiority of the new procedure over existing well-known procedures for non-asymptotic cases are shown by extensive simulations. The false alarm probability of various change detection procedures is studied in the final section.

3.1 Problem Statement and Criterion

3.1.1 Problem Statement

Suppose $\{ X_i | i = 1, 2, ... \}$ are independent random variables observed sequentially and there exists some positive, non-random integer m such that the distribution function of X_i is $F_0(x_i)$ with density $f_0(x_i)$ for $i = 1, ... m - 1$, and $F_1(x_i)$ for $i = m, m + 1, ...$ with density $f_1(x_i)$, where F_0 and F_1 are known but m is unknown. Our objective is to stop

and raise alarm after m occurs, as quickly as possible, subject to a given false alarm constraint.

Let $N = n$ denote a stopping time. For convenience let $X_i^n \triangleq (X_i, X_{i+1}, ..., X_n)^t$ and $x_i^n \triangleq (x_i, x_{i+1}, ..., x_n)^t$ for $i \leq n$. When $i > n$, X_i^n or x_i^n means no observations have been taken. Let $dx_i^n \triangleq dx_i dx_{i+1}...dx_n$ for $i \leq n$. For any positive, non-random integer m, we define joint distributions P_m as

$$P_m\{X_i^n \leq x_i^n\} = \begin{cases} \int_{-\infty}^{x_i^n} f_0(x_i) f_0(x_{i+1})...f_0(x_n) dx_i^n & \text{if } n < m \\ \int_{-\infty}^{x_i^n} f_0(x_i)...f_0(x_{m-1}) f_1(x_m)...f_1(x_n) dx_i^n & \text{if } i < m \leq n \\ \int_{-\infty}^{x_i^n} f_1(x_i) f_1(x_{i+1})...f_1(x_n) dx_i^n & \text{if } m \leq i \leq n \end{cases} \quad (3.1.1)$$

and expectation E_m under P_m. Technically, the situation where no change ever occurs is P_∞. Often in the literature, the alternative notation $P_0 \triangleq P_\infty$ and E_0 are used. Let H_m denote the hypothesis that P_m is true. Let a_m denote the event that one accepts H_m. Define the likelihood ratio of the ith sample,

$$l_i \triangleq \frac{f_1(x_i)}{f_0(x_i)} \quad (3.1.2)$$

and the product form

$$S_k^n \triangleq \begin{cases} \prod_{i=k}^n l_i & \text{if } n \geq k \\ 1 & \text{if } n < k. \end{cases} \quad (3.1.3)$$

For any stopping time $N = n$, define

$$\Theta_n \triangleq \{X_n \sim f_1\} \qquad \overline{\Theta}_n \triangleq \{X_n \sim f_0\}. \quad (3.1.4)$$

This definition of Θ_n implies a *correct detection*, i.e., that one stops at or after the real change-time m ($n \geq m$). $\overline{\Theta}_n$ implies a *false alarm*, i.e., that one stops before m ($n < m$). Let

$$\pi \triangleq P(\Theta_N) \quad (3.1.5)$$

represent the probability that Θ_N is true for a stopping rule N, where $0 < \pi < 1$. We remark that π is the *correct detection probability*, i.e., the overall probability that the

36

change detection procedure terminates on or after the change-time m, and

$$\alpha \triangleq 1 - \pi \qquad (3.1.6)$$

is the overall *false alarm probability*. It should be noted that π and α depend on the decision rule and the unknown, non-random change time m.

3.1.2   A Modified Shiryayev Criterion

To compare the performance of procedures, Page [89] defined the average run length (ARL), i.e., that

$$N_d \triangleq E\{N|N \geq m\} = E\{N|\Theta_N\}, \qquad (3.1.7)$$

which is the detection ARL, and

$$N_{fm} \triangleq E\{N|N < m\} = E\{N|\overline{\Theta}_N\}, \qquad (3.1.8)$$

which is the false alarm ARL. The expected delay for any given change-time m is then

$$D_m \triangleq N_d - m + 1 = E\{N - m + 1|\Theta_N\}. \qquad (3.1.9)$$

The definition of D_m was first given by Moustakides [85].

Considering the disadvantages of the existing criteria, we introduce a criterion which concerns both the false alarm probability and false alarm average run length. This criterion is closely related to the Shiryayev criterion, and is henceforth referred to as the *modified Shiayayev criterion*.

- **Modified Shiryayev Criterion:** This criterion minimizes the expected delay

$$D_m = E\{N - m + 1|\Theta_N\} \qquad (3.1.10a)$$

or equivalently the average run length (ARL) under the condition of delay (which we refer to as correct detection ARL)

$$E\{N|\Theta_N\} \qquad (3.1.10b)$$

37

for any given distribution, subject to

$$E(N|\overline{\Theta}_N) = E(N|N < m) \geq t, \tag{3.1.10c}$$

and

$$\pi \geq \gamma, \tag{3.1.10d}$$

or equivalently

$$\alpha < 1 - \gamma, \tag{3.1.10e}$$

where t and γ are two constants, and m is *any* unknown *non-random* change-time.

It should be noted that a large false alarm ARL alone cannot guarantee a small false alarm rate if the false alarm probability α is very large, or vice versa. Therefore the overall false alarm rate of a procedure is described by α and $E(N|N < m)$ together.

An alternative expression of the modified Shiryayev criterion is to maximize

$$E(N|\overline{\Theta}_N) \tag{3.1.11a}$$

subject to

$$D_m \leq t \tag{3.1.11b}$$

and

$$\pi \geq \gamma. \tag{3.1.11c}$$

3.2 MAR: a new procedure for change detection

Now we introduce a new procedure, referred to as the minimum-asymptotic-risk (MAR) procedure to approach this signal change detection problem. A brief discussion of the new procedure is presented. Its properties are shown in the next section.

38

The MAR Procedure

Decision Statistic:
$$\hat{r}(x_1^n) = \frac{T_n \pi_0 c_0 - (1 - \pi_0)}{T_n \pi_0 + 1 - \pi_0} n \tag{3.2.1}$$

Decision Rule: stop and decide a change has occured as soon as

$$\hat{r}(x_1^n) > \hat{r}(x_1^{n-1}). \tag{3.2.2}$$

T_n is Page's CUSUM statistic, $T_n = \max\{T_{n-1}, 1\} l_n$ for $n \geq 1$ and $T_0 = 1$.

π_0 and c_0 are design parameters arbitrarily chosen to achieve certain performance. Practically one may choose $\pi_0 = c_0$.

l_n is the likelihood ratio for nth observation, $n = 1, 2, \ldots$

- **Remark 1.** Since it is well-known [89] that the CUSUM statistic T_n can be written recursively as above, the complexity for calculating $\hat{r}(x_1^n)$ is low and comparable to Page's CUSUM procedure.

- **Remark 2.** As shown in the next section, it is required that $c_0 < (1 - \pi_0)/\pi_0$. Experimentally, we have observed that the performance is insensitive to different choices of design parameters π_0 and c_0. For the simulations, we arbitrarily chose

$$c_0 = \pi_0.$$

This issue is discussed further in Section 3.5.

- **Remark 3.** It has been assumed that $X_1, X_2, \ldots$ are i.i.d. before and after the change-time and that the distributions f_0 and f_1 are completely known. In Chapter 4 the MAR procedure will be modified for more practical cases where the knowledge of the distributions may not be completely known or the i.i.d. assumption may not hold.

39

- **Remark 4.** For all procedures, the design parameters must be chosen differently for different change-times in order to have a constant false alarm probability (CFAP). In Section 3.6, some bounds on the false alarm probability of the change-detection procedures are presented. It will be seen from these bounds that the thresholds of procedures have to be varying with different change-times to ensure CFAPs. The problem of ensuring CFAP over a range of change-times is non-trivial and is not considered in this thesis.

In the next section, it will be shown that under certain asymptotic conditions the new procedure converges almost surely (a.s.) to a theoretically optimum procedure which minimizes a risk function. Therefore, we henceforth refer to the new procedure as the minimum asymptotic risk (MAR) procedure. The decision statistic $\hat{r}(x_1^n)$ may be referred to as the MAR statistic. For convenience we also use $\hat{r}(n)$ to represent $\hat{r}(x_1^n)$. The performance of the MAR procedure will be investigated in the following sections. The almost sure convergence of the MAR procedure to an optimum procedure under asymptotic conditions is to be shown in the next section. For the non-asymptotic situation, the superiority of the MAR procedure over existing procedures will be shown in Section 3.5 via extensive simulations.

The MAR procedure was first presented in [72, 73] by Liu and Blostein.

3.3 Almost Sure Convergence of the MAR Procedure to an Optimum Procedure under Asymptotic Conditions

In this section, it will be shown that the MAR procedure converges almost surely to an optimum procedure if (i) the signal-to-noise ratio approaches zero or (ii) the false alarm probability goes to zero. The major result is Propositions 3.1 and 3.2.

40

We first formulate a noncausal theoretically optimum procedure with respect to the modified Shiryayev criterion. This optimum decision rule can be never realized due to a number of reasons. However, we will show that the MAR procedure almost surely converges to the optimum procedure under certain asymptotic conditions.

LEMMA 3.1 *Let $N(> 0)$ denote the stopping variable for any stopping rule. There exists a noncausal theoretically optimum stopping rule $N_b(> 0)$ such that for any given distribution with $0 < l_i < +\infty$ for all $i \geq 1$, we have*

$$D_{bm} \leq D_m \qquad\qquad (3.3.1)$$

subject to

$$N_{bfm} \leq N_{fm} \qquad\qquad (3.3.2)$$

and

$$\alpha_b = \alpha, \qquad\qquad (3.3.3)$$

where D_{bm} and D_m denote the expected delay for the optimum procedure and any other procedure, respectively, N_{bfm} and N_{fm} denote the corresponding false alarm ARLs, and α_b and α denote the false alarm probabilities. $\square$

PROOF: The stopping rule N_b is derived from the following Bayesian analysis:

Let $N = n$ be a random stopping time. Define a risk function

$$R(x_1^n) \triangleq \begin{cases} cn & \text{if } \Theta_n \text{ is true} \\ -n & \text{if } \overline{\Theta}_n \text{ is true.} \end{cases} \qquad\qquad (3.3.4)$$

The posterior risk function after observing $X_1^n = x_1^n$ and stopping at $N = n$ is

$$
\begin{aligned}
r(x_1^n) &\triangleq E\{R(X_1^N)|X_1^N = x_1^n\} \\
&= E\{R(X_1^N)|X_1^N = x_1^n, \Theta_N\}P(\Theta_N|X_1^N = x_1^n) + \\
&\quad E\{R(X_1^N)|X_1^N = x_1^n, \overline{\Theta}_N\}P(\overline{\Theta}_N|X_1^N = x_1^n) \\
&= cn\pi_n - n(1 - \pi_n), \qquad\qquad (3.3.5)
\end{aligned}
$$

41

where we define

$$\{X_1^N = x_1^n\} \triangleq \{N = n, X_1^n = x_1^n\}, \tag{3.3.6}$$

and

$$\pi_n \triangleq P(\Theta_N | X_1^N = x_1^n). \tag{3.3.7}$$

The latter is the probability that Θ_N is true after observing $X_1^n = x_1^n$ and stopping at $N = n$. In Eq.(3.3.5), $cn\pi_n$ denotes the posterior cost of delay when Θ_N is true, while $-n(1 - \pi_n)$ represents the posterior false alarm cost when $\overline{\Theta}_N$ is true. The unit false alarm cost is normalized to unity, so $c > 0$ denotes the relative cost of delay with respect to false alarm. The above cost structure is different from that in [111]. It should be noted that the above function is used only for analytical purposes.

Applying Bayes' Theorem, it is easily shown that

$$\pi_n = \frac{S_m^n \pi}{S_m^n \pi + 1 - \pi}. \tag{3.3.8}$$

Substituting the above equation into (3.3.5) yields

$$r(x_1^n) = \frac{S_m^n \pi c - (1 - \pi)}{S_m^n \pi + 1 - \pi} n. \tag{3.3.9}$$

For convenience we also use $r(n)$ to represent $r(x_1^n)$.

Define the decision rule N_b to be

$$r(n_b) \triangleq \min_{1 \leq n < +\infty} r(n) \tag{3.3.10}$$

where

$$P_m(N_b = n) \triangleq \begin{cases} 1 & \text{if } n = n_b \\ 0 & \text{if } n \neq n_b. \end{cases} \tag{3.3.11}$$

Since N_b is noncausal and depends on unknown change time m, it can never be realized. N_b is used for analytical purposes only.

The expected risk for N_b is

$$E\{R(X_1^{N_b})\} = E\{\sum_{n=1}^{+\infty} P(N_b = n) r(X_1^n)\}. \tag{3.3.12}$$

42

On the other hand, from (3.3.4)(3.3.5) we have

$$
\begin{aligned}
E\{R(X_1^{N_b})\} &= E\{R(X_1^{N_b})|\Theta_{N_b}\}P(\Theta_{N_b}) + E\{R(X_1^{N_b})|\overline{\Theta}_{N_b}\}P(\overline{\Theta}_{N_b}) \\
&= cE(N_b|\Theta_{N_b})\pi - E(N_b|\overline{\Theta}_{N_b})(1 - \pi).
\end{aligned}
\tag{3.3.13}
$$

For any alternative procedure the expected risk is

$$
E\{R(X_1^N)\} = E\{\sum_{n=1}^{+\infty} P(N = n)r(X_1^n)\} = cE(N|\Theta_N)\pi - E(N|\overline{\Theta}_N)(1 - \pi).
\tag{3.3.14}
$$

According to (3.3.10)–(3.3.12), the expected Bayesian risk for N_b is

$$
\begin{aligned}
E\{R(X_1^{N_b})\} &= E\{\sum_{n=1}^{+\infty} r(X_1^n)P(N_b = n)\} \\
&= E\{r(X_1^{n_b})\} \\
&= E\{r(X_1^{n_b})\sum_{n=1}^{+\infty} P(N = n)\} \\
&= E\{\sum_{n=1}^{+\infty} r(X_1^{n_b})P(N = n)\} \\
&\leq E\{\sum_{n=1}^{+\infty} r(X_1^n)P(N = n)\} \\
&= E\{R(X_1^N)\}
\end{aligned}
\tag{3.3.15}
$$

where we should notice that

$$
\sum_{n=1}^{+\infty} P(N = n) = 1.
$$

On the other hand, according to (3.3.13)–(3.3.15)

$$
cE(N_b|\Theta_{N_b})\pi - E(N_b|\overline{\Theta}_{N_b})(1 - \pi) \leq cE(N|\Theta_N)\pi - E(N|\overline{\Theta}_N)(1 - \pi).
\tag{3.3.16}
$$

From the constraints (3.3.2) and (3.3.3), the above equation yields (3.3.1).

We note that the above argument is similar to the proof of the Wald-Wolfowitz Theorem (see [33, 70, 120]).

QED

The optimum decision rule (3.3.9)(3.3.10) cannot be realized due to the following two reasons: (i) the decision statistic $r(n)$ depends on unknown change-time m, (ii) the

43

decision rule (3.3.10) is noncausal. To solve the first problem, we approximate $r(n)$ using a statistic independent of m. The following lemma may provide some hint towards this direction.

LEMMA 3.2 *For any given m and any n $(1 \leq n < +\infty)$,*

$$E_m(S_m^n) = \max_{1 \leq i \leq n} E_m(S_i^n). \tag{3.3.17}$$

□

PROOF: From definition (3.1.3), we have two cases.

First for $n < m$, we have $S_m^n = 1$, so

$$E(S_m^n) = 1.$$

For all i with $1 \leq i \leq n < m$,

$$
\begin{aligned}
E(S_i^n) &= \int S_i^n f_0(x_i) f_0(x_{i+1})...f_0(x_n) dx_i^n \\
&= \int f_1(x_i) f_1(x_{i+1})...f_1(x_n) dx_i^n \\
&= 1
\end{aligned}
$$

which implies that

$$\max_{1 \leq i \leq n} E(S_i^n) = E(S_m^n).$$

On the other hand, for $n \geq m$, $S_m^n = \prod_{i=m}^{n} l_i$. Notice that $E_0(l_i) = 1$ for $1 \leq i < m$. Also, using the fact that $E_m(\ln l_i) \geq 0$ (see, e.g. Siegmund [113], p.12) and Jensen's inequality, $E_m(l_i) \geq 1$ for $i \geq m$. For any $1 \leq i < m$,

$$0 < E(S_i^n) = \{\prod_{j=i}^{m-1} E_0(l_j)\}\{\prod_{j=m}^{n} E_m(l_j)\} = \prod_{j=m}^{n} E_m(l_j) = E(S_m^n).$$

For any $m \leq i \leq n$,

$$0 < E(S_i^n) = \frac{\prod_{j=m}^{n} E_m(l_j)}{\prod_{j=m}^{i-1} E_m(l_j)} \leq \prod_{j=m}^{n} E_m(l_j) = E(S_m^n).$$

44

In any case,

$$\max_{1 \le i \le n} E(S_i^n) = \prod_{j=m}^{n} E_m(l_j) = E(S_m^n).$$

QED

Lemma 3.2 suggests that Page's CUSUM statistic may be used as an estimator of S_m^n in $r(x_1^n)$. However, this lemma provides only an intuitive understanding of the relationship between T_n and S_m^n. In Proposition 3.1 we will establish a stronger result with regard to the almost sure convergence between the decision statistic $\hat{r}(x_1^n)$ and the optimum decision statistic $r(x_1^n)$ (not the convergence between T_n and S_m^n).

Substituting T_n for S_m^n into (3.3.9) results in the decision statistic, $\hat{r}(n) \triangleq \hat{r}(x_1^n)$, defined by Eq. (3.2.1). The following proposition establishes almost sure convergence between the MAR statistic $\hat{r}(n)$ and the theoretically optimum statistic $r(n)$:

PROPOSITION 3.1 *If*

$$C3.1: \qquad\qquad l_i \to 1 \qquad\qquad\qquad (3.3.18a)$$

for all i, or the false alarm probability

$$C3.2: \qquad\qquad \alpha \to 0 \qquad\qquad\qquad (3.3.18b)$$

and

$$0 < c_0 < \frac{1 - \pi_0}{\pi_0}, \qquad\qquad (3.3.19)$$

then the following convergences hold almost surely:

$$r(n) \to \bar{r}(n) \triangleq \frac{e^{(n-m+1)I_f} \pi c - (1 - \pi)}{e^{(n-m+1)I_f} \pi + 1 - \pi} n \qquad (3.3.20)$$

$$\hat{r}(n) \to \bar{r}(n) \qquad\qquad (3.3.21)$$

and

$$\hat{r}(n) \to r(n) \qquad\qquad (3.3.22)$$

where $I_f \triangleq E_m(\ln l_m)$, $0 < I_f < \infty$, $\pi = \pi_0$ and $c = c_0$. It is assumed that $N_b = n > 0$ and $\hat{N} = n > 0$. □

45

PROOF: When $l_i \rightarrow 1$ for any i, choosing $\pi_0 = \pi_b = \pi$ and $0 < c_0 = c < \frac{1-\pi_0}{\pi_0}$ yields

$$r(x_1^n) = \frac{S_m^n \pi c - (1 - \pi)}{S_m^n \pi + 1 - \pi} n \rightarrow [\pi c - (1 - \pi)]n < 0.$$

Therefore for the stopping rule $N_b = n_b$

$$r(n_b) = \min_{1 \leq n < +\infty} r(x_1^n) \rightarrow -\infty$$

for $n_b \rightarrow +\infty$. It should be noted that if we chose $c \geq \frac{1-\pi_0}{\pi_0}$, then

$$r(x_1^n) \rightarrow [\pi c - (1 - \pi)]n \geq 0$$

and

$$r(n_b) = \min_{1 \leq n < +\infty} r(x_1^n) \rightarrow 0$$

only if $n_b = 0$, which contradicts the condition that $N_b = n_b > 0$.

Similarly for the stopping rule $\hat{N} = \hat{n}$ we have

$$\hat{r}(x_1^n) = \frac{T_n \pi c - (1 - \pi)}{T_n \pi + 1 - \pi} n \rightarrow [\pi c - (1 - \pi)]n < 0,$$

and

$$\hat{r}(\hat{n}) = \min_{1 \leq n < +\infty} \hat{r}(x_1^n) \rightarrow -\infty$$

for $\hat{n} \rightarrow +\infty$. When $\alpha \rightarrow 0$, it can be similarly shown that $c = c_0 \rightarrow 0$ such that $n_b \rightarrow +\infty$ and $\hat{n} \rightarrow +\infty$.

We now show that (3.3.20) and (3.3.21) hold almost surely for $n \rightarrow +\infty$. Eq.(3.3.22) follows from (3.3.20) and (3.3.21) easily. It must be noted that T_n does not converge to S_m^n in general since both of them approach $+\infty$ exponentially at unknown speed. Fortunately, it can be shown that both $r(n)$ and $\hat{r}(n)$ approach $\bar{r}(n)$ almost surely. We will first show (3.3.20) holds almost surely. After (3.3.20) is established, (3.3.21)(3.3.22) are easily shown.

Proof of $r(n) \rightarrow \bar{r}(n)$: Defining

$$z_i \triangleq \ln \frac{f_1(x_i)}{f_0(x_i)} \tag{3.3.23}$$

46

for all $i \geq 1$. then

$$
\begin{aligned}
r(n) &= \frac{e^{z_m + z_{m+1} + \cdots + z_n} \pi c - (1 - \pi)}{e^{z_m + z_{m+1} + \cdots + z_n} \pi + 1 - \pi} n \\
&= \frac{\pi c - (1 - \pi) e^{-(z_m + z_{m+1} + \cdots + z_n)}}{\pi + (1 - \pi) e^{-(z_m + z_{m+1} + \cdots + z_n)}} n
\end{aligned}
\tag{3.3.24}
$$

Simple algebraic manipulation yields

$$
\bar{r}(n) = \frac{\pi c - (1 - \pi) e^{-(n-m+1)I_f}}{\pi + (1 - \pi) e^{-(n-m+1)I_f}} n.
\tag{3.3.25}
$$

Then we have

$$
0 \leq \quad |r(n) - \bar{r}(n)| = n \left| e^{-(n-m+1)I_f} - e^{-(z_m + z_{m+1} + \cdots + z_n)} \right| \cdot
$$
$$
\frac{\pi(1 - \pi)(1 + c)}{[\pi + (1 - \pi) e^{-(z_m + z_{m+1} + \cdots + z_n)}][\pi + (1 - \pi) e^{-(n-m+1)I_f}]}.
\tag{3.3.26}
$$

Since $(1 - \pi) e^{-(z_m + z_{m+1} + \cdots + z_n)} \geq 0$ and $(1 - \pi) e^{-(n-m+1)I_f} \geq 0$, then the second term of the above equation is bounded according to

$$
\frac{\pi(1 - \pi)(1 + c)}{[\pi + (1 - \pi) e^{-(z_m + z_{m+1} + \cdots + z_n)}][\pi + (1 - \pi) e^{-(n-m+1)I_f}]} \leq \frac{(1 - \pi)(1 + c)}{\pi}.
\tag{3.3.27}
$$

Therefore in order to show that $|r(n) - \bar{r}(n)|$ appraoaches zero almost surely, we only need to show that

$$
n \left| e^{-(n-m+1)I_f} - e^{-(z_m + z_{m+1} + \cdots + z_n)} \right| \to 0
\tag{3.3.28}
$$

almost surely. Notice that

$$
n \left| e^{-(n-m+1)I_f} - e^{-(z_m + z_{m+1} + \cdots + z_n)} \right| \leq \max\{ n e^{-(n-m+1)I_f}, n e^{-(z_m + z_{m+1} + \cdots + z_n)} \}.
\tag{3.3.29}
$$

It is easily seen that

$$
n e^{-(n-m+1)I_f} \to 0
\tag{3.3.30}
$$

almost surely as $n \to +\infty$ since $I_f > 0$. Now we show that

$$
n e^{-(z_m + z_{m+1} + \cdots + z_n)} \to 0
\tag{3.3.31}
$$

47

almost surely as $n \to +\infty$. Using the Strong Law of Large Numbers yields that the following convergence holds almost surely:

$$\frac{z_m + z_{m+1} + \cdots + z_n}{n - m + 1} \to I_f = E_m(\ln l_m) \tag{3.3.32}$$

as $n \to +\infty$, where $0 < I_f < +\infty$. According to the definition of the limit and the fact that $I_f > 0$, there always exists a large integer $n_m > 0$ such that the inequality

$$\frac{z_m + z_{m+1} + \cdots + z_n}{n - m + 1} \geq \frac{I_f}{2} \tag{3.3.33}$$

holds almost surely for all $n \geq n_m$. That is

$$z_m + z_{m+1} + \cdots + z_n \geq (n - m + 1)\frac{I_f}{2} \tag{3.3.34}$$

holds almost surely. Then

$$0 < ne^{-(z_m + z_{m+1} + \cdots + z_n)} \leq ne^{-(n-m+1)I_f/2} \tag{3.3.35}$$

holds almost surely. It is clear that

$$ne^{-(n-m+1)I_f/2} \to 0 \tag{3.3.36}$$

almost surely as $n \to +\infty$ since $I_f > 0$.

The above argument establishes the almost sure convergence $r(n) \to \bar{r}(n)$.

Proof of $\hat{r}(n) \to \bar{r}(n)$: It is easily seen that

$$0 \leq |\hat{r}(n) - \bar{r}(n)| = n|e^{-(n-m+1)I_f} - e^{-\ln T_n}| \cdot$$
$$\frac{\pi(1 - \pi)(1 + c)}{[\pi + (1 - \pi)e^{-\ln T_n}][\pi + (1 - \pi)e^{-(n-m+1)I_f}]}. \tag{3.3.37}$$

The second term of the above equation is bounded by

$$\frac{\pi(1 - \pi)(1 + c)}{[\pi + (1 - \pi)e^{-\ln T_n}][\pi + (1 - \pi)e^{-(n-m+1)I_f}]} \leq \frac{(1 - \pi)(1 + c)}{\pi} \tag{3.3.38}$$

Since

$$T_n = \max_{1 \leq i \leq n} S_i^n \geq S_m^n = e^{z_m + z_{m+1} + \cdots + z_n} \tag{3.3.39}$$

48

then

$$0 \leq n|e^{-(n-m+1)I_f} - e^{-\ln T_n}| \leq \max\{ne^{-(n-m+1)I_f}, ne^{-\ln T_n}\}$$

$$\leq \max\{ne^{-(n-m+1)I_f}, ne^{-(z_m+z_{m+1}+\cdots+z_n)}\}. \tag{3.3.40}$$

Then using the exactly same argument as that used before yields that $\hat{r}(n) \to \bar{r}(n)$ almost surely.

The proof that $r(n) \to \hat{r}(n)$ almost surely is straightforward if we notice that

$$0 < |r(n) - \hat{r}(n)| \leq |r(n) - \bar{r}(n)| + |\hat{r}(n) - \bar{r}(n)|. \tag{3.3.41}$$

QED

The asymptotic condition C3.1 $l_i \to 1$ for all i implies that $f_1(x) \to f_0(x)$, i.e., that the signal change becomes very small, and corresponds to the case of low SNR. In the alternative asymptotic condition C3.2, the new procedure is optimal as $\alpha = 1 - \pi$, the probability of stopping too soon, approaches zero. Since m is fixed, this latter statement can be intuitively interpreted as the high SNR case.

We now try to solve the second problem of the theoretically optimum decision rule (3.3.10), i.e., noncausality. This problem is addressed by showing that $\bar{r}(n)$ is convex under certain conditions.

PROPOSITION 3.2 *If $I_f < 1/m$, then $\bar{r}(n)$ is a convex function of n for $n \leq 2/I_f$. If $I_f < 1/m$ and $\alpha < \sqrt{c/(1+c)}$, then there exists a unique minimum point of $\bar{r}(n)$.*

PROOF: Writing $\bar{r}(n)$ as

$$\bar{r}(n) = cn - (1 - \pi)(1 + c)\frac{n}{e^{(n-m+1)I_f}\pi + 1 - \pi}. \tag{3.3.42}$$

The first derivative of $\bar{r}(n)$ with respect to n is

$$\bar{r}'(n) = c - (1 - \pi)(1 + c)\frac{e^{(n-m+1)I_f}\pi(1 - nI_f) + 1 - \pi}{(e^{(n-m+1)I_f}\pi + 1 - \pi)^2} \tag{3.3.43}$$

49

and the second derivative is

$$\bar{r}''(n) = (1-\pi)(1+c)\frac{e^{(n-m+1)I_f}\pi I_f}{(e^{(n-m+1)I_f}\pi + 1 - \pi)^3}[e^{(n-m+1)I_f}\pi(2-nI_f) + (2+nI_f)(1-\pi)].$$
(3.3.44)

Therefore the condition $n \le 2/I_f$ guarantees $\bar{r}''(n) > 0$, i.e., that $\bar{r}(n)$ is convex.

Substituting $n = 1/I_f$ into (3.3.43) yields that

$$\bar{r}'(n)|_{n=1/I_f} = c - (1-\pi)^2(1+c)\frac{1}{(e^{(-m+1+1/I_f)I_f\pi}+1-\pi)^2} > c - (1-\pi)^2(1+c) > 0$$
(3.3.45)

when $\alpha = 1 - \pi < \sqrt{c/(1+c)}$, i.e., that $\bar{r}(n)$ is an increasing function of n at $n = 1/I_f$. From (3.3.43) it is seen that for $n > 1/I_f$

$$
\begin{aligned}
\bar{r}'(n) \;&>\; c - (1-\pi)(1+c)\frac{e^{(n-m+1)I_f}\pi(1 - I_f/I_f) + 1 - \pi}{(e^{(n-m+1)I_f}\pi + 1 - \pi)^2} \\
&=\; c - (1-\pi)^2(1+c)\frac{1}{(e^{(n-m+1)I_f}\pi + 1 - \pi)^2} \\
&>\; c - (1-\pi)^2(1+c) \\
&>\; 0
\end{aligned}
$$
(3.3.46)

i.e., that $\bar{r}(n)$ is an increasing function for all $n \ge 1/I_f$. Combining the above fact with the convexity of $\bar{r}(n)$ yields that $\bar{r}(n)$ must have a unique minimum point.

QED

To provide some intuitive understanding of the MAR procedure, we plot several runs of the MAR and CUSUM statistics in Figure 3.1, where $f_0(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x_i^2}{2\sigma^2}}$ and $f_1(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_i-\mu_1)^2}{2\sigma^2}}$ with $\sigma = 0.25$ and $\mu_1 = 0.025$ (the signal-to-noise ratio is then $SNR \triangleq 20\log\frac{\mu_1}{\sigma}(db) = -20db$). The change-time is fixed at $m = 100$. To have a fixed false alarm probability $\alpha = 0.1$, our experiments (the same experiment is repeated 50000 times) show that the decision threshold for the CUSUM must be $B = 4.61$, and the design parameters for the MAR are chosen to be $\pi_0 = c_0 = 0.0162966$. Ten randomly chosen experiments are plotted in Figure 3.1. The numbers $1, 2, ..., 10$ at the end of each curve represent the position of the stopping-times for the MAR procedure. From Figure

50

3.1, it is clear that for $B = 4.61$, the CUSUM has much larger delays in the experiments No.1, No.2, No.4, No.5, No.6, No.7, No.8, No.9. In No.10 the CUSUM has a false alarm (the new procedure does not have any false alarm in the ten randomly chosen experiments). Only in No.3 is the CUSUM better than the new procedure.

It should be noted that the overall false alarm rate may sometimes be described by the weighted false alarm ARL αN_{fm} and the overall delay performance can be described by the weighted correct detection ARL πD_m, where α is the overall false alarm probability. It can be easily shown that the optimality in Lemma 3.1 can be modified to be

$$\pi_b D_{bm} \leq \pi D_m \tag{3.3.47}$$

subject to

$$\alpha_b N_{bfm} \leq \alpha N_{fm}. \tag{3.3.48}$$

The optimality defined above is necessary for that in Lemma 3.1. When $\alpha_b = \alpha$, the above optimality is the same as that in Lemma 3.1. Practically, in comparing the performance of two procedures, sometimes it is impossible to satisfy both $\alpha_1 = \alpha_2$ and $N_{bfm} \leq N_{fm}$. For these cases, weighted ARLs may be reasonable measures of performance.

3.4 Sensitivity with Respect to Design Parameters

Ideally, a procedure should not be sensitive to different choices of design parameters. For CUSUM, this is not an issue since the only design parameter is a decision threshold, B, which has a one-to-one correspondence with the false alarm probability, α, for a given distribution. On the other hand, both FSS and GRS procedures require more than one design parameter. It has been shown [92] that the performance of the FSS procedure, is very sensitive to different pairs of window sizes, w, and thresholds, d. In the case of the GRS procedure, it is proposed in [102, 94] that the parameter p be set to $p = 0$, corresponding to an approximately uniform prior change-time distribution in $(0, +\infty)$.

51

For the MAR procedure described by Eqs. (3.2.1) and (3.2.2), two design parameters π_0 and c_0 are required, and an analytical test design procedure is an open problem. However, from extensive simulation, we have observed that different pairs of π_0 and c_0 do not alter the performance of the new procedure significantly. In fact, simply choosing

$$\pi_0 = c_0 \tag{3.4.1}$$

does not degrade performance significantly from an optimized design. A precise analytical justification for this observation has not been found. Intuitively, this may be expected since $\hat{r}(n)$ is an approximation to overall risk $r(n)$. Therefore, π_0 and c_0 cannot be interpreted, respectively, as detection probability and unit cost of delay. Only in the asymptotic situation $n \to +\infty$, $\hat{r}(n)$ approaches $r(n)$, and c_0 may be interpreted as the unit cost of each delayed sample. Thus, for a finite change time, m, if $n \to +\infty$, the value of c_0 becomes less critical to performance.

Typical simulation results, shown in Table 3.1 and Figures 3.2–3.2c support the arbitrary choice of c_0 given in Eq. (3.4.1). This insensitivity is easily seen in a wide neighborhood around $\pi_0 = c_0$. In Figure 3.2, when c_0 is large (greater than .01), a slightly higher false alarm probability, α, is traded off by a smaller average delay, D_m, since π_0 is held fixed. Similar behavior was also observed for a wide range of SNR's. In practice, therefore, the new procedure seems to require only one design parameter π_0, and as a result, effectively shares the desirable insensitivity of the CUSUM procedure.
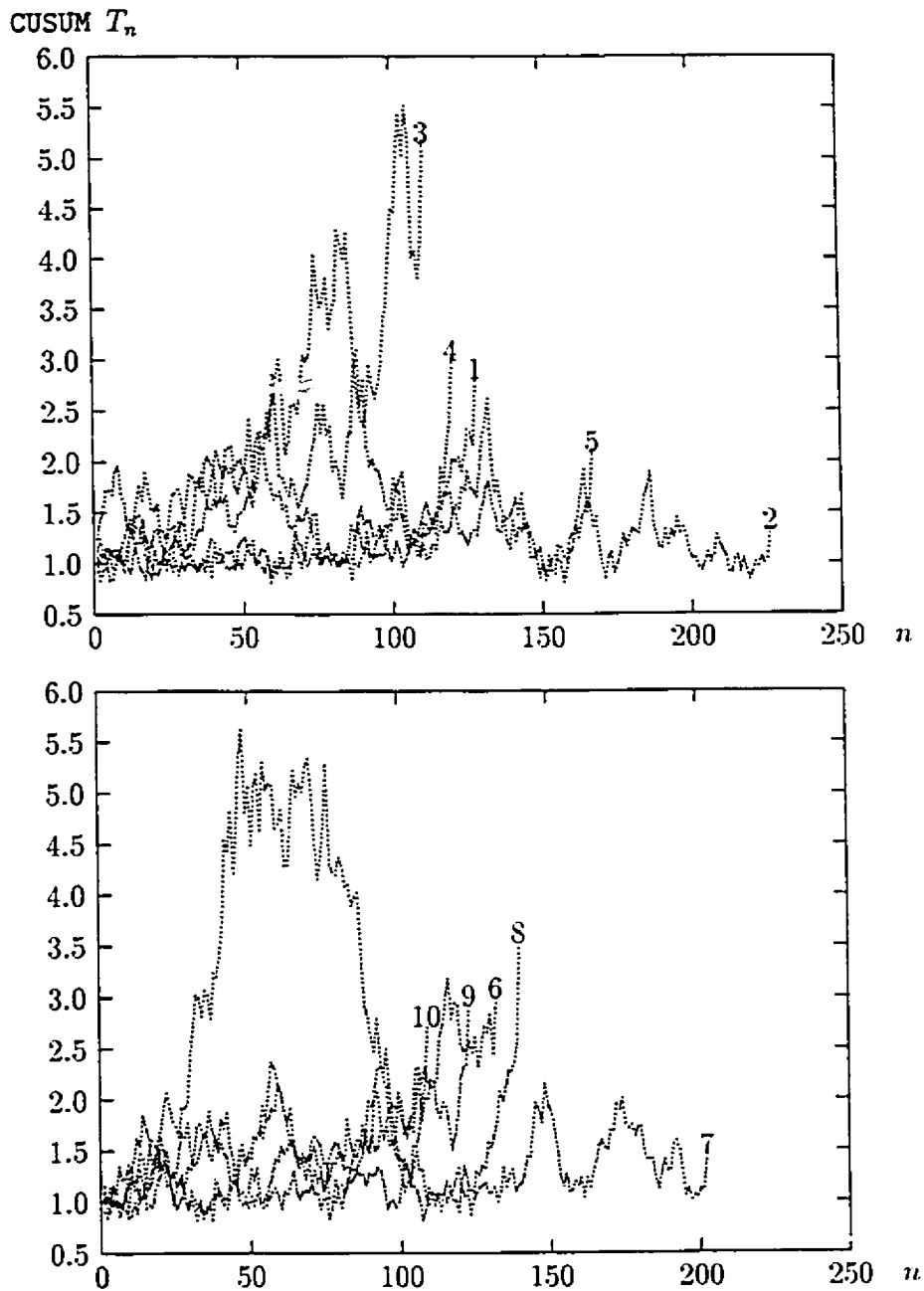
Figure 3.1: *Plot of the CUSUM statistics for 10 experiments which are chosen randomly. SNR = −20dB. The change-time m = 100. The numbers 1,2,...,10 at the end of each curve represent the position of the stopping times of the New Procedure. To have an overall false alarm probability $\alpha = 0.1$, a constant decision threshold must be chosen as $B = 4.61$ for the CUSUM, and $\pi_0 = c_0 = 0.0162966$ for the new procedure.*

53
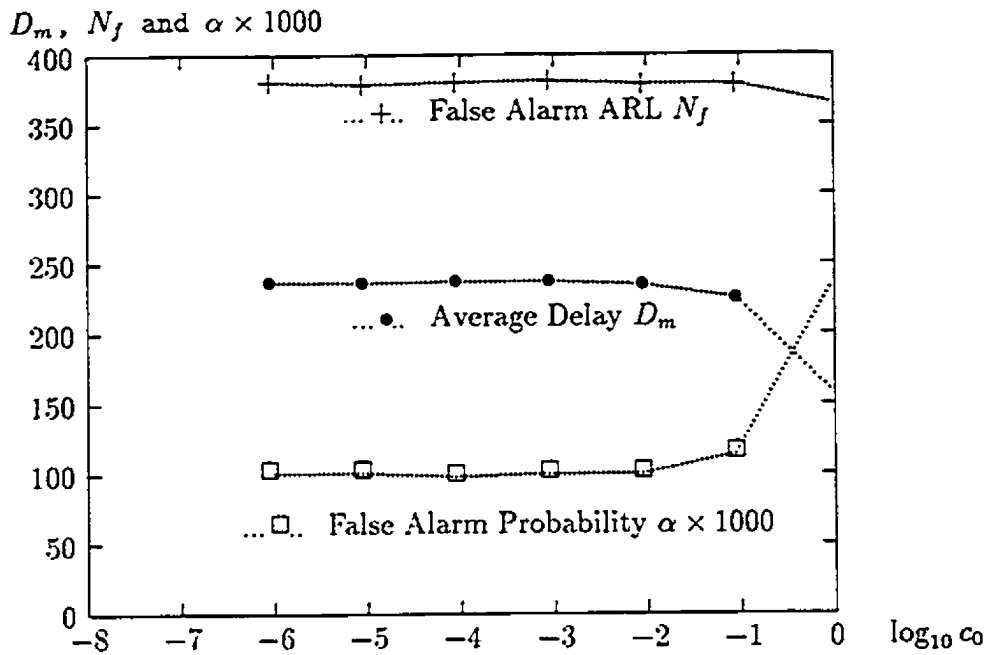
Figure 3.2: *Performance of the new procedure for a wide range of c_0 values. Here, $m =$ 500, $SNR = -20dB$, $\pi_0 = 8.7 \times 10^{-4}$ (i.e. $\log_{10} \pi_0 = -3.06$). As shown, performance is insensitive to different c_0.*
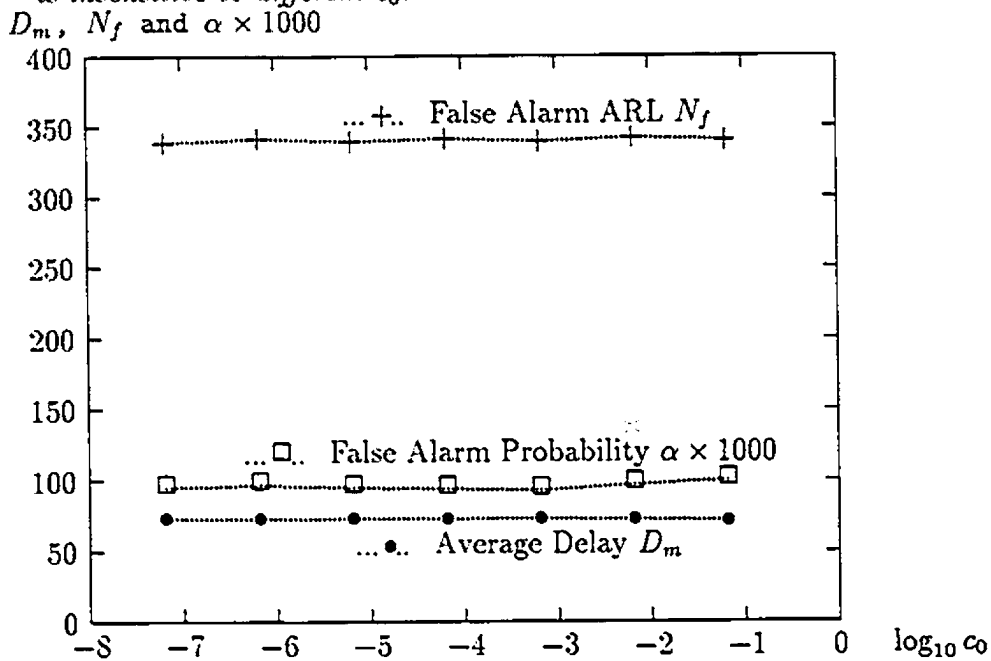
Figure 3.2a: *Performance of the new procedure for a wide range of c_0 values. Here, $m = 500$, $SNR = -10.5dB$, $\pi_0 = 6.52 \times 10^{-5}$ (i.e. $\log_{10} \pi_0 = -4.186$). As shown, performance is insensitive to different c_0.*

54

D_m, N_f and $\alpha \times 1000$
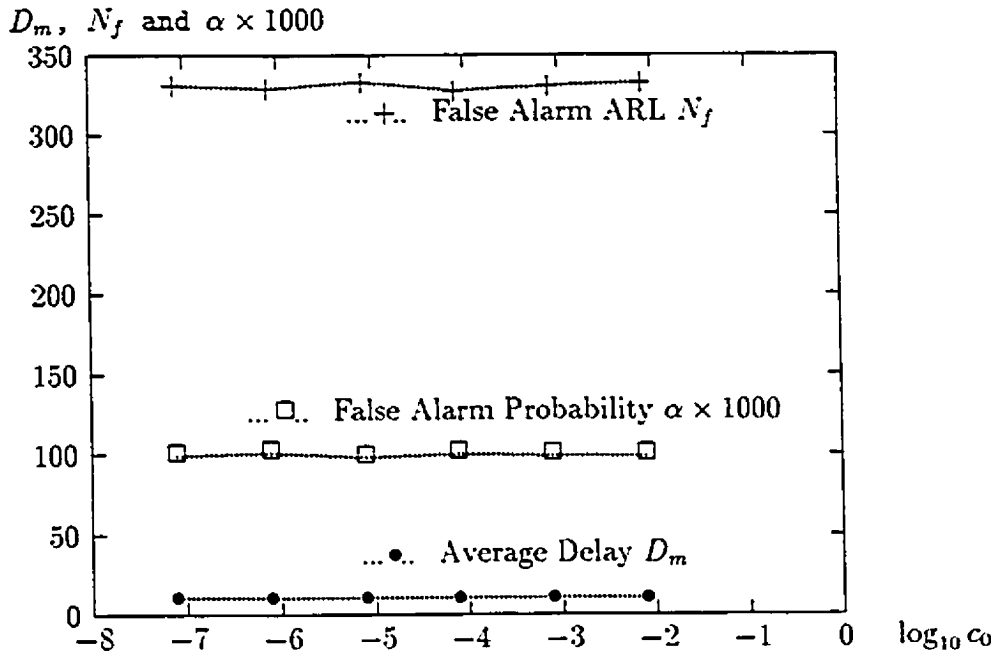
Figure 3.2b: *Performance of the new procedure for a wide range of c_0 values. Here, $m = 500$, $SNR = 0.0dB$, $\pi_0 = 7.84 \times 10^{-6}$ (i.e. $\log_{10} \pi_0 = -5.106$). As shown, performance is insensitive to different c_0.*

D_m, N_f and $\alpha \times 1000$
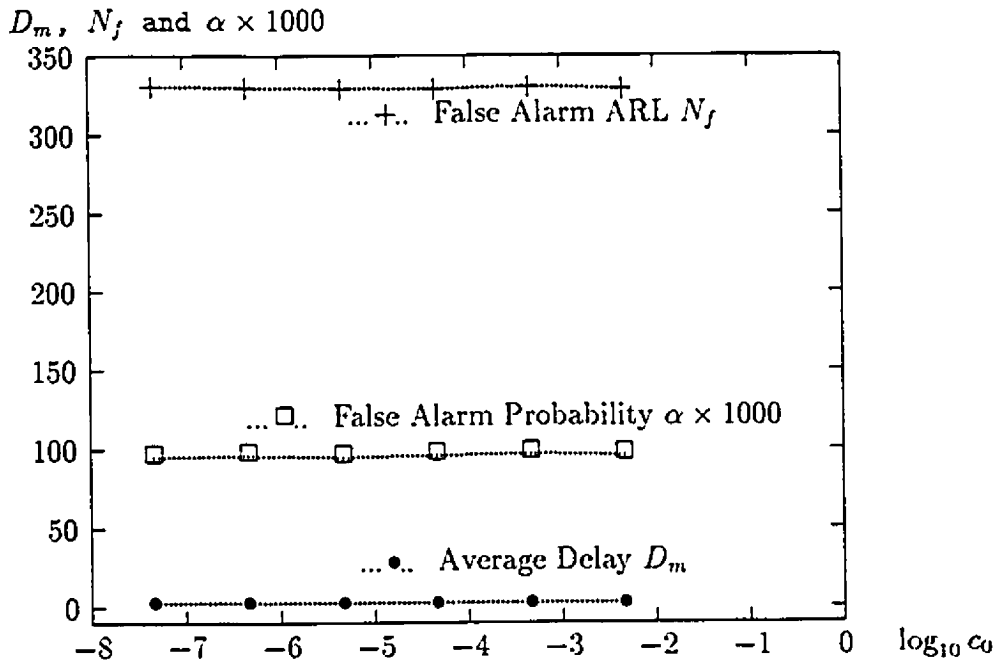
Figure 3.2c: *Performance of the new procedure for a wide range of c_0 values. Here, $m = 500$, $SNR = 6.0dB$, $\pi_0 = 4.62 \times 10^{-6}$ (i.e. $\log_{10} \pi_0 = -5.335$). As shown, performance is insensitive to different c_0.*

3.5 Nonasymptotic Performance: Simulation Results

In this section we are going to simulate and compare the performance of four different procedures: (i) the MAR procedure defined by Eqs.(3.2.1)(3.2.2) ; (ii) Page's CUSUM procedure (described in Subsection 2.2.1); (iii) a moving-window FSS procedure (described in Subsection 2.2.2); (iv) the GRS procedure. The four procedures have similar computational complexity.

3.5.1 Simulation Methodology

The modified Shiryayev criterion given by (3.1.10) is used to compare the performance of the four procedures. Given f_0, f_1, and fixed m, design parameters are chosen to ensure that the overall false alarm probability, $\alpha = 1 - \pi$, is the same for all four procedures. The average delay, D_m, and false alarm average run length $N_f = E(N|\overline{\Theta}_N)$ are compared.

Ensuring the same false alarm probability across the four different quickest detection algorithms is nontrivial, since each procedure lacks an exact, analytical threshold design procedure (the threshold bounds derived in section 3.6 may only be used as a guideline). Given change-time, m, and W identical simulations, let W_1 and W_2, respectively, denote the number of trials resulting in false alarm and correct detection, with corresponding sample sizes given by $N_1, N_2, ..., N_{W_1}$ and $N_1^*, N_2^*, ..., N_{W_2}^*$. For sufficiently large W_1 and W_2, we have

$$\pi \approx \frac{W_2}{W}, \tag{3.5.1}$$

$$N_f = E(N|\overline{\Theta}_N) \approx \frac{N_1 + N_2 + ... + N_{W_1}}{W_1}, \tag{3.5.2}$$

and

$$N_d = E(N|\Theta_N) \approx \frac{N_1^* + N_2^* + ... + N_{W_2}^*}{W_2}, \tag{3.5.3}$$

where N_f denotes the false alarm average run length and N_d the correct detection average

56

run length. Then the average delay is

$$D_m = N_d - m + 1. \qquad (3.5.4)$$

The 90% confidence intervals for the above quantities are $(\pi - \delta\pi, \pi + \delta\pi)$, $(N_f - \delta N_f, N_f + \delta N_f)$ and $(N_d - \delta N_d, N_d + \delta N_d)$ with (see [30])

$$\delta\pi = 1.645\sqrt{\frac{\pi(1-\pi)}{W}}, \qquad (3.5.5)$$

$$\delta N_f = 1.282\sqrt{\frac{S_f^2}{W_1}}, \qquad (3.5.6)$$

and

$$\delta N_d = 1.282\sqrt{\frac{S_d^2}{W_2}}, \qquad (3.5.7)$$

where

$$S_f^2 \approx \frac{\sum_{i=1}^{W_1}(N_i - N_f)^2}{W_1} = \frac{\sum_{i=1}^{W_1} N_i^2}{W_1} - N_f^2 \qquad (3.5.8)$$

and

$$S_d^2 \approx \frac{\sum_{i=1}^{W_1}(N_i^- - N_d)^2}{W_2} = \frac{\sum_{i=1}^{W_1} N_i^{-2}}{W_2} - N_d^2. \qquad (3.5.9)$$

The above relations are used to determine adequate sample sizes W_1 and W_2. Some typical values are listed in Table 3.2. Since π can never be exactly equated for the four procedures in experiments, we set the value of π for the MAR procedure to be no smaller than that of CUSUM, FSS, and GRS.

As mentioned in Section 3.4, CUSUM has only one design parameter, and the MAR procedure also has effectively one design parameter due to its insensitivity to the design parameter c_0. However, the FSS procedure is very sensitive to d and w, and therefore search is used to optimize the FSS procedure for each case. As is suggested in [102, 94], the algorithm (2.2.31) is used for the GRS procedure, where p and threshold A are design parameters.

In Subsections 3.5.2 and 3.5.3, we compare the performance of the new procedure, the FSS, CUSUM and a special case of the GRS ($p = 0$) for Gaussian and non-Gaussian

57

distributions respectively. In Subsection 3.5.4, we compare the performance of the new procedure with the GRS procedure with $p \neq 0$.

3.5.2 Simulation Results: Detection of a Change in the mean of a Gaussian Distribution

The Model

First we simulate the case of a constant signal in Gaussian noise, which arises in a variety of applications. The discrete-time model is

$$X_i = \begin{cases} N_i & \text{if } i < m \\ N_i + \mu_1 & \text{if } i \geq m \end{cases} \tag{3.5.10}$$

where $i = 1, 2, ...$, μ_1 is a constant, and N_i is i.i.d additive white Gaussian noise (AWGN). The pdfs for this model are

$$f_0(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x_i^2}{2\sigma^2}} \quad \text{if } i < m \tag{3.5.11}$$

and

$$f_1(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu_1)^2}{2\sigma^2}} \quad \text{if } i \geq m \tag{3.5.12}$$

where σ^2 is the variance (and also the energy) of the AWGN N_i.

The Performance for Different SNR's

For fixed change-time $m = 500$, we first compare the performance of three procedures for different SNRs. The simulation results are shown in Table 3.2 and Figures 3.3-3.6, where the overall false alarm probability $\alpha = 0.1$, the noise amplitude $\sigma = 0.25$, and $SNR \triangleq 20 \log \frac{\mu_1}{\sigma} (db)$. The corresponding 90% confidence intervals are shown in Table 3.2.

As predicted by the asymptotic convergence property, it is seen that the new procedure outperforms the other procedures significantly as the $SNR \to -\infty (dB)$. For -20

58

db SNR, the new procedure is over 31.8% faster than CUSUM and 38.6% faster than moving-window FSS procedure. As is evident in Figure 3.6, the false alarm ARL of the new procedure is significantly longer than the other procedures, which together with the average delay performance results shown in Figures 3.3–3.5, shows that the new procedure outperforms the CUSUM, FSS, and GRS procedures in these cases. It may seem surprising that the GRS procedure outperforms CUSUM. However, this comparison is based on the modified Shiryayev criterion rather than Lorden's criterion.

The Performance for Different Change Times

The performance also varies with change-time, m, as shown in Figures 3.7-3.9. With the overall false alarm probability $\alpha = 0.1$, the noise amplitude $\sigma = 0.25$, and $SNR = -6.0db$, we observe that for all change-times tried, D_m for the new procedure is the shortest, while N_f is the longest. By examining Fig.3.7, we note that for small m, CUSUM is outperformed by the FSS procedure, which confirms earlier results reported in [31, 41, 42]. Again it is observed that GRS has shorter average delays than CUSUM. Also, note that when $m \to +\infty$, D_m should go to $+\infty$ for a fixed $\alpha(< 1)$. Particularly, when $m \to +\infty$ and $\alpha(< 1)$ is fixed, the decision thresholds for GRS, FSS and CUSUM approach $+\infty$, while the design parameters for the new procedure tend to zero.

The Performance for Different False Alarm Probabilities

The performance also varies with different α, as shown in Figures 3.10 and 3.11. Here the change time is $m = 500$, the noise amplitude $\sigma = 0.25$, and $SNR = -6.0db$. It is clear that for all false alarm probabilities tried, D_m for the new procedure is the shortest, while N_f is the longest. It is noteworthy that the relative superiority of the new procedure improves as α goes to 0, as predicted by the asymptotic convergence property.

3.5.3   Simulation Results: Detection of a Change in non-Gaussian Distribution

We also considered the case of quickest detection of a change in a non-Gaussian distribution where

$$f_0(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, \tag{3.5.13}$$

and

$$f_1(x) = \frac{x}{\sigma^2} e^{-\frac{x^2 + A^2}{2\sigma^2}} I_0(\frac{Ax}{\sigma^2}). \tag{3.5.14}$$

The pdfs f_0 and f_1 are Rayleigh and Rician densities respectively, and I_0 is the modified Bessel function of the first kind and order zero. This model describes envelope detection of a sinusoidal signal in Gaussian noise [107]. Here, the signal-to-noise-ratio, defined by $SNR \triangleq 20 \log \frac{A}{\sigma}(db)$, is varied, while m and α are held fixed at 500 and 0.1, respectively. From Figures 3.12-3.15, it is clear that for all SNR's tried, D_m for the new procedure is the shortest, while N_f is the longest. We observe performance results similar to those described for the Gaussian case. Note that GRS also outperforms CUSUM for this case.

3.5.4   Performance Comparison of MAR and GRS with $p > 0$

In the last two subsections a special case of the GRS procedure when $p = 0$ was compared to the new procedure and other existing procedures. It was observed that the new procedure and the GRS outperforms CUSUM and FSS according to the modified Shiryayev criterion. Under the same criterion, it appears that the new procedure has better performance than the special case of GRS. However, this is not the case for the GRS procedure if $p \neq 0$. When $p > 0$, the GRS procedure assumes that the change occurs with greater probability for smaller change-time m. Since the GRS procedure is optimal for a geometric change-time distribution, we would expect that the GRS procedure may perform well if the actual change-time is not much greater than the mean of

60

the prior geometric distribution. i.e.,

$$
\begin{aligned}
E(M) &= \sum_{m=1}^{+\infty} m P(M=m) \\
&= \sum_{m=1}^{+\infty} m(1-p)^{m-1}p \\
&= \frac{1}{p}.
\end{aligned}
\tag{3.5.15}
$$

Using the same Gaussian distribution model, we simulated the GRS procedure for $p = 0.01$, i.e., $E(M) = 100$. As shown in Figures 3.16-3.19, the MAR procedure outperforms GRS if the SNR is not very small. However, from Figures 3.20 and 3.21 it appears that GRS performs better than the MAR procedure if the SNR is very small and the actual change-time m is less than 3500.

61

μ_1	SNR	$\alpha \pm \delta\alpha$	$N_f \pm \delta N_f$	$D_m \pm \delta D_m$	π_0	c_0
.025	−20.0	$.232 \pm .003$	365.7 ± 1.1	161.0 ± 0.8	8.7×10^{-4}	.87
		$.114 \pm .002$	379.3 ± 1.4	226.2 ± 1.0	8.7×10^{-4}	.087
		$.100 \pm .002$	379.1 ± 1.5	235.7 ± 1.0	8.7×10^{-4}	.0087
		$.100 \pm .002$	381.6 ± 1.5	238.2 ± 1.0	$\pi_0 = c_0 = 8.7 \times 10^{-4}$	
		$.098 \pm .002$	380.6 ± 1.5	237.9 ± 1.0	8.7×10^{-4}	8.7×10^{-5}
		$.101 \pm .002$	379.0 ± 1.5	237.3 ± 1.0	8.7×10^{-4}	8.7×10^{-6}
		$.101 \pm .002$	380.9 ± 1.5	237.6 ± 1.0	8.7×10^{-4}	8.7×10^{-7}
.075	−10.5	$.100 \pm .002$	340.5 ± 2.0	$72.6 \pm .3$	6.52×10^{-5}	.0652
		$.097 \pm .002$	342.4 ± 2.0	$73.7 \pm .3$	6.52×10^{-5}	.00652
		$.093 \pm .002$	339.7 ± 2.0	$74.1 \pm .3$	6.52×10^{-5}	.000652
		$.094 \pm .002$	341.4 ± 2.0	$73.9 \pm .3$	$\pi_0 = c_0 = 6.52 \times 10^{-5}$	
		$.095 \pm .002$	339.8 ± 2.0	$74.1 \pm .3$	6.52×10^{-5}	6.52×10^{-6}
		$.097 \pm .002$	341.8 ± 2.0	$74.2 \pm .3$	6.52×10^{-5}	6.52×10^{-7}
		$.095 \pm .002$	338.7 ± 2.0	$74.1 \pm .3$	6.52×10^{-5}	6.52×10^{-8}
.025	0.0	$.099 \pm .002$	332.8 ± 2.1	$11.6 \pm .02$	7.84×10^{-6}	.00784
		$.099 \pm .002$	330.5 ± 2.1	$11.6 \pm .02$	7.84×10^{-6}	.000784
		$.100 \pm .002$	327.7 ± 2.1	$11.5 \pm .02$	7.84×10^{-6}	7.84×10^{-5}
		$.098 \pm .002$	333.1 ± 2.1	$11.6 \pm .02$	$\pi_0 = c_0 = 7.84 \times 10^{-6}$	
		$.101 \pm .002$	329.0 ± 2.1	$10.5 \pm .02$	7.84×10^{-6}	7.84×10^{-7}
		$.099 \pm .002$	330.9 ± 2.1	$11.5 \pm .02$	7.84×10^{-6}	7.84×10^{-8}
		$.100 \pm .002$	330.9 ± 2.1	$11.6 \pm .02$	7.84×10^{-6}	7.84×10^{-9}
.499	6.0	$.096 \pm .002$	329.1 ± 2.2	$3.76 \pm .05$	4.62×10^{-6}	.00462
		$.097 \pm .002$	330.6 ± 2.2	$3.76 \pm .05$	4.62×10^{-6}	.000462
		$.096 \pm .002$	329.1 ± 2.2	$3.76 \pm .05$	4.62×10^{-6}	4.62×10^{-5}
		$.095 \pm .002$	329.4 ± 2.2	$3.76 \pm .05$	$\pi_0 = c_0 = 4.62 \times 10^{-6}$	
		$.096 \pm .002$	330.1 ± 2.2	$3.75 \pm .05$	4.62×10^{-6}	4.62×10^{-7}
		$.095 \pm .002$	330.9 ± 2.2	$3.76 \pm .04$	4.62×10^{-6}	4.62×10^{-8}
		$.097 \pm .002$	329.3 ± 2.2	$3.78 \pm .05$	4.62×10^{-6}	4.62×10^{-9}

Table 3.1: *Insensitivity of design parameter c_0.*

62

	μ_1	SNR	$\alpha \pm \delta\alpha$	$N_f \pm \delta N_f$	$D_m \pm \delta D_m$	Design Parameters
New	.025	−20.0	.10 ± .002	381.6 ± 1.5	238.2 ± 1.0	$\pi_0 = c_0 = 8.7 \times 10^{-4}$
CUSUM	.025	−20.0	.103 ± 0.002	322.6 ± 1.9	350.5 ± 1.6	$\ln B = 2.878$
FSS	.025	−20.0	.101 ± 0.002	333.2 ± 1.6	388.5 ± 1.8	$\ln d = 2.259, w = 218$
GRS	.025	−20.0	.104 ± .002	372.8 ± 1.5	276.3 ± 1.2	$p = 0, A = 2002.9$
New	.044	−15.0	.098 ± 0.002	355.9 ± 1.8	143.2 ± 0.6	$\pi_0 = c_0 = 2.13 \times 10^{-4}$
CUSUM	.044	−15.0	.103 ± 0.002	297.2 ± 2.1	175.6 ± 0.7	$\ln B = 3.944$
FSS	.044	−15.0	.104 ± 0.002	297.3 ± 2.1	196.2 ± 0.9	$\ln d = 3.139, w = 121$
GRS	.044	−15.0	.111 ± .002	324.4 ± 1.8	159.4 ± .6	$p = 0, A = 2727.6$
New	.074	−10.5	.094 ± 0.002	341.4 ± 2.0	73.9 ± 0.3	$\pi_0 = c_0 = 6.52 \times 10^{-5}$
CUSUM	.074	−10.5	.099 ± 0.002	278.4 ± 2.3	86.4 ± 0.3	$\ln B = 4.928$
FSS	.074	−10.5	.101 ± 0.002	289.0 ± 2.1	91.3 ± 0.3	$\ln d = 3.42, w = 106$
GRS	.074	−10.5	.114 ± .002	285.2 ± 2.1	81.7 ± .3	$p = 0, A = 3078.3$
New	.125	−6.0	.093 ± .002	332.7 ± 2.2	34.6 ± 0.1	$\pi_0 = c_0 = 2.29 \times 10^{-5}$
CUSUM	.125	−6.0	.099 ± 0.002	261.1 ± 2.5	39.3 ± 0.1	$\ln B = 5.781$
FSS	.125	−6.0	.103 ± 0.002	267.0 ± 2.3	42.3 ± 0.1	$\ln d = 4.041, w = 51$
GRS	.125	−6.0	.102 ± .002	266.9 ± 2.4	38.5 ± .1	$p = 0, A = 3405.3$
New	.25	0.0	.098 ± 0.002	333.1 ± 2.1	11.6 ± .02	$\pi_0 = c_0 = 7.84 \times 10^{-6}$
CUSUM	.25	0.0	.100 ± 0.002	253.6 ± 2.6	12.8 ± .02	$\ln B = 6.611$
FSS	.25	0.0	.104 ± 0.002	249.4 ± 2.5	13.8 ± .03	$\ln d = 5.440, w = 12$
GRS	.25	0.0	.097 ± .002	251.3 ± 2.6	12.7 ± .02	$p = 0, A = 2678.6$
New	.5	6.0	.095 ± .002	329.4 ± 2.2	3.8 ± .05	$\pi_0 = c_0 = 4.62 \times 10^{-6}$
CUSUM	.5	6.0	.112 ± 0.002	243.6 ± 2.4	4.0 ± .05	$\ln B = 6.768$
FSS	.5	6.0	.101 ± 0.002	247.0 ± 2.6	4.4 ± .06	$\ln d = 5.878, w = 4$
GRS	.5	6.0	.095 ± .002	246.0 ± 2.6	4.1 ± .05	$p = 0, A = 1624.7$

Table 3.2: *Simulation results for the new procedure, CUSUM, FSS and GRS procedures. The number of simulations $W = 50000$. The change-time $m = 500$. f_0 and f_1 are Gaussian with mean 0 and μ_1 respectively, and a common variance $\sigma^2 = 0.25^2$. $SNR \triangleq 20\log\frac{\mu_1}{\sigma}(db)$. N_f is the false alarm ARL and D_m is average delay.*

63
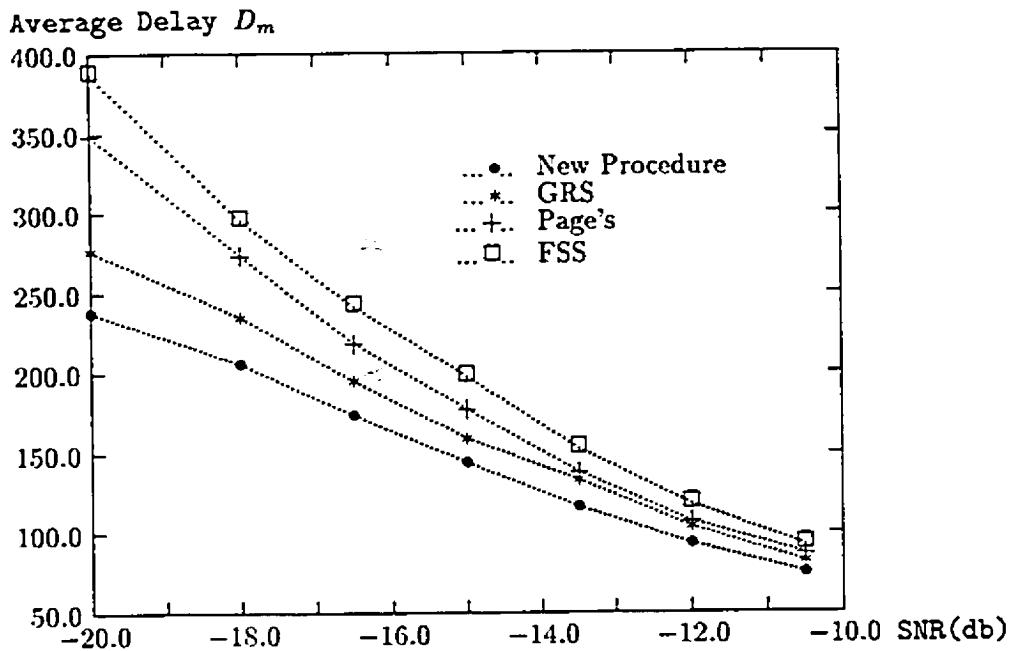
Average Delay D_m

Figure 3.3: *Quickest detection of a change in the mean of Gaussian distribution. The false alarm probability is* $\alpha = 0.1$. *The change-time is* $m = 500$. *The average delay v.s. SNR (weak signal).*
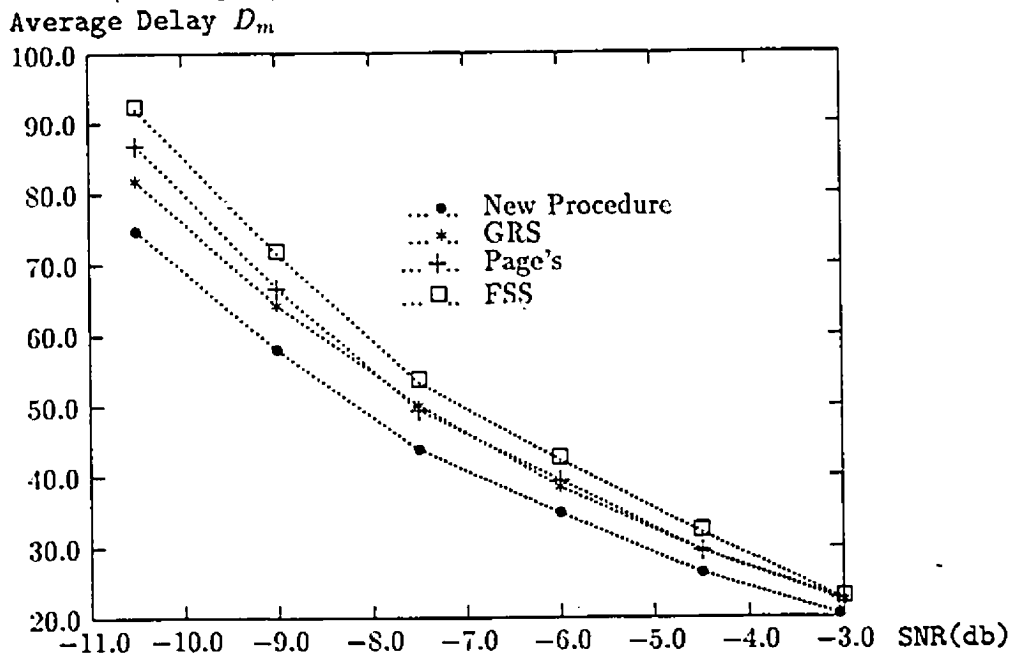
Average Delay D_m

Figure 3.4: *Quickest detection of a change in the mean of Gaussian distribution. The false alarm probability is* $\alpha = 0.1$. *The change-time is* $m = 500$. *The average delay v.s. SNR (weak signal).*

64
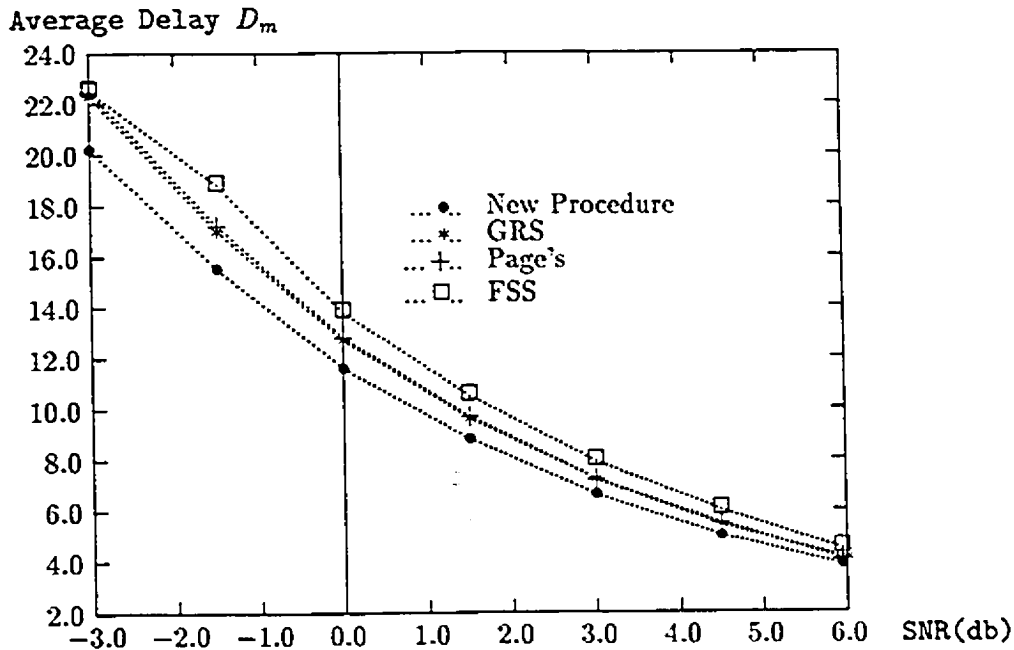
Average Delay D_m

Figure 3.5: *Quickest detection of a change in the mean of Gaussian distribution.* $\alpha = 0.1$.

$m = 500$. *The average delay v.s. SNR (strong signal).*

False Alarm ARL N_f
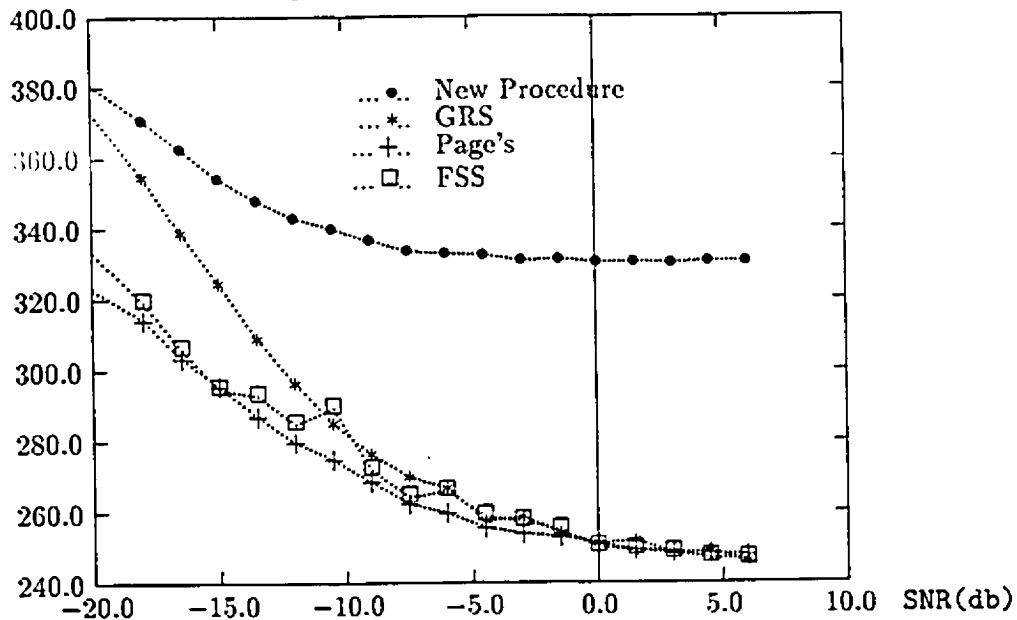
Figure 3.6: *Quickest detection of a change in the mean of Gaussian distribution. The false alarm ARL N_f v.s. SNR. $m = 500$. $\alpha = 0.1$. The false alarm rate is described by α and N_f together. For given α, a larger N_f implies a smaller false alarm rate.*

65
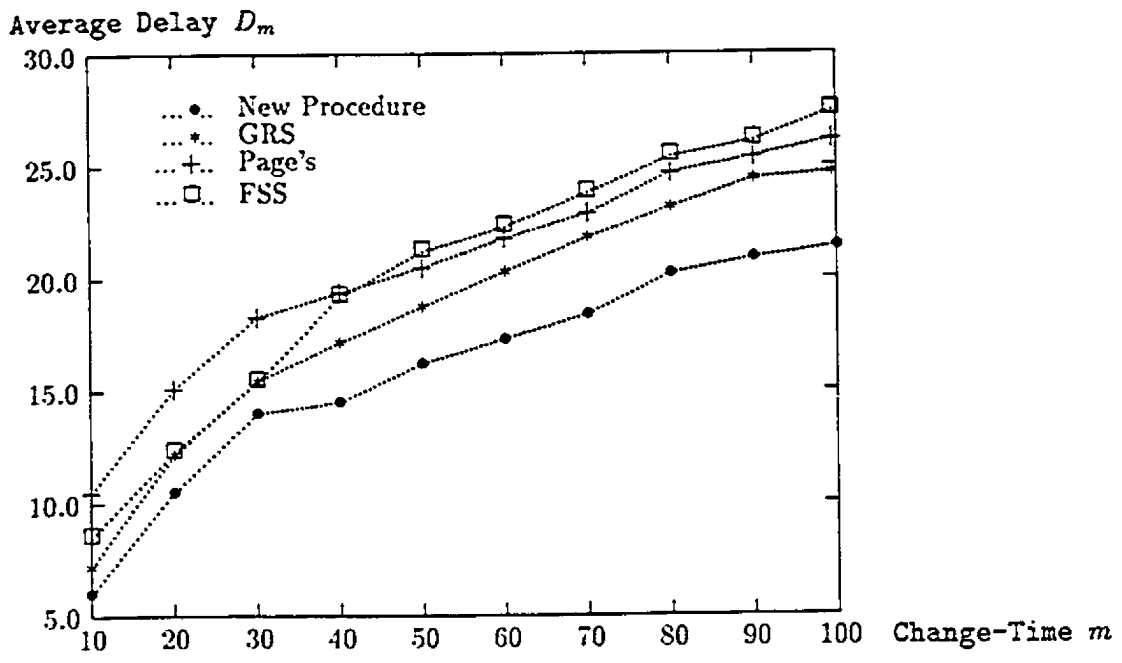
Average Delay D_m

Figure 3.7: *Quickest detection of a change in the mean of Gaussian distribution. The average delay D_m v.s. change-time m. The signal-to-noise ratio $SNR = -6.0db$. The false alarm probability is $\alpha = 0.1$.*
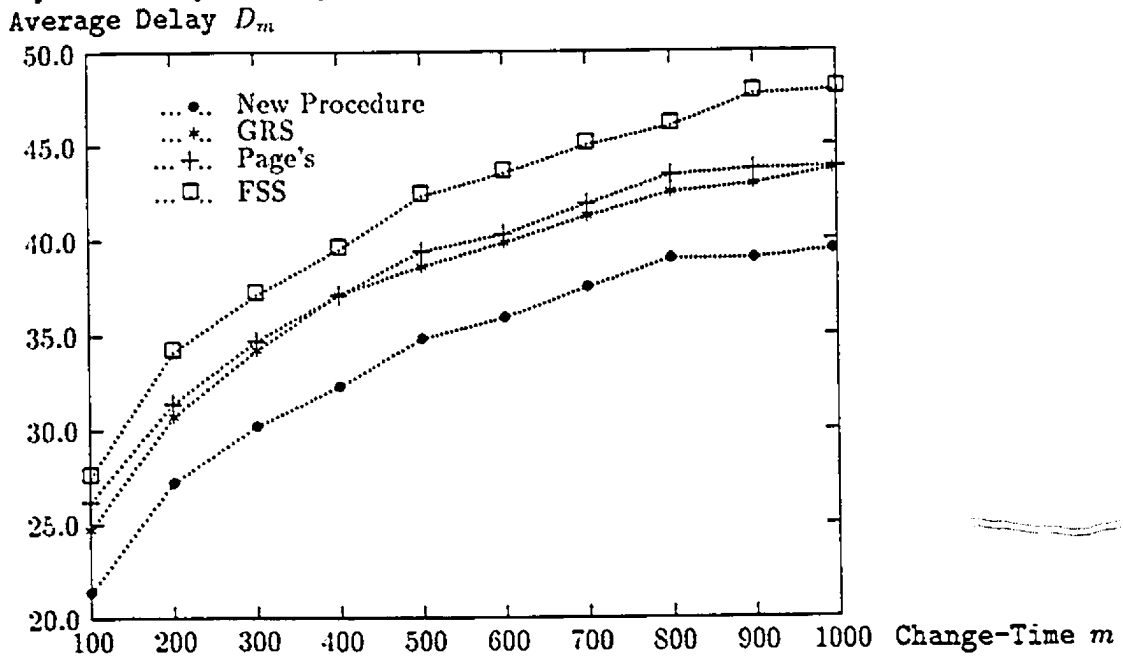
Average Delay D_m

Figure 3.8: *Quickest detection of a change in the mean of Gaussian distribution. The average delay D_m v.s. change-time m. The signal-to-noise ratio $SNR = -6.0db$. The false alarm probability is $\alpha = 0.1$.*

66
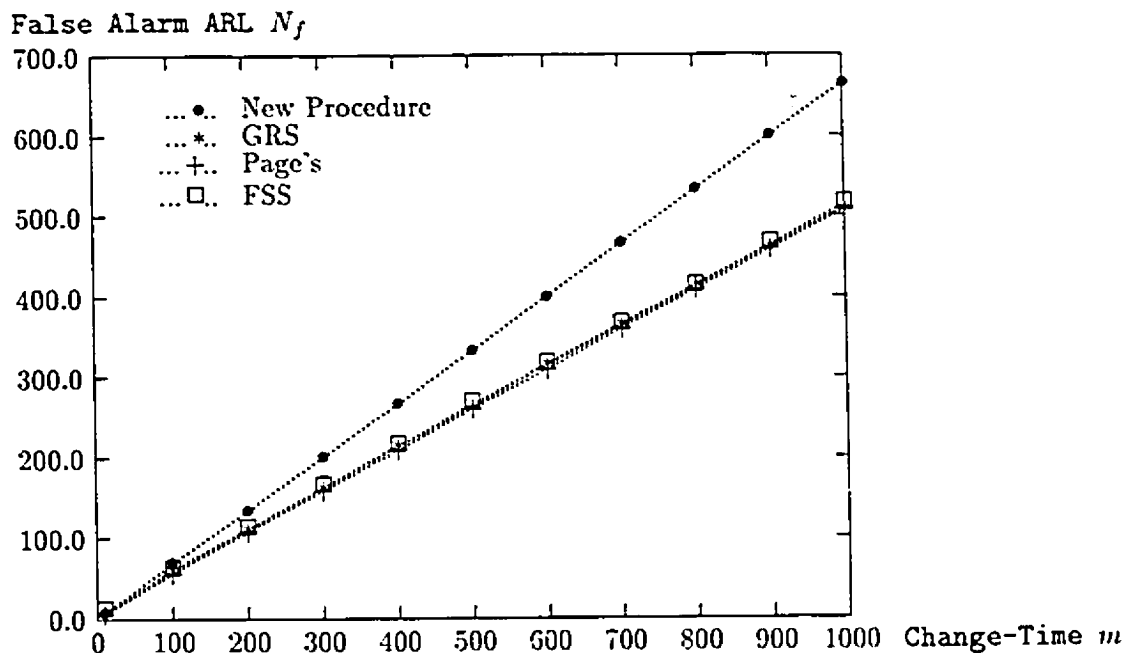
False Alarm ARL N_f

Figure 3.9: *Quickest detection of a change in the mean of Gaussian distribution. The false alarm average run length N_f v.s. change-time m. The signal-to-noise ratio $SNR = -6.0db$. The false alarm probability is $\alpha = 0.1$.*

67
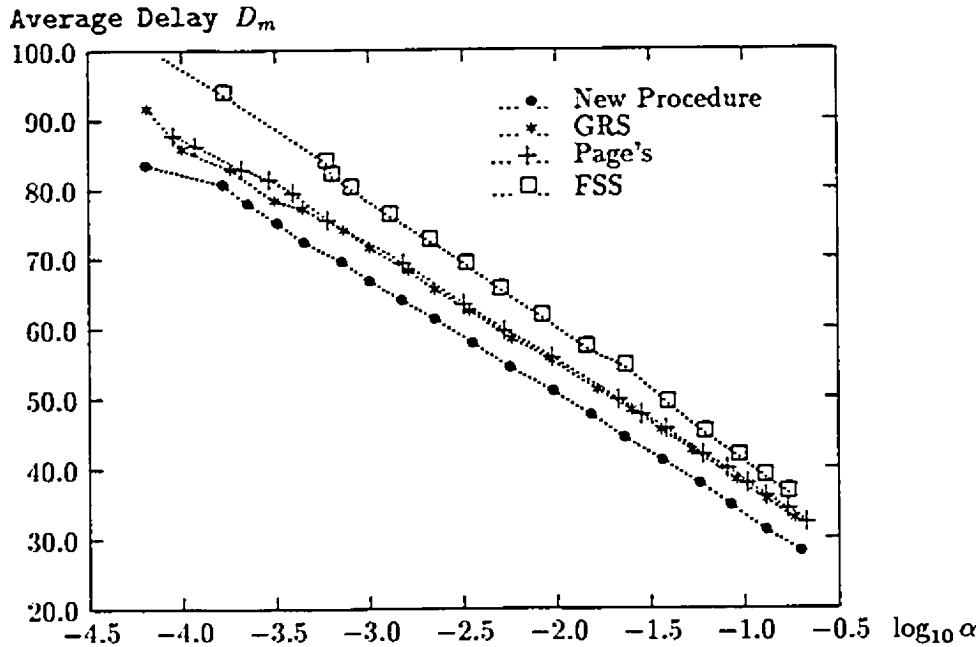
Average Delay D_m

Figure 3.10: *Quickest detection of a change in the mean of Gaussian distribution. The average delay D_m v.s. false alarm probability α. The SNR is $-6.0db$. $m = 500$.*
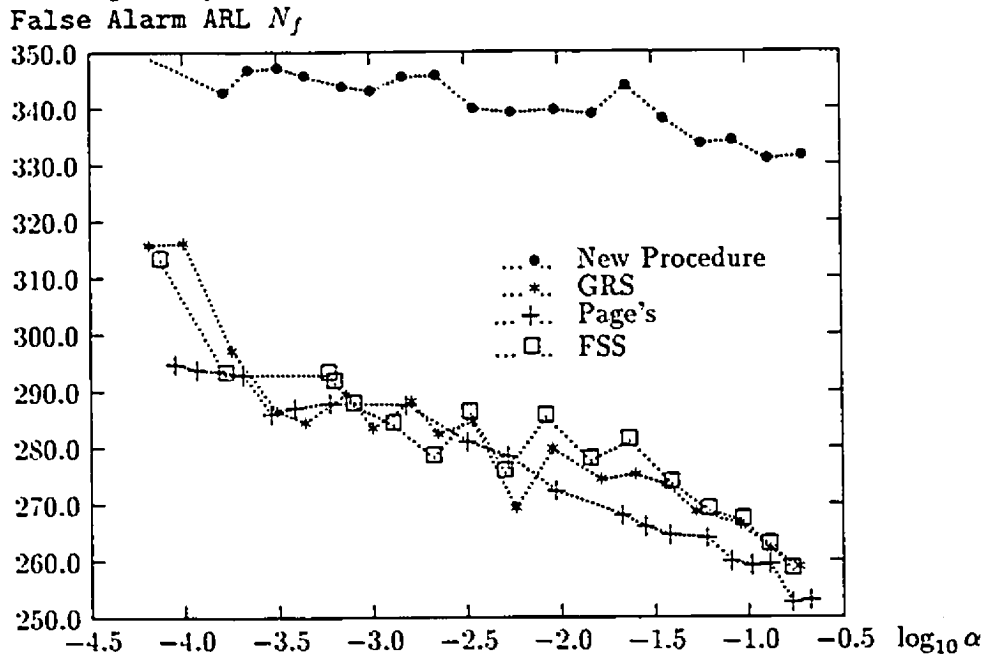
False Alarm ARL N_f

Figure 3.11: *Quickest detection of a change in the mean of Gaussian distribution. The false alarm ARL N_f v.s. false alarm probability α. The SNR is $-6.0db$. $m = 500$.*

68
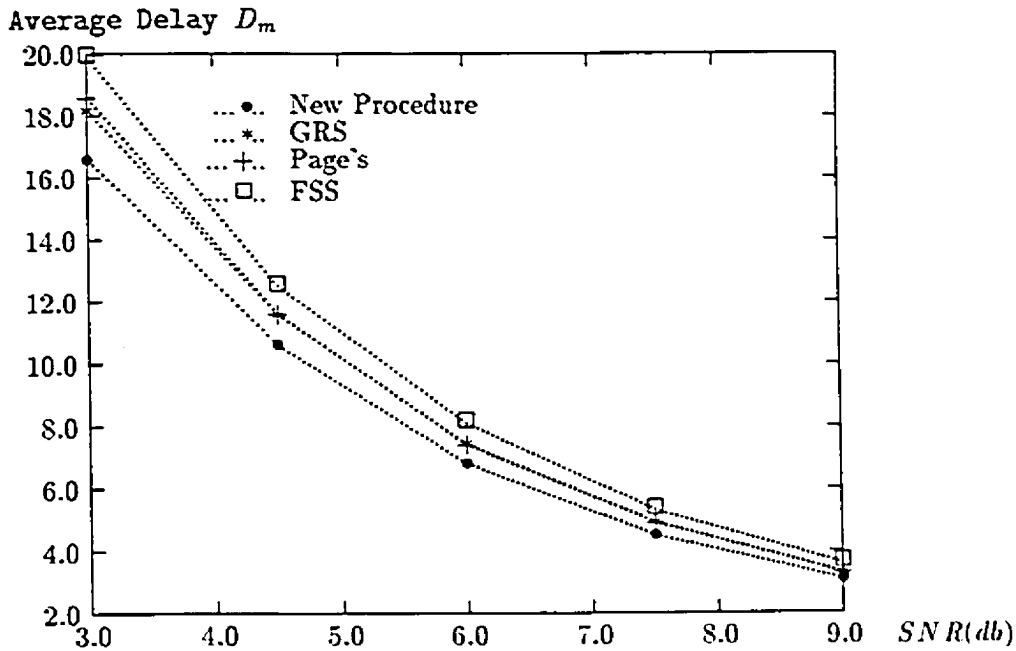
Average Delay D_m

Figure 3.12: *Quickest detection of a change in non-Gaussian distribution. The average delay D_m v.s. SNR (strong signal). The change-time is $m = 500$. The false alarm probability is $\alpha = 0.1$.*
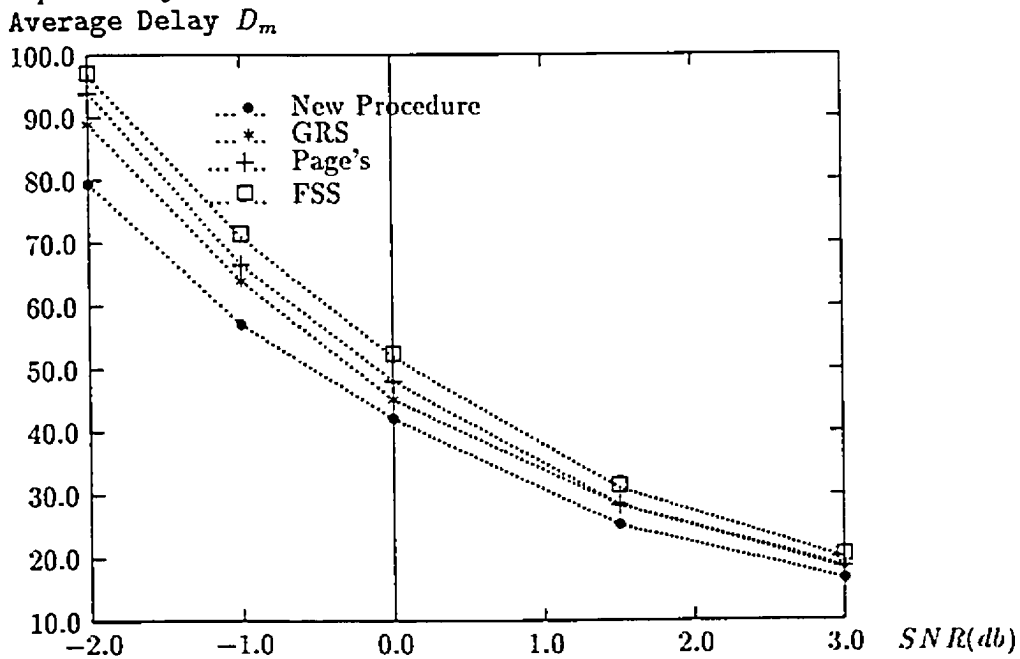
Average Delay D_m

Figure 3.13: *Quickest detection of a change in non-Gaussian distribution. The average delay D_m v.s. SNR (weak signal). The change-time is $m = 500$. The false alarm probability is $\alpha = 0.1$.*

69
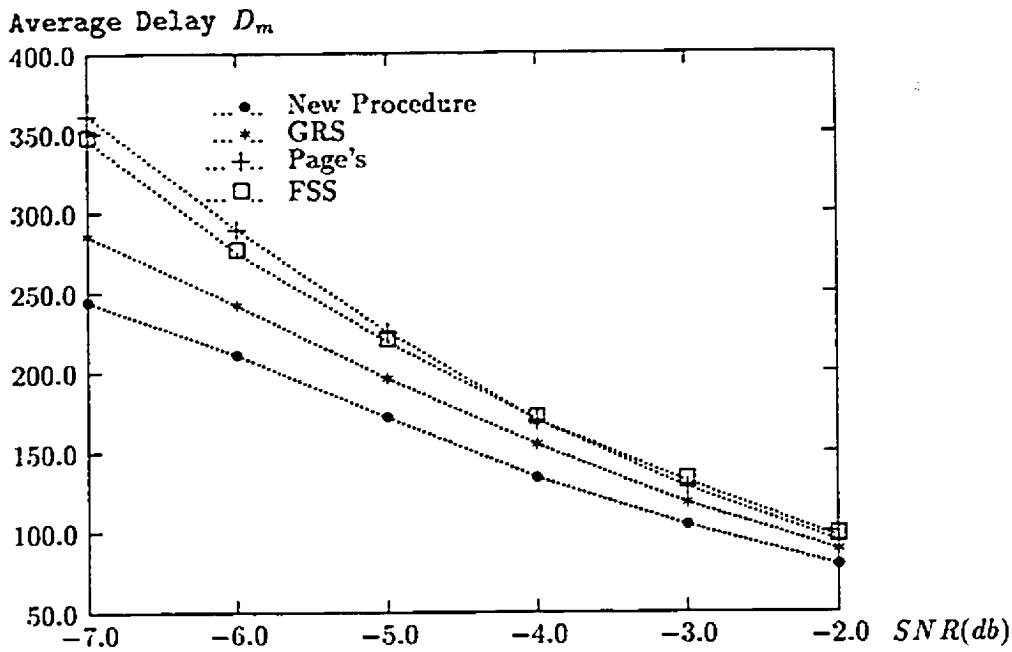
Average Delay D_m

Figure 3.14: *Quickest detection of a change in non-Gaussian distribution. The average delay D_m v.s. SNR (weak signal). The change-time is $m = 500$. The false alarm probability is $\alpha = 0.1$.*
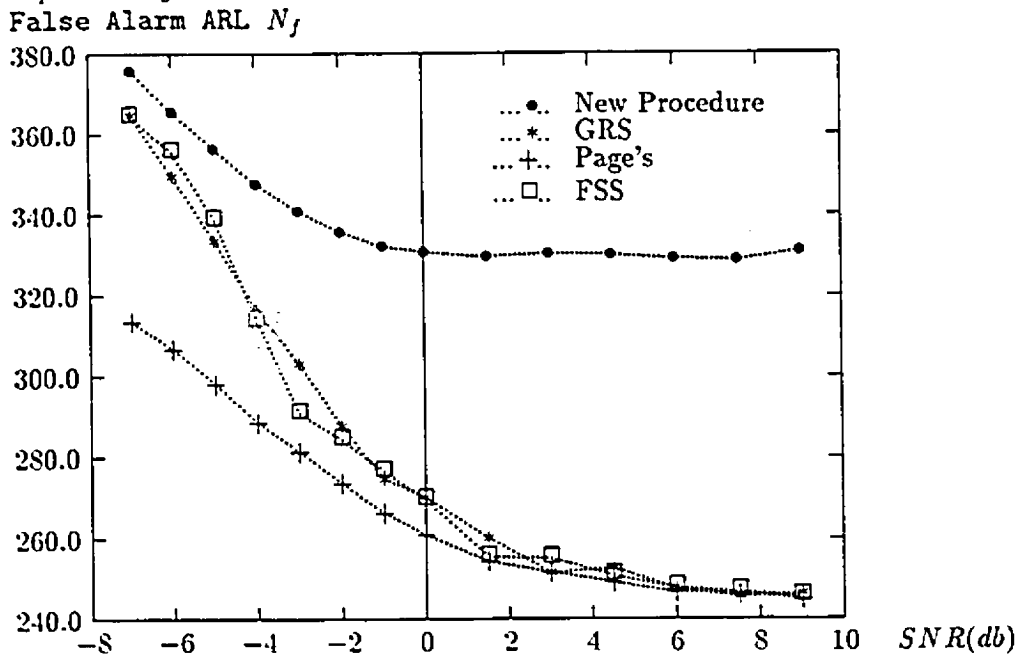
False Alarm ARL N_f

Figure 3.15: *Quickest detection of a change in non-Gaussian distribution. The false alarm ARL N_f v.s. SNR. The change-time is $m = 500$. The false alarm probability is $\alpha = 0.1$.*

70
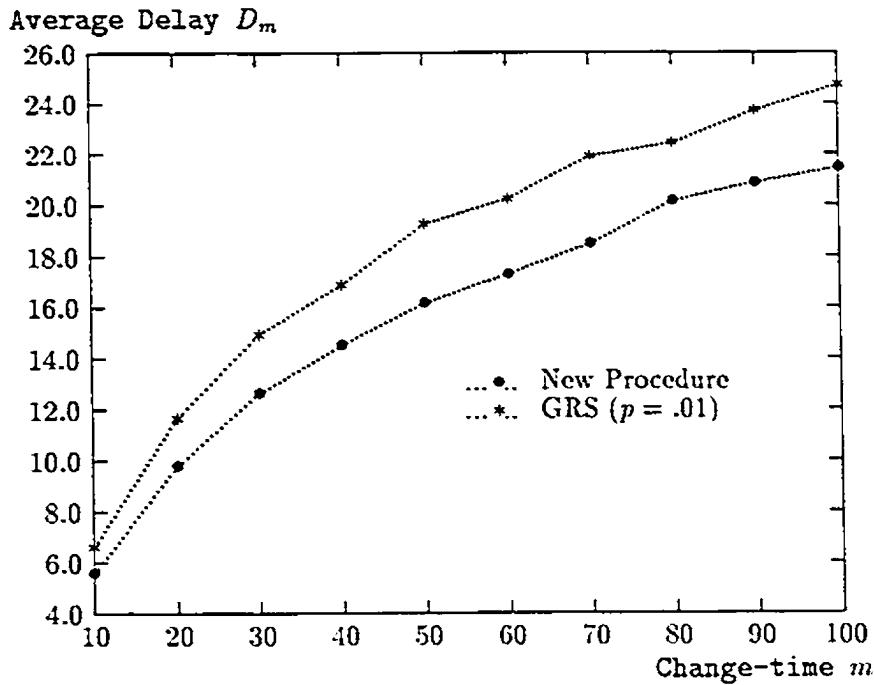
Figure 3.16: *Quickest detection of a change in the mean of Gaussian distribution.* $\alpha = 0.1$. $SNR = -6.dB$.
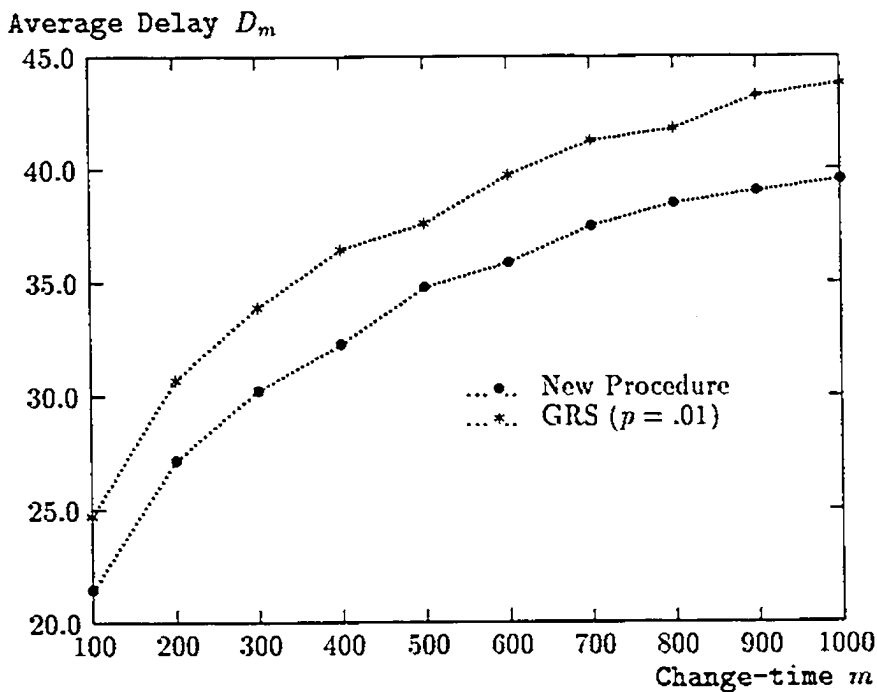
Figure 3.17: *Quickest detection of a change in the mean of Gaussian distribution.* $\alpha = 0.1$. $SNR = -6.dB$.

Average Delay D_m
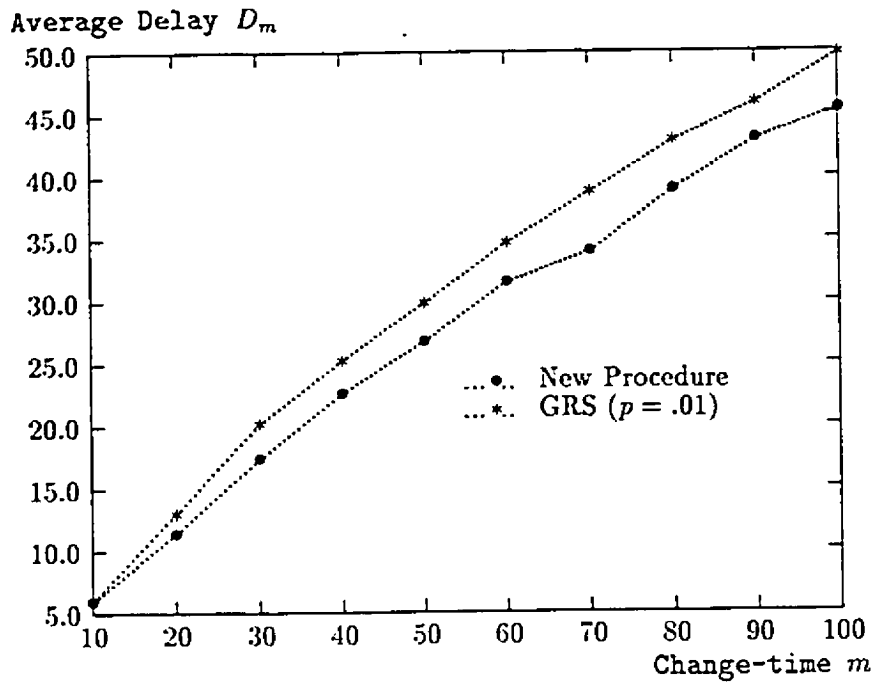
Figure 3.18: *Quickest detection of a change in the mean of Gaussian distribution.* $\alpha = 0.1. \ SNR = -12.dB.$

Average Delay D_m
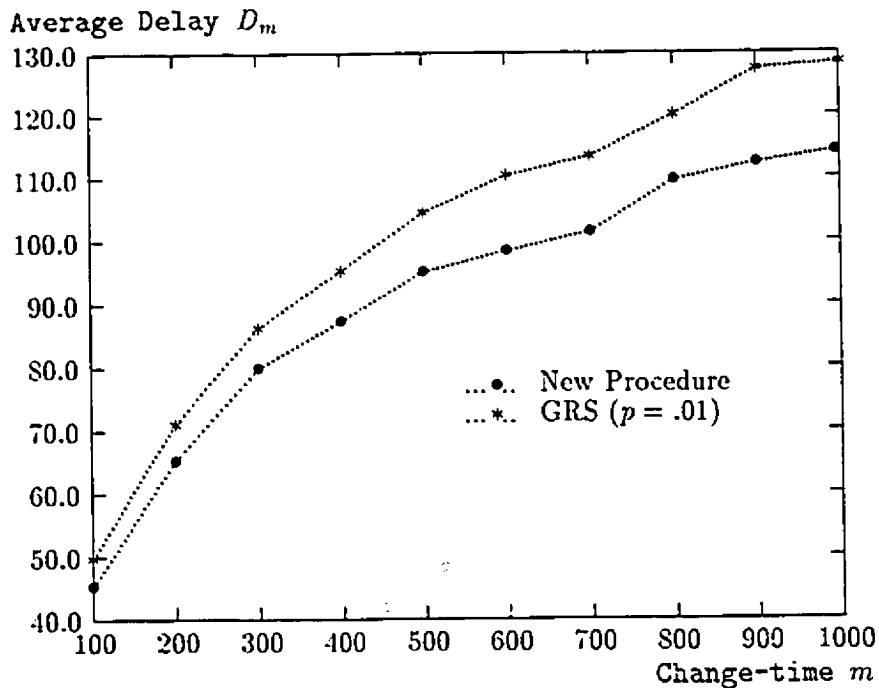
Figure 3.19: *Quickest detection of a change in the mean of Gaussian distribution.* $\alpha = 0.1. \ SNR = -12.dB.$

72

Average Delay D_m
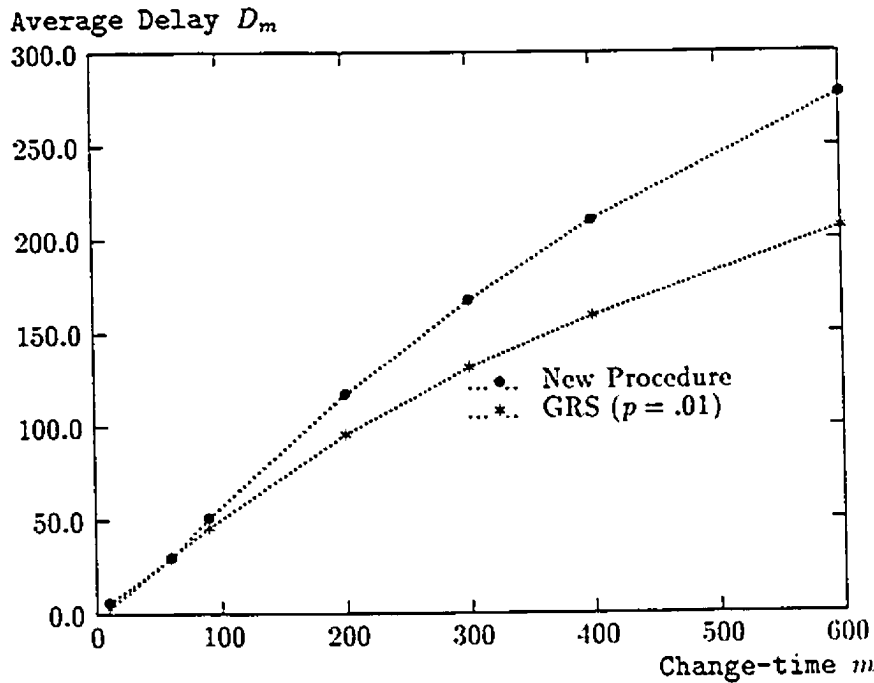
Figure 3.20: *Quickest detection of a change in the mean of Gaussian distribution.* $\alpha =$ 0.1. $SNR = -20.dB$.

Average Delay D_m
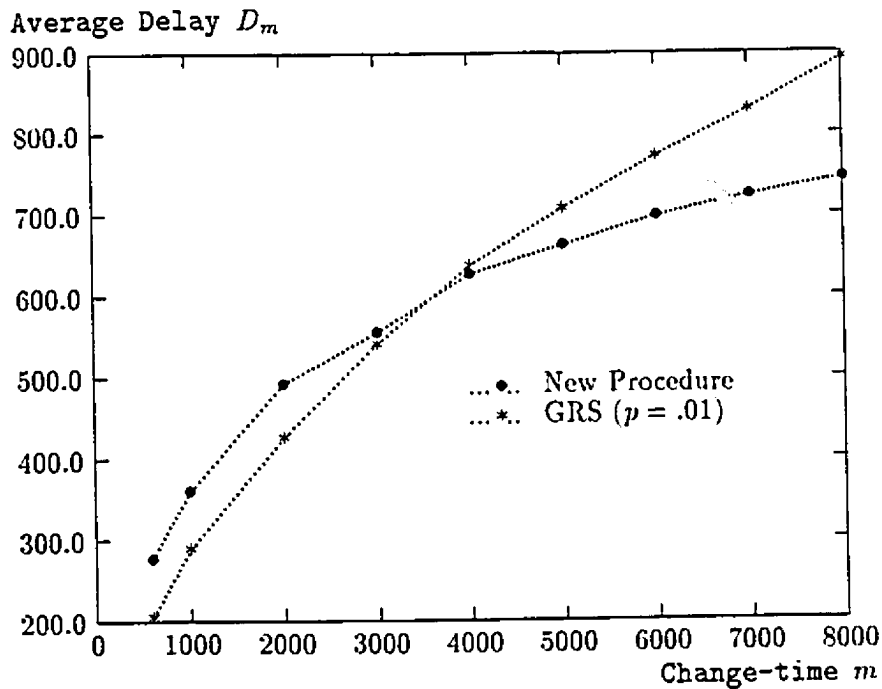
Figure 3.21: *Quickest detection of a change in the mean of Gaussian distribution.* $\alpha =$ 0.1. $SNR = -20.dB$.

73

3.6 False Alarm Probability of Change Detection Procedures

When no change ever occurs (i.e. H_0 is true for all $m \geq 1$), the false alarm average run length (ARL) has been studied by a number of researchers [89, 102, 41, 77, 92], mainly for designing thresholds under the Lorden or QC criteria. However, the problem of computing false alarm probabilities (FAP) for change-detection procedures has not been carefully studied in the existing literature. In this section we will present a number of properties concerning the false alarm probability of three change-detection procedures. These properties may be helpful for choosing thresholds and for understanding the relationship of the change-detection procedures with respect to their thresholds and model parameters.

Two reasonably tight FAP upper bounds for the CUSUM and MAR procedures are derived. The almost sure (a.s.) termination of CUSUM, MAR and GRS is proved. The CUSUM procedure will be first studied, followed by a study of the MAR procedure defined by (3.2.1) and (3.2.2). The GRS procedure will be addressed in the last subsection.

3.6.1 False Alarm Probability of CUSUM

I. Almost Sure Termination of CUSUM

We first present two propositions showing the almost sure (a.s.) termination of CUSUM.

PROPOSITION 3.3 *If a change never occurs, i.e.*

$$m = +\infty, \tag{3.6.1}$$

then the false alarm probability of the CUSUM procedure equals 1, i.e.

$$\alpha = P_0\{N < +\infty\} = P_\infty\{N < +\infty\} = 1. \tag{3.6.2}$$

74

PROOF: Notice that the CUSUM is equivalent to a repeated SPRT with upper threshold B and lower threshold 1, which we denote by $SPRT(1,B)$ (see Figure 2.1). Let a_{0k} denote the event that the kth SPRT accepts H_0. For $m = +\infty$, the false alarm probability

$$\alpha = 1 - P_0\{\bigcap_{k=1}^{\infty} a_{0k}\} = 1 - \lim_{n \to \infty} \prod_{k=1}^{n} P_0\{a_{0k}\} \qquad (3.6.3)$$

where $a_{01}, a_{20}, \ldots$ are mutually independent. It should be noted that an $SPRT(1,B)$ terminates with probability 1 [119, 44]. Since

$$P_0\{a_{0k}\} = P_0\{SPRT(1,B) \text{ accepts } H_0\} = 1 - \alpha_{SPRT(1,B)} \qquad (3.6.4)$$

where the fact that an $SPRT(1,B)$ terminates with probability 1 is again used. The error probability $\alpha_{SPRT(1,B)}$ of the first type of the $SPRT(1,B)$ satisfies $0 < \alpha_{SPRT(1,B)} < 1$ for $f_0 \neq f_1$. Therefore

$$\alpha = 1 - \lim_{n \to \infty} (1 - \alpha_{SPRT(1,B)})^n = 1. \qquad (3.6.5)$$

QED

The following proposition states that the CUSUM terminates with probability one for any change-time $m \geq 1$.

PROPOSITION 3.4 *For any change-time $m \geq 1$, the CUSUM procedure terminates with probability 1, i.e.*

$$P_m\{N < +\infty\} = 1. \qquad (3.6.6)$$

PROOF: There are two cases: (i) $m = +\infty$, and (ii) $m < +\infty$.

For $m = +\infty$, the proposition is actually a re-statement of Proposition 3.3, since

$$\alpha = P_m\{N < m\}|_{m=+\infty} = P_{\infty}\{N < +\infty\} \qquad (3.6.7)$$

which equals 1 according to Proposition 3.3.

For $m < +\infty$, let $D_1, D_2, \ldots, D_n$ denote the event that the first, second, ... and nth $SPRT(1,B)$ start before the change-time m and accept H_0 respectively, and $E_1, E_2, \ldots$

75

denote the event that the first, second, ... SPRT(1,B) start after or at m and accept H_0 respectively (see Figure 2.1). Then

$$P_m\{N < +\infty\} = 1 - P_m\{\bigcap_{i=1}^{n} D_i \bigcap_{j=1}^{+\infty} E_j\}$$

$$\geq 1 - P_m\{\bigcap_{j=1}^{+\infty} E_j\}$$

$$= 1 - \lim_{n' \to \infty} \prod_{j=1}^{n'} P_1\{E_j\} \qquad (3.6.8)$$

Since

$$P_1\{E_j\} = P_1\{\text{SPRT(1,B) accepts } H_0\} = \beta_{SPRT(1,B)} < 1 \qquad (3.6.9)$$

then

$$1 \geq P_m\{N < +\infty\} \geq 1 - \lim_{n' \to \infty} \beta_{SPRT(1,B)}^{n'} = 1 \qquad (3.6.10)$$

i.e.

$$P_m\{N < +\infty\} = 1. \qquad (3.6.11)$$

QED

II. False Alarm Probability Bound of CUSUM

Now we present a lower and an upper bound for the false alarm probability (FAP) of the CUSUM procedure.

PROPOSITION 3.5 *The false alarm probability $\alpha(m, B)$ of the CUSUM procedure satisfies the following inequalities*

$$\alpha(m, B) \geq 1 - [1 - P_0(l_1 \geq B)]^{m-1} = \alpha_L(m, B) \qquad (3.6.12)$$

and

$$\alpha(m, B) \leq \alpha_U(m, B) = 1 - [1 - P_0(l_1 \geq B)]^{m-1} + \sum_{k=2}^{m-1} (m - k)P_0\{S_1^k \geq B\}$$

$$\approx 1 - [1 - P_0(z_1 \geq \ln B)]^{m-1} + \sum_{i=2}^{k'-1} (m - i)P_0\{Z_i \geq \ln B\} +$$

$$\sum_{k=k'}^{m-1} (m - k)Q(\frac{\ln B - k\mu_z}{\sqrt{k}\sigma_z}) \qquad (3.6.13)$$

76

where B is the decision threshold, m is the real change-time, l_i is the likelihood ratio of the ith sample, S_i^j is the likelihood ratio of the ith through jth samples as defined in Section 2.1, $z_i \triangleq \ln l_i$ is the log-likelihood-ratio of the ith sample, $Z_i \triangleq z_1 + ... + z_i$, k' is a large integer, $Q()$ is the tail function of the standard normal distribution, and μ_z and σ_z^2 are mean and variance of z_i respectively.

PROOF: According to the definition, the false alarm probability of the CUSUM with threshold B is

$$\alpha(m, B) = P_m\{(T_1 \geq B) \cup (T_2 \geq B) \cup ... \cup (T_{m-1} \geq B)\}$$

Recall the definition (2.2.1) of the CUSUM statistic

$$T_n \triangleq \max_{1 \leq i \leq n} S_i^n,$$

It is easily seen that the event $\{T_n \geq B\}$ is equivalent to $\{\bigcup_{i=1}^n (S_i^n \geq B)\}$, i.e.

$$\{T_n \geq B\} = \{\bigcup_{i=1}^n (S_i^n \geq B)\} \qquad (3.6.14)$$

Therefore

$$
\begin{aligned}
\alpha(m, B) &= P_m\{[S_1^1 \geq B] \cup [(S_1^2 \geq B) \cup (S_2^2 \geq B)] \cup \\
&\quad ... \cup [(S_1^{m-1} \geq B) \cup (S_2^{m-1} \geq B) \cup ... \cup (S_{m-1}^{m-1} \geq B)]\} \\
&= P_0\{[(S_1^1 \geq B) \cup (S_2^2 \geq B) \cup ... \cup (S_{m-1}^{m-1} \geq B)] \cup \\
&\quad [(S_1^2 \geq B) \cup (S_2^3 \geq B) \cup ... \cup (S_{m-2}^{m-1} \geq B)] \cup ... \cup [S_1^{m-1} \geq B]\}
\end{aligned}
$$

$$(3.6.15)$$

From the above equation it is easily seen that

$$\alpha(m, B) \geq P_0\{(S_1^1 \geq B) \cup (S_2^2 \geq B) \cup ... \cup (S_{m-1}^{m-1} \geq B)\} \qquad (3.6.16)$$

and

$$
\begin{aligned}
\alpha(m, B) &\leq P_0\{(S_1^1 \geq B) \cup (S_2^2 \geq B) \cup ... \cup (S_{m-1}^{m-1} \geq B)\} + \\
&\quad P_0\{S_1^2 \geq B\} + P_0\{S_2^3 \geq B\} + ... + P_0\{S_{m-2}^{m-1} \geq B\} \\
&\quad + ... + P_0\{S_1^{m-1} \geq B\}
\end{aligned}
$$

$$(3.6.17)$$

77

Since the observations $X_1, X_2, \ldots, X_{m-1}$ are i.i.d., the right-hand-side of (3.6.16) is

$$
\begin{aligned}
\alpha_L(m, B) &= P_0\{(S_1^1 \geq B) \cup (S_2^2 \geq B) \cup \ldots \cup (S_{m-1}^{m-1} \geq B)\} \\
&= P_0\{(l_1 \geq B) \cup (l_2 \geq B) \cup \ldots \cup (l_{m-1} \geq B)\} \\
&= \sum_{k=1}^{m-1} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq m-1} P_0\{l_{i_1} \geq B\} P_0\{l_{i_2} \geq B\} \ldots P_0\{l_{i_k} \geq B\} \\
&= \sum_{k=1}^{m-1} (-1)^{k+1} C_{m-1}^k P_0\{l_1 \geq B\} \\
&= 1 - [1 - P_0(l_1 \geq B)]^{m-1} \tag{3.6.18}
\end{aligned}
$$

which is the lower bound.

Using the i.i.d. assumption and the above equation yields that the right-hand-side of (3.6.17) is

$$
\begin{aligned}
\alpha_U(m, B) &= 1 - [1 - P_0(l_1 \geq B)]^{m-1} + (m-2) P_0\{S_1^2 \geq B\} + (m-3) P_0\{S_1^3 \geq B\} + \\
&\quad \ldots + P_0\{S_1^{m-1} \geq B\} \tag{3.6.19}
\end{aligned}
$$

which is the upper bound.

The approximation of $\alpha_U(m, B)$ can be easily derived using the central limit theorem.

QED

It is easily seen that when $m = 1$, the lower bound $\alpha_L(m, B)|_{m=1} = 0$; and when $m \to +\infty$, the lower bound $\alpha_L(m, B)|_{m \to \infty} = 1$. That is, the lower bound $\alpha_L(m, B)$ equals the exact FAP for $m = 1$ and $m \to +\infty$. This result is also consistent with Proposition 3.3.

In many applications, $P_0(l_1 \geq B) \ll 1$. For this case, the lower and upper bounds may be written as

$$
\alpha_L(m, B) \approx (m-1) P_0(l_1 \geq B) \tag{3.6.20}
$$

and

$$
\alpha_U(m, B) \approx \sum_{k=1}^{m-1} (m-k) P_0\{S_1^k > B\} \tag{3.6.21}
$$

78

The lower bound $\alpha_L(m, B)$ can be easily computed for a real change-time m, a given threshold B and any distributions.

The numerical computation of the upper bound $\alpha_U(m, B)$ is less straightforward. When the pdfs f_0 and f_1 are both Gaussian with a common variance σ^2 (such as a model of constant signal plus Gaussian noise), a closed-form solution of the upper and lower bounds (α_L and α_U) can be easily obtained. Let

$$z_k \triangleq \ln l_k \tag{3.6.22}$$

and

$$Z_k \triangleq \ln S_1^k = z_1 + z_2 + ... + z_k. \tag{3.6.23}$$

Let the pdfs be

$$f_1(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu_1)^2}{2\sigma^2}} \tag{3.6.24}$$

and

$$f_0(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu_0)^2}{2\sigma^2}} \tag{3.6.25}$$

Then the log-likelihood ratio of the kth sample is

$$z_k = (X_k - \frac{\mu_1 + \mu_0}{2})(\frac{\mu_1 - \mu_0}{\sigma^2}) \tag{3.6.26}$$

Under H_0, z_k is Gaussian with mean

$$E\{z_k\} = -\frac{(\mu_1 - \mu_0)^2}{2\sigma^2} \tag{3.6.27}$$

and variance

$$Var\{z_k\} = \frac{(\mu_1 - \mu_0)^2}{\sigma^2} \tag{3.6.28}$$

Therefore Z_k is Gaussian with mean

$$E\{Z_k\} = -\frac{(\mu_1 - \mu_0)^2}{2\sigma^2} k \tag{3.6.29}$$

and variance

$$Var\{Z_k\} = \frac{(\mu_1 - \mu_0)^2}{\sigma^2} k. \tag{3.6.30}$$

Then

$$P_0\{S_1^k > B\} = P_0\{Z_k \geq \ln B\} = Q(\frac{\ln B + \frac{(\mu_1 - \mu_0)^2}{2\sigma^2}k}{\frac{|\mu_1 - \mu_0|}{\sigma}\sqrt{k}}) \tag{3.6.31}$$

where $Q(y)$ is the tail of the standard normal distribution, i.e.

$$Q(y) \triangleq \int_y^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \tag{3.6.32}$$

Therefore the closed-form solutions of the lower and upper bounds are

$$\alpha_L(m, B) = 1 - [1 - Q(\frac{\ln B + \frac{(\mu_1 - \mu_0)^2}{2\sigma^2}}{\frac{|\mu_1 - \mu_0|}{\sigma}})]^{m-1} \tag{3.6.33}$$

and

$$\alpha_U(m, B) = 1 - [1 - Q(\frac{\ln B + \frac{(\mu_1 - \mu_0)^2}{2\sigma^2}}{\frac{|\mu_1 - \mu_0|}{\sigma}})]^{m-1} + \sum_{k=2}^{m-1}(m - k)Q(\frac{\ln B + \frac{(\mu_1 - \mu_0)^2}{2\sigma^2}k}{\frac{|\mu_1 - \mu_0|}{\sigma}\sqrt{k}}) \tag{3.6.34}$$

Particularly, when $Q(\frac{\ln B + \frac{(\mu_1 - \mu_0)^2}{2\sigma^2}}{\frac{|\mu_1 - \mu_0|}{\sigma}}) \ll 1$, the above bounds may be written as

$$\alpha_L(m, b) \approx (m - 1)Q(\frac{\ln B + \frac{(\mu_1 - \mu_0)^2}{2\sigma^2}}{\frac{|\mu_1 - \mu_0|}{\sigma}}) \tag{3.6.35}$$

and

$$\alpha_U(m, B) \approx \sum_{k=1}^{m-1}(m - k)Q(\frac{\ln B + \frac{(\mu_1 - \mu_0)^2}{2\sigma^2}k}{\frac{|\mu_1 - \mu_0|}{\sigma}\sqrt{k}}) \tag{3.6.36}$$

Some numerical results are shown in Table 3.3. It is seen that the above bounds are reasonably tight if the signal to noise ratio is large. However, when the SNR is small, the bounds are too loose.

In any case, it appears that taking the average of the lower and upper bounds

$$\overline{\alpha}(m, B) = \frac{\alpha_L(m, B) + \alpha_U(m, B)}{2}$$

improves the estimated false alarm probability.

The closed-form solution (3.6.33)-(3.6.36) holds only for Gaussian f_0 and f_1 with common variance. For other distributions, computing the upper bound is less straightforward. The upper bound $\alpha_U(m, B)$ in Proposition 3.5 is equivalent to

$$\begin{aligned}
\alpha_U(m, B) &= 1 - [1 - P_0(z_1 \geq \ln B)]^{m-1} + \sum_{k=2}^{m-1}(m - k)P_0\{z_1 + z_2 + \ldots + z_k > \ln B\} \\
&= 1 - [1 - P_0(z_1 \geq \ln B)]^{m-1} + \sum_{k=2}^{m-1}(m - k)P_0\{Z_k > \ln B\} \tag{3.6.37}
\end{aligned}$$

80

Table 3.3: *Lower and upper bounds of the false alarm probability for CUSUM. f_0 is Gaussian with mean 0, and variance 0.25^2, f_1 is Gaussian with mean μ_1 and the same variance, and $SNR \triangleq 20 \log \frac{\mu_1}{\sigma} (db)$. The simulated $\alpha(m, B)$ is also listed.*

μ_1	SNR	m	$\alpha \pm 5\alpha$	α_L	α_U	$\ln B$
.79	10.	100	$.099 \pm .002$	.077	.134	5.
		300	$.092 \pm .002$	.059	.126	6.2
		800	$.110 \pm .002$	.059	.151	7.0
.79	10.	500	$.073 \pm .002$	.039	.098	6.96
			$.0181 \pm .001$	.0060	.0238	8.352
			$.00334 \pm .0004$	.00051	.00439	10.02
.499	6.0		$.112 \pm .002$	.0028	.2857	6.768
.420	4.5	500	$.090 \pm .002$	.00015	.325	6.98
.25	0.0		$.10 \pm .002$	2.87×10^{-10}	1.29	6.611

where z_i is the log-likelihood ratio of the ith sample. For any distributions f_0 and f_1, the pdf of z_i can be easily derived. Due to the i.i.d. assumption of X_i, it is easily seen that z_i are i.i.d. and the pdf of Z_k is the convolution of the pdfs of $z_1, z_2, ..., z_k$, i.e.

$$f_{Z_k}(t) = f_{z_1}(t) * f_{z_2}(t) * ... * f_{z_k}(t) = \underbrace{f_z(t) * f_z(t) * ... * f_z(t)}_{k} \qquad (3.6.38)$$

where the operator '$*$' stands for 'convolution' and f is the pdf of the corresponding variable. When k is small, the above convolution can be numerically computed. However, when k is large, it is not practical to compute the convolution numerically. Therefore an alternative approach has to be found. Fortunately, applying the central limit theorem to $Z_k = z_1 + z_2 + ... + z_k$ implies that for $k \geq k'$, Z_k is approximately Gaussian with mean

$$\mu_k = E_0\{Z_k\} = kE_0\{z_k\} = k\mu_z \qquad (3.6.39)$$

81

Table 3.4: *Lower and upper bounds of the false alarm probability for CUSUM. f_0 and f_1 are Rayleigh densities with variances σ_0^2 and σ_1^2 respectively. The significance of the change is $s = \frac{\sigma_1^2}{\sigma_0^2}$. The simulated $\alpha(m, B)$ is also listed.*

σ_1	σ_0	s	m	$\alpha \pm \delta\alpha$	α_L	α_U	$\ln B$
.45		20.25	100	$.00992 \pm .0007$	.00759	.01498	6.
.4	.1	16	100	$.0320 \pm .0013$	.0248	.0540	5.
			500	$.0215 \pm .0011$	.0148	.0371	7.0
			1000	$.0153 \pm .0009$	.0102	.0273	8
.3		9.	100	$.0065 \pm .0006$	.0031	.0152	7.
.25		6.25	100	$.0203 \pm .0010$	.0088	.0693	6.
.2		4.	100	$.0631 \pm .0018$	.0198	.3876	5.

and variance

$$\sigma_k = Var_0\{Z_k\} = kVar_0\{z_k\} = k\sigma_z^2 \qquad (3.6.40)$$

where $k'(< m-1)$ is a reasonably large integer and depends on the rate of convergence of the central limit property. Therefore the upper bound can be written as

$$\alpha_U(m, B) \approx 1 - [1 - P_0(z_1 \geq \ln B)]^{m-1} + \sum_{i=2}^{k'-1}(m-i)\int_{\ln B}^{\infty} f_{Z_i}(t)dt + \sum_{k=k'}^{m-1}(m-k)Q(\frac{\ln B - k\mu_z}{\sqrt{k}\sigma_z}) \qquad (3.6.41)$$

where $Q()$ is the tail of the standard normal distribution.

As an example, let the pdfs be

$$f_0(x) = \frac{x}{\sigma_0^2}e^{-\frac{x^2}{2\sigma_0^2}} \qquad x \geq 0 \qquad (3.6.42)$$

and

$$f_1(x) = \frac{x}{\sigma_1^2}e^{-\frac{x^2}{2\sigma_1^2}} \qquad x \geq 0 \qquad (3.6.43)$$

which are Rayleigh densities [98](pp.29-30), and may represent the amplitude distribution of a Rayleigh fading channel. The quantity σ_i represents the noise power and the

82

average depth of the fade. Then the one-sample log-likelihood-ratio is

$$z_i = \ln \frac{f_1(x)}{f_0(x)} = \ln \frac{\sigma_0^2}{\sigma_1^2} + \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right) \frac{x^2}{2} \tag{3.6.44}$$

Assume that

$$\sigma_0 < \sigma_1 \tag{3.6.45}$$

i.e. that a change occurs from a lighter average fade to a deeper average fade. Then the pdf of z under H_0 is

$$f_{z_i}(z) = \frac{1}{\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right)} e^{-\left(z - \ln \frac{\sigma_0^2}{\sigma_1^2}\right)/\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right)} \qquad z > \ln \frac{\sigma_0^2}{\sigma_1^2} \tag{3.6.46}$$

where

$$s \triangleq \frac{\sigma_1^2}{\sigma_0^2} \tag{3.6.47}$$

represents the significance of the change. Using this notation and the fact $f_{z_i}(z) = 0$ for $z \leq -\ln s$, the pdf of z can be also written as

$$f_{z_i}(z) = \begin{cases} \frac{1}{(1-1/s)} e^{-(z+\ln s)/(1-1/s)} & z > -\ln s \\ 0 & z \leq -\ln s \end{cases} \tag{3.6.48}$$

where $s > 1$. Using the properties of the Fourier Transform, it can be easily shown that the pdf of Z_k is

$$f_{Z_k}(z) = \begin{cases} \frac{1}{(k-1)! \cdot (1-1/s)^k} (z + k \ln s)^{k-1} e^{-(z+k\ln s)/(1-1/s)} & z > -k \ln s \\ 0 & z \leq -k \ln s \end{cases} \tag{3.6.49}$$

for $1 \leq k \leq m - 1$.

The mean of z_i is

$$\mu_z = E_0\{z_i\} = 1 - \frac{1}{s} - \ln s \tag{3.6.50}$$

and its variance is

$$\sigma_z^2 = \left(1 - \frac{1}{s}\right)^2 \tag{3.6.51}$$

The upper and lower bounds are computed numerically. Several results are listed in Table 3.4. Again it is found that the bounds are quite tight if the change is significant.

83

3.6.2 False Alarm Probability of MAR Procedure

I. Almost Sure Termination of MAR

As a counterpart of Propositions 3.3 and 3.4 in the last subsection, we now present the propositions showing the a.s. termination of the MAR procedure.

PROPOSITION 3.6 *If a change never occurs, i.e.*

$$m = +\infty, \tag{3.6.52}$$

then the false alarm probability of the new MAR procedure equals 1, i.e.

$$\alpha = P_0\{N < +\infty\} = P_\infty\{N < +\infty\} = 1. \tag{3.6.53}$$

PROOF: Recall that the stopping rule of the MAR procedure is $\hat{r}_n < \hat{r}_{n+1}$. Therefore its false alarm probability when $m = +\infty$ is

$$\alpha = 1 - P_0\{\bigcap_{k=1}^{\infty} \hat{r}_n \geq \hat{r}_{n+1}\} \tag{3.6.54}$$

Using the definition (2.3.1), it is easily shown that the event $\{\hat{r}_n \geq \hat{r}_{n+1}\}$ is equivalent to

$$\{T_n \geq T_{n+1} + \frac{c_0(T_n\pi_0 + 1 - \pi_0)(T_{n+1}\pi_0 + 1 - \pi_0)}{\pi_0(1 - c_0)(1 - \pi_0)n}\} \tag{3.6.55}$$

where T_n is the CUSUM statistic. Using the conditions $0 < c_0 < 1$ and $0 < \pi_0 < 1$, it is easily shown that

$$T_{n+1} + \frac{c_0(T_n\pi_0 + 1 - \pi_0)(T_{n+1}\pi_0 + 1 - \pi_0)}{\pi_0(1 - c_0)(1 - \pi_0)n} > T_{n+1} \tag{3.6.56}$$

therefore

$$\{\hat{r}_n \geq \hat{r}_{n+1}\} \subset \{T_n \geq T_{n+1}\} \tag{3.6.57}$$

That is

$$\alpha \geq 1 - P_0\{\bigcap_{k=1}^{\infty} T_n \geq T_{n+1}\} \tag{3.6.58}$$

84

In addition,

$$T_n \geq T_{n+1} = \max\{T_n, 1\} l_n \geq T_n l_n \qquad (3.6.59)$$

where l_n is the likelihood ratio of the nth sample. Then

$$\{T_n \geq T_{n+1}\} \subset \{T_n \geq T_n l_n\} = \{l_n \leq 1\} \qquad (3.6.60)$$

Finally we have

$$
\begin{aligned}
\alpha &\geq 1 - P_0\{\bigcap_{k=1}^{\infty} T_n \geq T_{n+1}\} \\
&\geq 1 - P_0\{\bigcap_{k=1}^{\infty} l_n \leq 1\} \\
&= 1 - \prod_{k=1}^{\infty} P_0\{l_k \leq 1\} \\
&= 1 - \lim_{n \to \infty} [P_0(l_1 \leq 1)]^n \\
&= 1 \qquad (3.6.61)
\end{aligned}
$$

where the i.i.d. assumption of X_i is used. **QED**

The following proposition is a counterpart of Proposition 3.4 for CUSUM.

PROPOSITION 3.7 *For any change-time $m \geq 1$, the MAR procedure terminates with probability 1, i.e.*

$$P_m\{N < +\infty\} = 1. \qquad (3.6.62)$$

PROOF: There are two cases: (i) $m = +\infty$, and (ii) $m < +\infty$.

For $m = +\infty$, the proposition is actually a re-statement of Proposition 3.6 and has been proved.

For $m < +\infty$, using the result of Proposition 3.1 yields that

$$\hat{r}_n \sim \frac{e^{(n-m+1)I_f}\pi c - (1-\pi)}{e^{(n-m+1)I_f}\pi + 1 - \pi} n \qquad (3.6.63)$$

when $n \to +\infty$, where I_f is defined in Section 3.3. Using Proposition 3.2 yields that $\hat{r}_n$ is increasing with n for sufficiently large n and finite m. Therefore the stopping rule $\hat{r}_n < \hat{r}_{n+1}$ will be satisfied with probability 1 as n is sufficiently large. **QED**

85

II. False Alarm Probability Bound of MAR

We now present an upper bound for the false alarm probability of the MAR procedure. Numerical implementation of this upper bound is discussed later.

PROPOSITION 3.8 *The false alarm probability* $\alpha(m, B)$ *of the MAR procedure satisfies the following inequalities*

$$\alpha(m, c_0) \leq \sum_{n=1}^{m-2} P_0\{l_{n+1} > \frac{1}{(n+1)c_0} + \frac{n}{n+1}\} + \sum_{n=1}^{m-2} P_0\{U_n^n > \frac{1}{nc_0}\} +$$

$$\sum_{n=2}^{m-2} P_0\{U_{n-1}^n > \frac{1}{nc_0}\} + \ldots + \sum_{n=n'}^{m-2} P_0\{U_{n-n'+1}^n > \frac{1}{nc_0}\} + \ldots +$$

$$\sum_{n=m-2}^{m-2} P_0\{U_1^n > \frac{1}{nc_0}\} \triangleq \alpha_U(m, c_0) \tag{3.6.64}$$

with

$$U_k^n = [\frac{n+1}{n}l_{n+1} - 1]S_k^n \tag{3.6.65}$$

where $0 < c_0 = \pi_0 \ll 1$ *is the design parameter,* m *is the real change-time,* l_i *is the likelihood ratio of the ith sample and* S_i^j *is the likelihood ratio of the ith through jth samples as defined in Section 2.1.*

PROOF: When $0 < c_0 = \pi_0 \ll 1$, the MAR statistic $\hat{r}_n$ may be written as

$$\hat{r}_n = \frac{T_n c_0^2 - (1 - c_0)}{T_n c_0 + 1 - c_0} n \sim -n + T_n n c_0 \tag{3.6.66}$$

where the Taylor expansion is used. Therefore

$$\begin{aligned}\alpha(m, c_0) &= P_0\{\bigcup_{n=1}^{m-1} (\hat{r}_n < \hat{r}_{n+1})\} \\ &\sim P_0\{\bigcup_{n=1}^{m-1} (\frac{n+1}{n}T_{n+1} - T_n > \frac{1}{nc_0})\}\end{aligned} \tag{3.6.67}$$

Since $T_{n+1} = \max_{1 \leq i \leq n+1} S_i^{n+1}$, it is easily seen that

$$\{\frac{n+1}{n}T_{n+1} - T_n > \frac{1}{nc_0}\} = \{\bigcup_{k=1}^{n+1} (\frac{n+1}{n}S_k^{n+1} - T_n > \frac{1}{nc_0})\} \tag{3.6.68}$$

86

For any $1 \le k \le n$,

$$T_n = \max_{1 \le k \le n} S_k^n \ge S_k^n \qquad (3.6.69)$$

therefore

$$\{\frac{n+1}{n}T_{n+1} - T_n > \frac{1}{nc_0}\} = \{\bigcup_{k=1}^{n+1}(\frac{n+1}{n}S_k^{n+1} - T_n > \frac{1}{nc_0})\}$$

$$\subset \{\bigcup_{k=1}^{n+1}(\frac{n+1}{n}S_k^{n+1} - S_k^n > \frac{1}{nc_0})\} = \{\bigcup_{k=1}^{n+1}(U_k^n > \frac{1}{nc_0})\} \qquad (3.6.70)$$

Substituting the above equation into (3.6.68) yields the upper bound. **QED**

The central part of computing the upper bound (3.6.64) is to calculate the density of U_k^n, or that of

$$V_{kn} \triangleq \ln U_k^n = \ln(\frac{n+1}{n}l_{n+1} - 1) + \sum_{i=k}^{n} z_i = y_n + Z_k^n \qquad (3.6.71)$$

under the condition that

$$C_n \triangleq \{\frac{n+1}{n}l_{n+1} - 1 > 0\} = \{z_{n+1} \ge \ln\frac{n}{n+1}\} \qquad (3.6.72)$$

where

$$y_n \triangleq \ln(\frac{n+1}{n}l_{n+1} - 1) \qquad (3.6.73)$$

and

$$Z_k^n \triangleq \sum_{i=k}^{n} z_i \qquad (3.6.74)$$

with z_i to be the log-likelihood-ratio of the ith sample. It can be easily shown that the density of V_{kn} conditioned by $C_n = \{\frac{n+1}{n}l_{n+1} - 1 > 0\}$ is

$$f_{kn|C_n}(t) = f_{y_n|C_n}(t) * f_{z_k}(t) * ... * f_{z_n}(t) \qquad (3.6.75)$$

where

$$f_{y_n|C_n}(t) \triangleq \frac{\partial}{\partial t}P_0(y_n \le t|C_n) \qquad (3.6.76)$$

For small $n - k$, such as $n - k \le n'$, the above convolution may be evaluated numerically. When $n - k$ is large, using the central limit theorem yields that

$$f_{kn|C_n}(t) \approx \phi(t|(n - k + 1)\mu_z + \mu_{y_n}, \sqrt{(n - k + 1)\sigma_z^2 + \sigma_{y_n}^2}) \qquad (3.6.77)$$

87

where $\phi(t|\mu,\sigma)$ is the Gaussian density with mean μ and variance σ^2. The notations μ_{y_n} and $\sigma_{y_n}^2$ are mean and variance of y_n conditioned by C_n, and μ_z and σ_z^2 are mean and variance of the one-sample log-likelihood-ratio z_i. Therefore the upper bound may be written as

$$\alpha_U(m,c_0) \approx \sum_{n=1}^{m-2}\int_{\ln\frac{1}{(n+1)c_0}}^{\infty} f_z(t)dt + \sum_{n=1}^{m-2}\int_{\ln\frac{1}{nc_0}}^{\infty} P_0(C_n)f_{nn|C_n}(t)dt +$$

$$\sum_{n=2}^{m-2}\int_{\ln\frac{1}{nc_0}}^{\infty} P_0(C_n)f_{(n-1)n|C_n}(t)dt + ... + \sum_{n=n'-1}^{m-2}\int_{\ln\frac{1}{nc_0}}^{\infty} P_0(C_n)f_{[n-(n'-2)]n|C_n}(t)dt +$$

$$\sum_{n=n'}^{m-2} P_0(C_n)Q(\frac{\ln[1/(nc_0)] - n'\mu_z - \mu_{y_n}}{\sqrt{n'\sigma_z^2 + \sigma_{y_n}^2}}) + ...$$

$$+ \sum_{n=m-2}^{m-2} P_0(C_n)Q(\frac{\ln[1/(nc_0)] - (m-2)\mu_z - \mu_{y_n}}{\sqrt{(m-2)\sigma_z^2 + \sigma_{y_n}^2}}) \tag{3.6.78}$$

Define

$$f_{kn,C_n}(t) \triangleq P_0(C_n)f_{kn|C_n}(t) = f_{y_n,C_n}(t) * f_{z_k}(t) * ... * f_{z_n}(t) \tag{3.6.79}$$

where

$$f_{y_n,C_n}(t) \triangleq \frac{\partial}{\partial t}P_0(y_n \leq t, C_n)$$

$$= \frac{\partial}{\partial t}P_0\{\ln\frac{n}{n+1} \leq z_{n+1} \leq \ln\frac{n}{n+1} + \ln(e^t+1)\}$$

$$= \frac{e^t}{e^t+1}f_{z_{n+1}}(\ln\frac{n}{n+1} + \ln(e^t+1)) \tag{3.6.80}$$

Notice that because of the i.i.d. assumption, $f_{z_i}() = f_{z_j}() = f_z()$ for any $i,j < m$. Then

$$f_{kn,C_n}(t) = [\frac{e^t}{e^t+1}f_z(\ln\frac{n}{n+1} + \ln(e^t+1))] * \underbrace{f_z(t) * ... * f_z(t)}_{n-k+1} \triangleq f_{(n-k+1)n}(t) \tag{3.6.81}$$

Therefore the upper bound may be written as

$$\alpha_U(m,c_0) \approx \sum_{n=1}^{m-2}\int_{\ln\frac{1}{(n+1)c_0}}^{\infty} f_z(t)dt + \sum_{n=1}^{m-2}\int_{\ln\frac{1}{nc_0}}^{\infty} f_{1n}(t)dt +$$

$$\sum_{n=2}^{m-2}\int_{\ln\frac{1}{nc_0}}^{\infty} f_{2n}(t)dt + ... + \sum_{n=n'-1}^{m-2}\int_{\ln\frac{1}{nc_0}}^{\infty} f_{(n'-1)n}(t)dt +$$

$$\sum_{n=n'}^{m-2} P_0(C_n)Q(\frac{\ln[1/(nc_0)] - n'\mu_z - \mu_{y_n}}{\sqrt{n'\sigma_z^2 + \sigma_{y_n}^2}}) + ...$$

$$+ \sum_{n=m-2}^{m-2} P_0(C_n)Q(\frac{\ln[1/(nc_0)] - (m-2)\mu_z - \mu_{y_n}}{\sqrt{(m-2)\sigma_z^2 + \sigma_{y_n}^2}}) \tag{3.6.82}$$

88

with

$$P_0(C_n) = P_0(z_{n+1} > \ln \frac{n}{n+1}) = \int_{\ln \frac{n}{n+1}}^{\infty} f_z(t)dt \qquad (3.6.83)$$

and

$$f_{kn}(t) = [\frac{e^t}{e^t+1} f_z(\ln \frac{n}{n+1} + \ln(e^t+1))] * \underbrace{f_z(t) * ... * f_z(t)}_{k} \qquad (3.6.84)$$

The computational complexity of the convolution-integration operation in (3.6.82) is still quite large, particularly when m is large. Notice that for a large number n'', when $n \geq n''$

$$f_{kn}(t) \approx f_{k(n-1)}(t)$$

Therefore

$$\sum_{n=k}^{m-2} \int_{\ln \frac{1}{nc_0}}^{\infty} f_{kn}(t)dt \approx \sum_{n=k}^{n''-1} \int_{\ln \frac{1}{nc_0}}^{\infty} f_{kn}(t)dt + (m-1-n'') \int_{\ln \frac{1}{n''c_0}}^{\infty} f_{kn''}(t)dt +$$

$$(m-2-n'') \int_{\ln \frac{1}{(n''+1)c_0}}^{\ln \frac{1}{n''c_0}} f_{k(n''+1)}(t)dt +$$

$$(m-3-n'') \int_{\ln \frac{1}{(n''+2)c_0}}^{\ln \frac{1}{(n''+1)c_0}} f_{k(n''+2)}(t)dt +$$

$$... + \int_{\ln \frac{1}{(m-2)c_0}}^{\ln \frac{1}{(m-3)c_0}} f_{k(m-2)}(t)dt \qquad (3.6.85)$$

Eqs. (3.6.82)-(3.6.85) may be used to estimate the upper bound of false alarm probability of the MAR procedure.

As an example, consider the change-detection model (3.6.24) (3.6.25), i.e. the detection of a change in the mean of Gaussian distribution. The mean and variance of z_k under H_0 are

$$\mu_z = E\{z_k\} = -\frac{(\mu_1 - \mu_0)^2}{2\sigma^2} \qquad (3.6.86)$$

and

$$\sigma_z^2 = Var\{z_k\} = \frac{(\mu_1 - \mu_0)^2}{\sigma^2} \qquad (3.6.87)$$

respectively. It is easily seen that

$$f_{kn}(t) = [\frac{e^t}{e^t+1} \phi(\ln \frac{n}{n+1} + \ln(e^t+1)|\mu_z, \sigma_z)] * \underbrace{\phi(t|\mu_z, \sigma_z) * ... * \phi(t|\mu_z, \sigma_z)}_{k}$$

89

$$= [\frac{e^t}{e^t+1}\phi(\ln\frac{n}{n+1} + \ln(e^t+1)|\mu_z, \sigma_z)] * \phi(t|k\mu_z, \sqrt{k}\sigma_z) \qquad (3.6.88)$$

and

$$P_0(C_n) = Q(\frac{\ln\frac{n}{n+1} - \mu_z}{\sigma_z}) \qquad (3.6.89)$$

where $\phi(t|\mu, \sigma)$ is the Gaussian density with mean μ and variance σ^2, and $Q()$ is the tail area of the standard Gaussian distribution. The density of y_n conditioned by C_n is

$$
\begin{aligned}
f_{y_n|C_n}(t) &= \frac{f_{y_n, C_n}(t)}{P_0(C_n)} \\
&= \frac{e^t}{e^t+1} f_{z_{n+1}}(\ln\frac{n}{n+1} + \ln(e^t+1))/P_0(C_n) \\
&= \frac{e^t}{e^t+1}\phi(\ln\frac{n}{n+1} + \ln(e^t+1)|\mu_z, \sigma_z)/Q(\frac{\ln\frac{n}{n+1} - \mu_z}{\sigma_z}) \quad (3.6.90)
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\mu_{y_n} &= \int_{-\infty}^{\infty} t f_{y_n|C_n}(t)dt \\
&= \int_{-\infty}^{\infty} t\frac{e^t}{e^t+1}\phi(\ln\frac{n}{n+1} + \ln(e^t+1)|\mu_z, \sigma_z)dt/Q(\frac{\ln\frac{n}{n+1} - \mu_z}{\sigma_z}) \\
&\approx \int_{-\infty}^{\infty} t\frac{e^t}{e^t+1}\phi(\ln\frac{n''}{n''+1} + \ln(e^t+1)|\mu_z, \sigma_z)dt/Q(\frac{\ln\frac{n''}{n''+1} - \mu_z}{\sigma_z}) \quad (3.6.91)
\end{aligned}
$$

and

$$\sigma_{y_n}^2 \approx \int_{-\infty}^{\infty} t^2\frac{e^t}{e^t+1}\phi(\ln\frac{n''}{n''+1} + \ln(e^t+1)|\mu_z, \sigma_z)dt/Q(\frac{\ln\frac{n''}{n''+1} - \mu_z}{\sigma_z}) \quad (n \geq n') \qquad (3.6.92)$$

Substituting the above quantities into Eqs. (3.6.82)-(3.6.82), a number of numerical values of the upper bound α_U are computed and listed in Table 3.5. It is observed that for high SNRs the bound is reasonably tight. However, if the SNR is low the bound is quite loose.

Now we compute the bound for the Rayleigh model given by Eqs. (3.6.42)(3.6.43). Using the results in Section 3.6.1 yields that the pdf of the one-sample log-likelihood ratio z is

$$f_z(z) = \begin{cases} \frac{1}{(1-1/s)}e^{-(z+\ln s)/(1-1/s)} & z > -\ln s \\ 0 & z \leq -\ln s \end{cases} \qquad (3.6.93)$$

90

Table 3.5: *An upper bound α_U of the false alarm probability for MAR. f_0 is Gaussian with mean 0, and variance 0.25^2, f_1 is Gaussian with mean μ_1 and the same variance, and $SNR \triangleq 20 \log \frac{\mu_1}{\sigma} (db)$. The simulated $\alpha(m, B)$ is also listed.*

μ_1	SNR	m	$\alpha \pm \delta\alpha$	α_U	$\pi_0 = c_0$
.79	10.	100	$.114 \pm .002$	.149	.00017
		300	$.090 \pm .002$	.117	1.4×10^{-5}
		800	$.092 \pm .002$	.117	1.98×10^{-6}
.79	10.	500	$.101 \pm .002$	.130	5.6×10^{-6}
			$.0086 \pm .0007$	.0112	$5. \times 10^{-7}$
			$.0010 \pm .0002$	.0011	$5. \times 10^{-8}$
.499	6.0		$.095 \pm .002$	.199	4.62×10^{-6}
.420	4.5	500	$.117 \pm .002$	.321	$6. \times 10^{-6}$
.25	0.0		$.098 \pm .002$	.736	7.84×10^{-6}

where

$$s = \frac{\sigma_1^2}{\sigma_0^2} \tag{3.6.94}$$

represents the significance of the change. The pdf of $Z_k = z_1 + z_2 + ... + z_k$ is

$$f_{Z_k}(z) = \underbrace{f_z(t) * ... * f_z(t)}_{k}$$
$$= \begin{cases} \frac{1}{(k-1)! \cdot (1-1/s)^k}(z + k\ln s)^{k-1} e^{-(z+k\ln s)/(1-1/s)} & z > -k\ln s \\ 0 & z \le -k\ln s \end{cases} \tag{3.6.95}$$

for $1 \le k \le m - 1$.

It is easily computed that the mean and variance of z are

$$\mu_z = E_0\{z\} = 1 - \frac{1}{s} - \ln s \tag{3.6.96}$$

and

$$\sigma_z^2 = (1 - \frac{1}{s})^2 \tag{3.6.97}$$

Table 3.6: *An upper bound of the false alarm probability for MAR. f_0 and f_1 are Rayleigh densities with variances σ_0^2 and σ_1^2 respectively. The significance of the change is $s = \frac{\sigma_1^2}{\sigma_0^2}$. The simulated $\alpha(m, c_0)$ is also listed.*

σ_1	σ_0	s	m	$\alpha \pm \delta\alpha$	α_U	$\pi_0 = c_0$
.45		20.25	100	$.01032 \pm .0007$	.01369	5.5×10^{-5}
.4	.1	16	100	$.0223 \pm .0011$	.0317	.0001
			500	$.0215 \pm .0011$	.0399	$5. \times 10^{-6}$
			1000	$.0105 \pm .0008$	.0160	$5. \times 10^{-7}$
.3		9.	100	$.0047 \pm .0005$	.0087	1.5×10^{-5}
.25		6.25	100	$.0187 \pm .0010$	.0434	$5. \times 10^{-5}$
.2		4.	100	$.0645 \pm .0018$	.2338	.00016

respectively. It can be shown that

$$f_{kn}(t) = [(\frac{n+1}{ns})^{s/(s-1)} \frac{s}{(s-1)} \frac{e^t}{(e^t+1)^{(2s-1)/(s-1)}}] * f_{Z_k}(t) \quad s \geq 2 \qquad (3.6.98)$$

When $s < 2$, a slightly different form of f_{kn} can be derived. This situation is not considered here. Similarly

$$P_0(C_n) = (\frac{n+1}{ns})^{s/(s-1)} \qquad (3.6.99)$$

and

$$f_{y_n|C_n}(t) = \frac{f_{y_n,C_n}(t)}{P_0(C_n)} = \frac{s}{(s-1)} \frac{e^t}{(e^t+1)^{(2s-1)/(s-1)}} \qquad (3.6.100)$$

The mean and variance of y_n conditioned by C_n are

$$\mu_{y_n} = \int_{-\infty}^{\infty} t \cdot \frac{s}{(s-1)} \frac{e^t}{(e^t+1)^{(2s-1)/(s-1)}} dt \qquad (3.6.101)$$

and

$$\sigma_{y_n}^2 = \int_{-\infty}^{\infty} t^2 \cdot \frac{s}{(s-1)} \frac{e^t}{(e^t+1)^{(2s-1)/(s-1)}} dt \qquad (3.6.102)$$

respectively.

92

Combining Eqs. (3.6.82)-(3.6.85) with the above quantities, we computed a number of upper bounds α_u of the FAP for the Rayleigh model. The results are listed in Table 3.6. It is again observed that the bound is quite tight for large SNRs and loose for small SNRs.

3.6.3   Almost Sure Termination of GRS Procedure

In this subsection, we present two propositions showing almost sure termination of the GRS procedure. As is pointed out in Subsection 2.2.3, it appears difficult to compute an upper bound of the FAP for the GRS procedure. Therefore this problem will not be studied in this thesis.

PROPOSITION 3.9 *If a change never occurs, i.e.*

$$m = +\infty, \tag{3.6.103}$$

then the false alarm probability of the GRS procedure equals 1, i.e.

$$\alpha = P_0\{N < +\infty\} = P_\infty\{N < +\infty\} = 1. \tag{3.6.104}$$

PROOF: It is easily seen that

$$\alpha = 1 - P_0\{\bigcap_{n=1}^{\infty} R_{pn} < A\} \tag{3.6.105}$$

where

$$R_{pn} = \sum_{k=1}^{n} \prod_{i=k}^{n} \frac{l_i}{1-p} \geq \frac{l_n}{1-p} \tag{3.6.106}$$

Therefore

$$
\begin{aligned}
\alpha &\geq 1 - P_0\{\bigcap_{n=1}^{\infty} l_n < A(1-p)\} \\
&= 1 - \lim_{n \to \infty} \{P_0[l_1 < A(1-p)]\}^n \\
&= 1 \tag{3.6.107}
\end{aligned}
$$

QED

93

The following proposition shows the almost sure termination of the GRS procedure for any $m \geq 1$.

PROPOSITION 3.10 *For any change-time $m \geq 1$, the MAR procedure terminates with probability 1, i.e.*

$$P_m\{N < +\infty\} = 1. \tag{3.6.108}$$

PROOF: There are two cases: (i) $m = +\infty$, and (ii) $m < +\infty$.

For $m = +\infty$, the proposition is actually a re-statement of Proposition 3.9 and has been proved.

If $m < +\infty$,

$$
\begin{aligned}
\alpha &= 1 - P\{(\bigcap_{i=1}^{m-1} R_{pi} < A)(\bigcap_{n=m}^{\infty} R_{pn} < A)\} \\
&\geq 1 - P\{\bigcap_{n=m}^{\infty} R_{pn} < A\} \\
&\geq 1 - P\{\bigcap_{n=m}^{\infty} l_n < A(1-p)\} \\
&= 1 - \lim_{n \to \infty}\{P_1[l_1 < A(1-p)]\}^{n-m+1} \\
&= 1
\end{aligned}
\tag{3.6.109}
$$

which shows the proposition.

QED

3.7 Summary

In this chapter, a new procedure, referred to as the minimum asymptotic risk (MAR) procedure, is proposed. It is shown that the MAR procedure converges almost surely to an optimum procedure under certain asymptotic conditions. Extensive simulations reveal that the MAR procedure outperforms existing procedures for nonasymptotic situations. The false alarm probability bounds of change detection procedures are also studied.

Chapter 4

GENERALIZATION OF CHANGE-DETECTION PROCEDURES

In Chapter 3, we studied the quickest detection of an abrupt change in a random sequence with independent and identical distributions (i.i.d.) before and after the change-time m. In addition, the distributions before and after the change were assumed to be completely known. Based upon these assumptions, a change-detection procedure with minimal complexity is developed. In some practical situations the assumptions are satisfied or approximately satisfied, and complexity is an important enough issue to prohibit considering more complicated algorithms. However, in a number of applications an unacceptable performance loss would result if the above assumptions are not satisfied. In this chapter, we consider the change detection problem for situations where the distributions are known but have unknown parameters, and for non-independent or non-identically distributed (non-i.i.d.) observations. The approach is to generalize procedures studied in Chapters 2 and 3, including CUSUM, GRS, and the new MAR procedure. We will also study some key open problems concerning the well-known sequential probability ratio test (SPRT) for non i.i.d. observations, which has application to the change detection problem for non-i.i.d. observations.

95

4.1 Quickest Detection of Change in a Random Sequence with Unknown Parameters

In this section, the procedure presented in Chapter 3 is modified to solve the change detection problem for i.i.d. observations with unknown distribution parameters. This modification involves two multiple hypothesis (MH) approaches and a moving-window maximum likelihood estimation (MLE) approach. Their properties are first investigated theoretically. Their non-asymptotic performance is then studied via simulations.

4.1.1 Problem Statement

In this chapter, we study the signal change detection problem for a practical case where the distribution function $F_0(x_i)$ and the density $f_0(x_i)$ of the observations before the change time m (i.e. $i < m$) are completely known while the distribution function $F_1(x_i, \theta)$ and the density $f_1(x_i, \theta)$ of the observations after the change time (i.e. $i \geq m$) have unknown parameter (possibly a vector) θ. The observations are assumed to be i.i.d. before and after the change time. This problem is practical since estimating the distribution parameters of f_0 before change-time is relatively straightforward, while estimating the distribution parameters θ of f_1 is traded off by the quickest detection objective and should therefore be studied carefully.

Similar to Chapter 2, we define joint distributions P_m as

$$
P_m\{X_i^n \leq x_i^n\} = \begin{cases} \int_{-\infty}^{x_i^n} f_0(x_i)f_0(x_{i+1})...f_0(x_n)dx_i^n & \text{if } n < m \\ \int_{-\infty}^{x_i^n} f_0(x_i)...f_0(x_{m-1})f_1(x_m,\theta)...f_1(x_n,\theta)dx_i^n & \text{if } i < m \leq n \\ \int_{-\infty}^{x_i^n} f_1(x_i,\theta)f_1(x_{i+1},\theta)...f_1(x_n,\theta)dx_i^n & \text{if } m \leq i \leq n \end{cases}
$$

$$(4.1.1)$$

and expectation E_m under P_m for $m \geq 1$. Let $P_0 \triangleq P_\infty$. Define one-sample likelihood

96

ratio

$$l_i(\theta) \triangleq \frac{f_1(x_i, \theta)}{f_0(x_i)} \tag{4.1.2}$$

and its product

$$S_k^n(\theta) \triangleq \begin{cases} \prod_{i=k}^n l_i(\theta) & \text{if } n \geq k \\ 1 & \text{if } n < k. \end{cases} \tag{4.1.3}$$

Let $H_m(\theta)$ denote the hypothesis that P_m is true. For any stopping variable $N = n$, define

$$\Theta_n \triangleq \{X_n \sim f_1(x_n, \theta)\} \tag{4.1.4}$$

and

$$\overline{\Theta}_n \triangleq \{X_n \sim f_0\}. \tag{4.1.5}$$

This definition of Θ_n implies a correct detection, i.e., that one stops at or after the actual change-time m ($n \geq m$), $\overline{\Theta}_n$ implies a false alarm, i.e., that one stops before m ($n < m$). The modified Shiryayev criterion introduced in Chapter 3 will be used, i.e., that our objective is to minimize the expected delay D_m subject to a given level of false alarm probability α and false alarm ARL N_f, where D_m and N_f are defined the same as in Chapter 2.

4.1.2 General Background for Detection of Signal with Unknown Parameters

Detection of signals with unknown parameters has been studied for binary and M-ary hypothesis testing problems [97]. However, this problem has not been solved satisfactorily for signal change detection, even though some ad hoc approaches have been proposed. Generally, the approaches to solving the unknown parameter problem may be classified into: (i) *multiple hypothesis (MH)* approaches, and (ii) *maximum likelihood estimation (MLE)* approaches.

I. Multiple Hypothesis Approach

One of the first MH schemes was derived by Magill [78] for an estimation-filtering problem. This method involves assuming that the unknown parameters may take on one of M values, corresponding to M different hypotheses. Such a formulation results in a standard M-ary detection (or estimation) problem with known parameters. The actual parameter may not be exactly the same as any one of M hypotheses. A survey of the MH approach for the estimation-filtering problem is found in [45].

For signal change detection, the well-known two-sided CUSUM (or V-mask) procedure [42] for detecting the mean shift of Gaussian distribution may be viewed as an ad hoc MH approach for this problem, where $M = 2$. This technique is described as follows: The signal amplitude may be positive or negative with absolute value not less than ν_m. The amplitudes $\pm\nu_m$ are viewed as the 'actual signal amplitudes' and a change detection scheme is applied to the two presumed parameters. Generally the traditional CUSUM procedure or FSS procedure is used. However, this technique has the following drawbacks: (i) If the actual signal amplitude is much greater than ν_m, then the technique may be too conservative in terms of false alarm probability if the thresholds are chosen according to ν_m. (ii) If the noise is non-Gaussian, this technique may not work well. For example, suppose f_0 is Gaussian with zero mean, while f_1 is uniform also with zero mean.

The MH approach does not add significant complexity to existing schemes if the number of unknown parameters is very small and the observations are i.i.d. before and after the change-time m. However, if there are many unknown parameters and the observations are dependent, the MH approaches may not be practical since the complexity of the algorithm increases significantly with the number of multiple hypotheses. In addition, our simulations have shown that the algorithm may not always converge to the true parameters and frequently results in poor parameter estimates. This latter problem makes it difficult to apply the algorithm to adaptive parameter estimation.

98

II. Maximum Likelihood Estimation Approach

For the binary detection problem with unknown parameters, a maximum likelihood estimation (MLE) technique, known as the generalized likelihood ratio test, is widely used [97]. The basic idea of the MLE technique is to replace the unknown parameters with their maximum likelihood estimators.

For the change detection problem, it has been suggested that the CUSUM statistic should be modified to [9]

$$T_n = \max_{1 \leq i \leq n} \max_{\theta} S_i^n(\theta). \tag{4.1.6}$$

However, the above statistic cannot be written in a recursive form, and therefore is too computationally complex to be practically implemented, particularly when the actual change time m is very large. A moving-window maximum likelihood estimation (MLE) technique combined with a conventional likelihood ratio procedure is an effective approach to solve this problem [9], which may be referred to as the *maximum likelihood ratio (MLR)* procedure. The MLR procedure is described as follows: Beginning from $n = w$, $X_1, ..., X_w$ are taken, an MLE θ^w of θ is computed from observation X_i^w, $i = 1, 2, ..., w$. Then a binary hypothesis test of H_0 v.s. $H_i(\theta^w)$ is performed over the most likely hypothesis (i.e. over H_i for $i = 1, 2, ..., w$). If $H_i(\theta^w)$ is accepted, then the test is terminated. If H_0 is accepted, one more sample X_{w+1} is taken and a new MLE θ^{w+1} of θ is computed from observations X_i^{w+1}, $i = 2, 3, ..., w + 1$. A binary hypothesis test of H_0 v.s. $H_i(\theta^{w+1})$ is performed over the most likely hypothesis. Again, if $H_i(\theta^{w+1})$ is accepted, the test is terminated. If H_0 is accepted, one more sample X_{w+2} is taken, and the procedure repeats until $H_i(\theta^n)$ is first accepted ($i \geq 1$). Let the total sample size be N_m. The decision threshold for the likelihood ratio test is denoted by d. Formally, the decision rule is

$$T_n(\theta^n) \triangleq \max_{n-w+1 \leq i \leq n} \max_{\theta} S_i^n(\theta) \begin{cases} \geq d & a_i \text{ for } n - w + 1 \leq i \leq n \\ < d & a_0 \end{cases} \tag{4.1.7}$$

for all $n = w, w + 1, ...$, and where $a_i \triangleq \{$accept $H_i\}$. We should notice that the decision

99

statistic $T_n(\theta^n)$ is a compromise between a CUSUM statistic and a moving-window FSS statistic.

This approach is interesting particularly for the situation where parameter estimation is the most important goal. This situation occurs in applications such as adaptive channel identification and adaptive parameter estimation. The major source of complexity of the algorithm is to adaptively estimate the unknown parameters. This estimation may not be an issue if the likelihood function is a simple quadratic function of the unknown parameter. For more complicated likelihood functions, search algorithms such as Widrow-McCool's linear random search [123] may be used to achieve reasonable performance at a more affordable complexity cost.

4.1.3 Multiple-Hypothesis Algorithms for Change Detection

In this subsection, we propose two multiple-hypothesis algorithms to solve the problem of change detection with unknown parameters, which are referred to as the multiple hypothesis (MH) MAR procedure and the alternative multiple hypothesis (AMH) MAR procedure respectively.

I. MH MAR Procedure

Let $H_m(\theta_k)$ denote the hypothesis that the change occurs at time m and the distribution parameter of f_1 is θ_k, i.e.

$$H_m(\theta_k): \quad X_i \sim \begin{cases} f_0(x_i) & \text{if } 1 \leq i < m \\ f_1(x_i, \theta_k) & \text{if } i \geq m, \end{cases} \quad (4.1.8)$$

where $m \geq 1$ and $k = 1, 2, ..., M$. It assumed that the distribution parameter (possibly a vector) θ of f_1 may take M possible values. Practically, the actual θ does not have to be exactly the same as any one of these θ_k.

A bank of M MAR procedures are applied with the parameters of f_1 to be $\theta_1, \theta_2, ..., \theta_M$

100

respectively. When any one of these procedures reaches a final decision, then H_m is accepted and the whole procedure is terminated. Formally, this MH MAR procedure is expressed as follows. The decision statistic of the MH MAR procedure is

$$\hat{r}(x_1^n, \theta_k) = \frac{T_n(\theta_k)\pi_{0k}c_{0k} - (1 - \pi_{0k})}{T_n(\theta_k)\pi_{0k} + 1 - \pi_{0k}}n, \qquad (4.1.9)$$

where

$$T_n(\theta_k) \triangleq \max_{1 \le i \le n} S_i^n(\theta_k) = \max\{T_{n-1}(\theta_k), 1\}l_n(\theta_k), \qquad (4.1.10)$$

where c_{0k} is design parameter chosen to achieve certain performance, $T_0(\theta_k) = 1$, $k = 1, 2, ..., M$, and $n \ge 1$. As soon as the following condition

$$\hat{r}_{mh}(x_1^{n+1}, \theta_k) > \hat{r}_{mh}(x_1^n, \theta_k) \qquad (4.1.11)$$

is first satisfied for any $k \in \{1, 2, ..., M\}$ and any $n \ge 1$, it is decided that the change m has occurred. Similar to the MAR procedure, the constraints

$$c_{0k} \le \frac{1 - \pi_{0k}}{\pi_{0k}} \qquad (4.1.12)$$

for all $k = 1, 2, ..., M$ are needed to assure that $N \ge 1$.

Notice that the above procedure is a generalization of the MAR procedure introduced in Chapter 3 to the multiple hypotheses situation.

II. AMH MAR Procedure

In this subsection, we present an alternative multiple hypothesis (AMH) MAR procedure.

Similar to the MH MAR procedure, the parameters (possibly vectors) of f_1 are assumed to be $\theta_1, \theta_2, ..., \theta_M$ respectively. Then an M-ary hypothesis

$$H_m(\theta_k): \quad X_i \sim \begin{cases} f_0(x_i) & \text{if } 1 \le i < m \\ f_1(x_i, \theta_k) & \text{if } i \ge m, \end{cases} \qquad (4.1.13)$$

101

is formulated, where $m \geq 1$ and $k = 1, 2, \ldots, M$. Similar to the MH MAR procedure, the CUSUM statistic for each presumed parameter is

$$T_n(\theta_k) = \max\{T_{n-1}(\theta_k), 1\} l_n(\theta_k),$$ (4.1.14)

where $k = 1, 2, \ldots, M$.

Instead of applying a bank of MAR procedures directly, we take

$$\hat{T}_n = \max_{k=1,2,\ldots,M} T_n(\theta_k),$$ (4.1.15)

as an estimate of the CUSUM statistic. Then the AMH MAR decision statistic is

$$\hat{r}_a(x_1^n) = \frac{\hat{T}_n \pi_0 c_0 - (1 - \pi_0)}{\hat{T}_n \pi_0 + 1 - \pi_0} n.$$ (4.1.16)

the AMH MAR procedure is that one stops and decides that a change has occurred as soon as

$$\hat{r}_a(n + 1, x_1^{n+1}) > \hat{r}_a(x_1^n)$$ (4.1.17)

is first satisfied for any $n \geq 1$. Similar to the MAR procedure, we require that

$$c_0 \leq \frac{1 - \pi_0}{\pi_0}$$ (4.1.18)

to assure that $N \geq 1$.

III. Convergence of the MH and AMH MAR Procedures

In this subsection the convergence of the MH and AMH MAR procedures is studied, which may be helpful for choosing θ_k to avoid possible divergence. For the MH and AMH MAR procedures, the almost sure (a.s.) convergence implies that the decision statistic for some presumed parameter is a convex function of n for sufficiently large n.

We first define an information term which measures the distance between two distributions.

$$I(\theta_j, \theta_k) \triangleq \int_{-\infty}^{+\infty} f_1(x, \theta_j) \ln \frac{f_1(x, \theta_k)}{f_0(x)} dx.$$ (4.1.19)

102

Obviously

$$I_k \triangleq I(\theta, \theta_k) = \int_{-\infty}^{+\infty} f_1(x,\theta) \ln \frac{f_1(x,\theta_k)}{f_0(x)} dx = E_m\{\ln \frac{f_1(X,\theta_k)}{f_0(X)}\}. \qquad (4.1.20)$$

It can be shown that

$$-\infty < I_k \leq \max_{k=1,2,...M} I_k \triangleq I_{\max} \leq I(\theta,\theta) \triangleq I_f. \qquad (4.1.21)$$

Practically it can be easily verified that for most distributions a smaller distance between θ and θ_k corresponds to a larger I_k, and vice versa. Therefore $I_f - I_k$ can be physically viewed as a reasonable measure of distance between θ and θ_k, where $k \in \{1, 2, ..., M\}$. The distance between the set $\{\theta_1..., \theta_M\}$ and the actual parameter θ may be measured by $I_f - I_{\max}$

The necessary and sufficient condition for the a.s. convergence of the AMH and MH MAR procedures is given by

PROPOSITION 4.1 *If the change time $m < +\infty$ and the stopping variable $N \geq 1$, then the necessary and sufficient condition for the MH and AMH MAR procedures to have the a.s. convergence property is that there exists some $k \in \{1, 2, ..., M\}$ such that*

$$I_k > 0 \qquad (4.1.22)$$

or equivalently

$$I_{\max} > 0 \qquad (4.1.23)$$

where

$$I_{\max} = \max_{k=1,2,...,M} I_k. \qquad (4.1.24)$$

□

PROOF: Using the same argument as that used in the proof of Propositions 3.1 and 3.2, it follows that, as $n \to +\infty$, the following convergences hold almost surely:

$$\hat{r}(x_1^n, \theta_k) \to \frac{e^{(n-m+1)I_k} \pi_{0k} c_{0k} - (1 - \pi_{0k})}{e^{(n-m+1)I_k} \pi_{0k} + 1 - \pi_{0k}} n \qquad (4.1.25)$$

103

and

$$\hat{r}_a(x_1^n) \to \frac{e^{(n-m+1)I_{max}}\pi_0 c_0 - (1 - \pi_0)}{e^{(n-m+1)I_{max}}\pi_0 + 1 - \pi_0} n. \tag{4.1.26}$$

Here we present the proof for the MH MAR procedure. A similar argument holds for the AMH MAR procedure.

i) Necessity: When $I_k \leq 0$ for all $k = 1, 2, ..., M$ and $n \to +\infty$,

$$\hat{r}(x_1^n, \theta_k) \to \frac{e^{(n-m+1)I_k}\pi_{0k} c_{0k} - (1 - \pi_{0k})}{e^{(n-m+1)I_k}\pi_{0k} + 1 - \pi_{0k}} n \to -n.$$

That is, $\hat{r}(x_1^n, \theta_k)$ is asymptotically a monotonically decreasing function of n. Therefore from Eq.(5.1.11)

$$P\{N = +\infty\} > 0$$

and the MH MAR procedure is not a.s. convergent.

ii) Sufficiency: If $I_k > 0$ for some $k \in \{1, 2, ..., M\}$,

$$\hat{r}(x_1^n, \theta_k) \to \frac{e^{(n-m+1)I_k}\pi_{0k} c_{0k} - (1 - \pi_{0k})}{e^{(n-m+1)I_k}\pi_{0k} + 1 - \pi_{0k}} n$$

which is increasing with n for $n \to +\infty$. Therefore

$$P\{N < +\infty\} = 1.$$

The MH MAR procedure is a.s. convergent.

A similar argument holds for the AMH MAR procedure and is omitted here.

QED

From the above proposition it is seen that the number of presumed parameters should not be too small if a.s. convergence has to be guaranteed.

IV. Other Multiple Hypothesis Algorithms

It is natural to combine the MH approach with the CUSUM procedure. Using the MH approach the CUSUM statistic can be easily written in a recursive form, as is suggested by Eq.(4.1.10), i.e.

$$T_n(\theta_k) = \max_{1 \leq i \leq n} S_i^n(\theta_k) = \max\{T_{n-1}(\theta_k), 1\} l_n(\theta_k). \tag{4.1.27}$$

104

Then the MH CUSUM procedure is proposed here as a decision rule that one stops and decides a change has occurred as soon as

$$T_n(\theta_k) \geq B_k \qquad (4.1.28)$$

for any $k = 1, 2, ..., M$.

The MH CUSUM procedure proposed here may be viewed as a generalization of the two-sided CUSUM (V-mask) procedure where $M = 2$. From the simulation results in Chapter 3 we conjecture that the MH CUSUM procedure proposed here may not be able to perform better than the MH MAR and AMH MAR procedures.

V. Number of Presumed Parameters

According to Proposition 4.1, a sufficiently large number of presumed parameters have to be chosen in order to satisfy the a.s. convergence condition. However, using a larger number of presumed parameters does not always imply better performance. This is true because too many presumed parameters would result in a large false alarm probability. The following analysis explores this trade-off.

Let a_i denote the event that the procedure using the ith presumed parameter causes a false alarm, i.e., that

$$\alpha_i = P_0\{a_i\} \qquad (4.1.29)$$

where $i = 1, 2, ..., M$. For small M, we may suppose the bank of M procedures are independent of each other. Then the overall false alarm probability is

$$
\begin{aligned}
\alpha &= P_0\{a_1 \bigcup a_2 \bigcup ... \bigcup a_M\} \\
&= \sum_{k=1}^{M}(-1)^{k+1} \sum_{i_1 < i_2 < ... < i_k} P_0\{a_{i_1}\} P_0\{a_{i_2}\}...P_0\{a_{i_k}\} \\
&= \sum_{k=1}^{M}(-1)^{k+1} \sum_{i_1 < i_2 < ... < i_k} \alpha_1 \alpha_2 ... \alpha_k. \qquad (4.1.30)
\end{aligned}
$$

If

$$\alpha_1 = \alpha_2 = ... = \alpha_M \qquad (4.1.31)$$

105

then

$$\alpha = \sum_{k=1}^{M}(-1)^{k+1}C_M^k\alpha_1^k$$
$$= 1 + \sum_{k=0}^{M}(-1)^{k+1}C_M^k\alpha_1^k$$
$$= 1 - (1-\alpha_1)^M \qquad (4.1.32)$$

where

$$C_M^k \triangleq \frac{M!}{k!(M-k)!}. \qquad (4.1.33)$$

Notice that using more presumed parameters (a larger M) may result in a smaller α_1 since I_{max} may be closer to I_f, i.e., that $\alpha_1(M)$ may be a (non-strictly) decreasing function of M. The smallest $\alpha_1(M)$ is the one when $\theta_k = \theta$ for some $k \in \{1,2,...,M\}$, which is denoted by α_f. Notice that

$$\alpha_1(M) \geq \alpha_f > 0. \qquad (4.1.34)$$

The situation that the presumed parameter θ_k equals the actual parameter θ may happen by chance if M is finite or may be a result of $M \to +\infty$. Rewriting (4.1.32) as

$$\alpha = 1 - (1 - \alpha_1(M))^M \qquad (4.1.35)$$

we notice that α is a convex function of M. Therefore, an optimum M may exist. However, it seems difficult to find this optimum M analytically.

Interestingly, when the number M of presumed parameters goes to $+\infty$, the overall false alarm probability α goes to 1 according to (4.1.32). For the change detection problem the detection probability $\pi = 1 - \alpha$. Therefore an extremely large M may result in both a very large false alarm probability and a very small detection probability.

106

4.1.4 Maximum Likehood Estimation Approaches for Change Detection

As is discussed in previous subsections, combining maximum likelihood estimation (MLE) with the CUSUM statistic yields [9]

$$T_n = \max_{1 \leq i \leq n} \max_{\theta} S_i^n(\theta).$$
(4.1.36)

However, the above statistic cannot be written in a recursive form, and therefore has too much computational complexity to be practically implemented, particularly when the actual change time m is very large. Therefore a moving-window MLE technique must be used, i.e., that

$$T_n(\theta^n) = \max_{n-w+1 \leq i \leq n} \max_{\theta} S_i^n(\theta)$$
(4.1.37)

for $n = w, w + 1, \ldots$ In contrast to the MLR approach discussed in Subsection 4.2.2 (see [9]), an MLE MAR procedure is proposed here. The MLE MAR statistic is

$$\hat{r}_{mle}(x_1^n) = \frac{T_n(\theta^n)\pi_0 c_0 - (1 - \pi_0)}{T_n(\theta^n)\pi_0 + 1 - \pi_0} n.$$
(4.1.38)

The decision rule of the MLE MAR procedure is that one stops and decides that a change has occurred as soon as

$$\hat{r}_{mle}(x_1^n) > \hat{r}_{mle}(x_1^{n-1}).$$
(4.1.39)

- **Remark 1:** The MLE MAR procedure can be easily combined with an adaptive parameter estimation algorithm to achieve combined detection-estimation. This is an advantage over the MH approaches discussed in the last subsection since the latter approaches often lead to poor estimation results according to our simulations.

- **Remark 2:** In practice, if w is not very large compared to m, the simulations in the next subsection show that the detection performance of the MLE MAR procedure is inferior to both the MH and AMH MAR procedures. If the window size w is very close to m, the performance of the MLE MAR procedure can be

close to either the MH or AMH MAR procedures. However, for a large change time m, it is not practical to make the window size w close to m.

- **Remark 3:** If the window-size w is not very large, the algorithm can be implemented with reasonable computation. However, if the window-size is chosen to be large to have better performance, the computational complexity of the MLE MAR procedure is prohibitive.

4.1.5 Simulation Results

I. Simulation Model

We use a model similar to one used in Chapter 3, i.e.

$$X_i = \begin{cases} N_i & \text{if } i < m \\ N_i + \theta & \text{if } i \geq m \end{cases} \qquad (4.1.40)$$

where $i = 1, 2, ...$, θ is an unknown constant, and N_i is i.i.d additive white Gaussian noise (AWGN). The densities for this model are

$$f_0(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x_i^2}{2\sigma^2}} \quad \text{if } i < m \qquad (4.1.41)$$

and

$$f_1(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\theta)^2}{2\sigma^2}} \quad \text{if } i \geq m \qquad (4.1.42)$$

where σ^2 is the variance (and also the energy) of the AWGN N_i. The signal-to-noise ratio (SNR db) is

$$SNR = 20 \log_{10} \frac{\theta}{\sigma}. \qquad (4.1.43)$$

II. Comparison between the MLE MAR Procedure and the MLR Procedure

We first compare the performance of the MLR procedure with that of the MLE MAR procedure. It is shown that the MLE MAR procedure is always better than the MLR procedure in all cases tried. In next subsection, it will be shown that if some presumed

108

parameters are not very far away from the actual parameter, then either the MH MAR procedure or the AMH MAR is always better than the MLE MAR procedure in all cases tried.

The MLR procedure has appeared in the published literature [9]. The MLE MAR procedure is first proposed in this thesis (Section 4.1.4). Both procedures use a moving window maximum likelihood estimation (MLE) technique. The MLR procedure combines the MLE with a likelihood ratio technique. The MLE MAR procedure combines the MLE with the MAR procedure introduced in this thesis. Both procedures use a window size $w = 100$. The design parameters for the MLE MAR procedure is $\pi_0 = c_0 = 1.1 \times 10^{-6}$. The threshold for the MLR procedure is $\ln B = 8.38$. For all signal-to-noise ratios (SNR), the design parameters do not change. The detection probability for both procedures is $\pi = 0.9$ and the false alarm probability $\alpha = 0.1$ with a 90% confidence interval of ± 0.002. The actual detection (false alarm) probability of the MLE MAR procedure is always slightly larger (smaller) than the MLR procedure.

From Figures 4.1–4.3, it is clear that the average delay of the MLE MAR procedure is always smaller than the MLR procedure. When the SNR is small (less than -8 db), the superiority of the MLE MAR procedure is significant. When the SNR is not small (greater than -8 db), the difference of the average delay between the two procedures is not significant. However, the false alarm ARL of the MLE MAR procedure is always significantly better than the MLR procedure. Therefore, the conclusion is that the MLE MAR procedure is always better than the MLR procedure in all cases tried.

Figure 4.1: *Quickest detection of a change in the mean of Gaussian distribution (with unknown amplitude). The false alarm probability is $\alpha = 0.1$. The change-time is $m = 500$. The average delay v.s. SNR (weak signal).*

110

Figure 4.2: *Quickest detection of a change in the mean of Gaussian distribution (with unknown amplitude). The false alarm probability is $\alpha = 0.1$. The change-time is $m = 500$. The average delay v.s. SNR.*

111
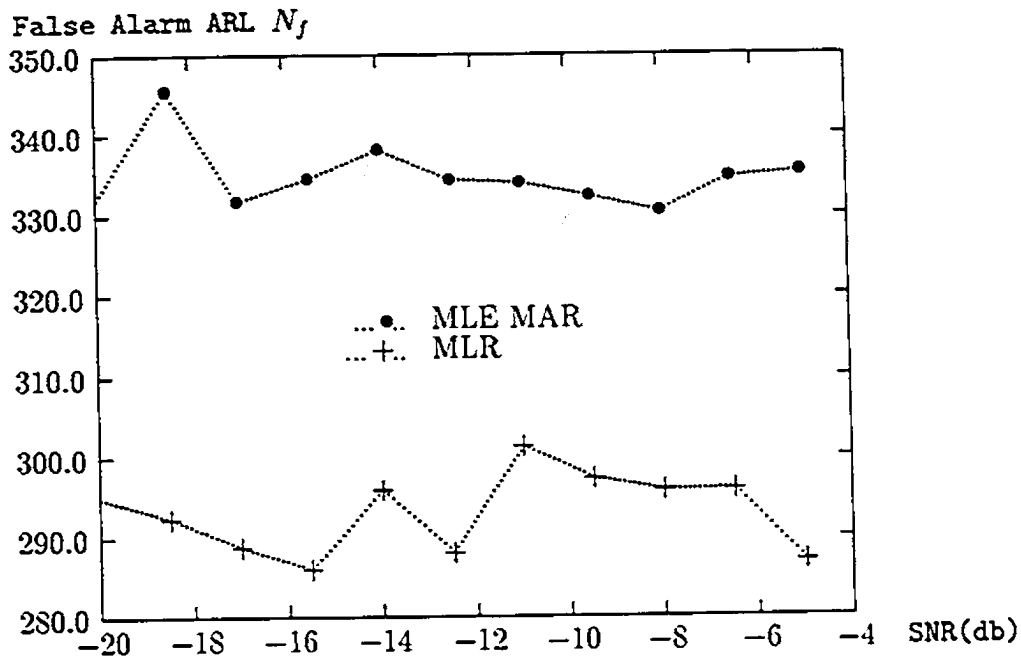
Figure 4.3: *Quickest detection of a change in the mean of Gaussian distribution (with unknown amplitude). The false alarm probability is $\alpha = 0.1$. The change-time is $m = 500$. The false alarm ARL N_f v.s. SNR.*

112

III. Comparison between the MH, AMH and MLE MAR Procedures

In the last subsection it was shown that the MLE MAR procedure is better than the MLR procedure. In this subsection we will show that either the MH MAR procedure or the AMH MAR procedure can be better than the MLE MAR procedure if the distance between some presumed parameter and the actual parameter is not very large. The results are shown in Figures 4.4–4.8.

The presumed parameters (the means of Gaussian distribution) are listed in Table 4.1. The actual parameters (the means of Gaussian distribution) used in the simulation are listed in Table 4.2. The distance measure $I_f - I_{max}$ is also listed, where

$$I_f = E_1\{\ln \frac{f_1(X,\theta)}{f_0(X)}\} = \frac{1}{2}(\frac{\theta}{\sigma})^2 \tag{4.1.44}$$

and

$$I_{max} = \max_{k=1,2,...,M} I_k = \max_{k=1,2,...,M} E_1\{\ln \frac{f_1(X,\theta_k)}{f_0(X)}\} = \max_{k=1,2,...,M} \frac{\theta_k(2\theta - \theta_k)}{2\sigma^2}. \tag{4.1.45}$$

In the simulation $\sigma = 0.25$. It is seen that the a.s. convergence condition $I_{max} > 0$ outlined in Proposition 4.1 is satisfied.

The design parameters for the AMH MAR procedure are $\pi_0 = c_0 = 3.1 \times 10^{-6}$ and do not change for all SNR's. The design parameters pi_{0k} and c_{0k} for the MH MAR procedure are listed in Table 4.3 and do not change for all SNR's.

From Figure 4.4, it is seen that for very weak signals (the SNR is less than -14.5db), the average delay performance of the MH MAR procedure is better than other two procedures. The MLE MAR procedure (and the MLR procedure according to Figures 4.1–4.3) is worse than both the MH and AMH MAR procedures. The difference between the MH and AMH MAR procedures is a result of the fact that $\theta_k \neq \theta$ for all k and that the MH MAR procedure is more robust than AMH MAR.

Figures 4.5–4.7 show that for moderately weak and strong signals, the average delay performance of the AMH MAR procedure is the best among all procedures. The AMH

113

MAR procedure seems to be a very good procedure if the actual parameter is quite close (but not necessarily equal) to a presumed parameter.

However, Figures 4.5-4.8 reveal that the false alarm ARL performance of the MLE MAR procedure is the best among all procedures. However, in simulations we found that this is the case only if the window size w is not much smaller that the actual change time m. In addition, since the false alarm probability α is generally very small, the false alarm ARL performance may not be very important.

In conclusion, we remark that for the situation where the average delay and the false alarm probability are more important than the false alarm ARL, either the AMH or MH MAR procedures would be preferred. If the false alarm ARL performance is much more important than the average delay and the false alarm probability, and if combined detection-estimation is very important, then the MLE MAR procedure may be preferred. However, all three procedures introduced in this thesis are preferable to the existing MLR procedure according to the simulation results.

114
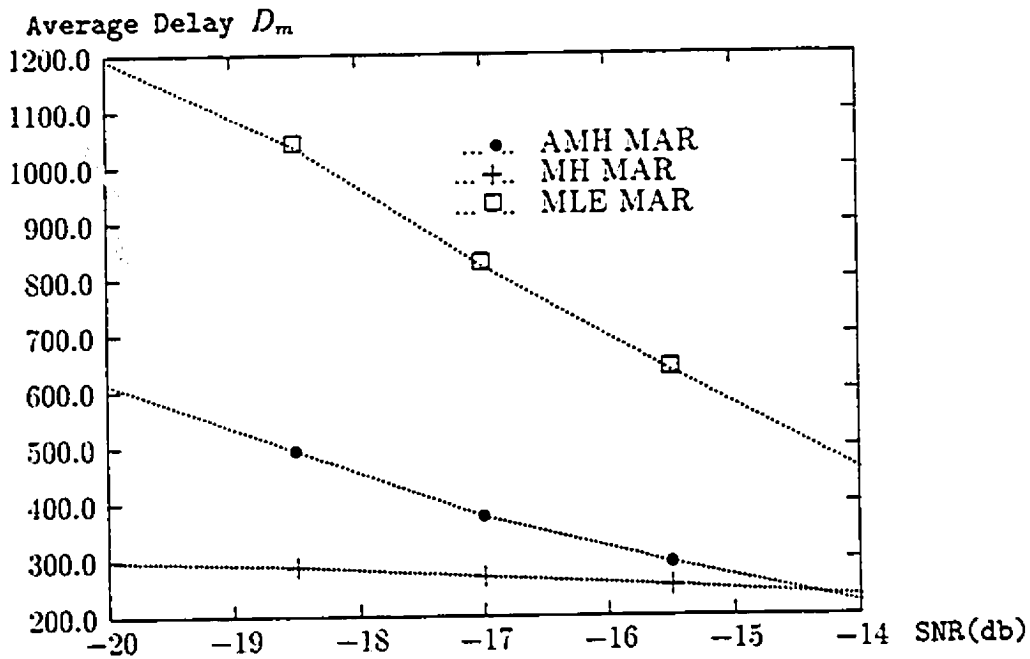
Figure 4.4: *Quickest detection of a change in the mean of Gaussian distribution (with unknown amplitude). The false alarm probability is $\alpha = 0.1$. The change-time is $m = 500$. The average delay D_m v.s. SNR (very weak signal).*
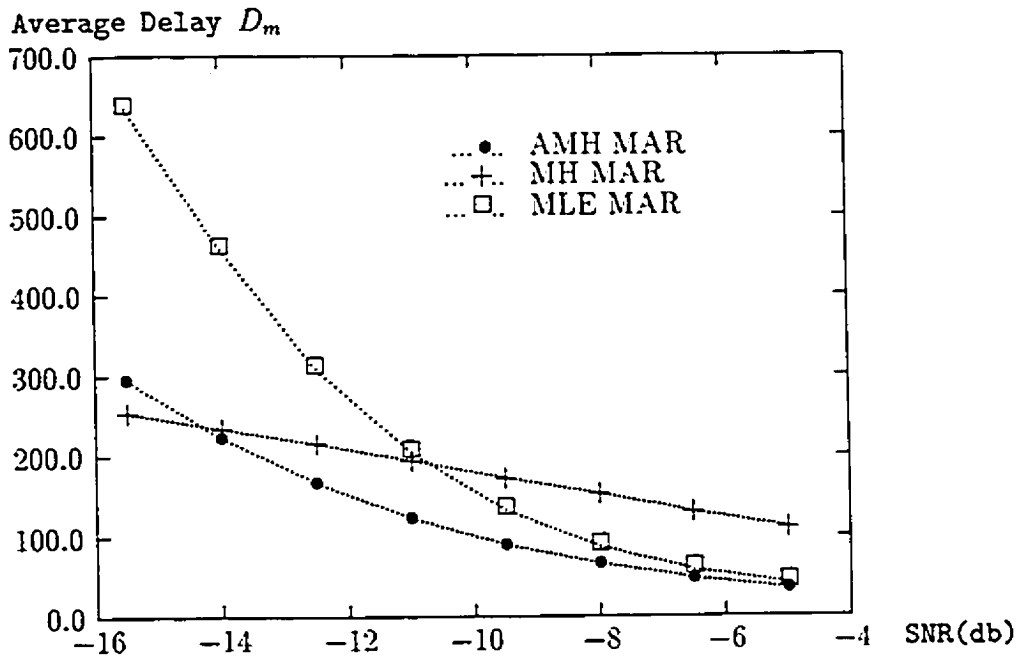
Average Delay D_m

Figure 4.5: *Quickest detection of a change in the mean of Gaussian distribution (with unknown amplitude). The false alarm probability is $\alpha = 0.1$. The change-time is $m = 500$. The average delay D_m v.s. SNR (moderately weak signal).*

116
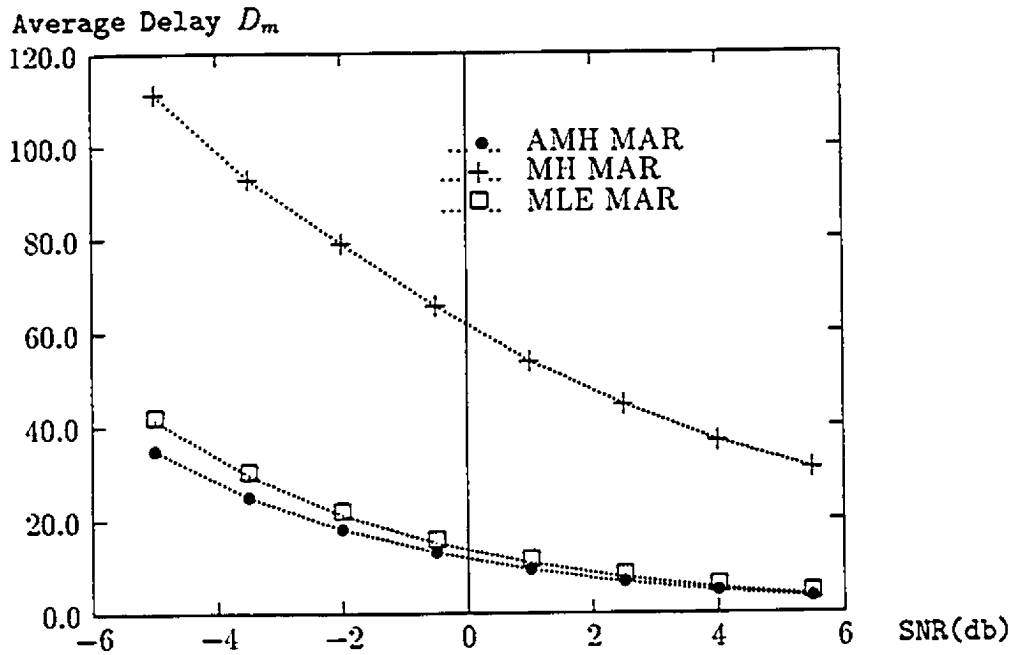
Figure 4.6: *Quickest detection of a change in the mean of Gaussian distribution (with unknown amplitude). The false alarm probability is $\alpha = 0.1$. The change-time is $m = 500$. The average delay D_m v.s. SNR (strong signal).*

117
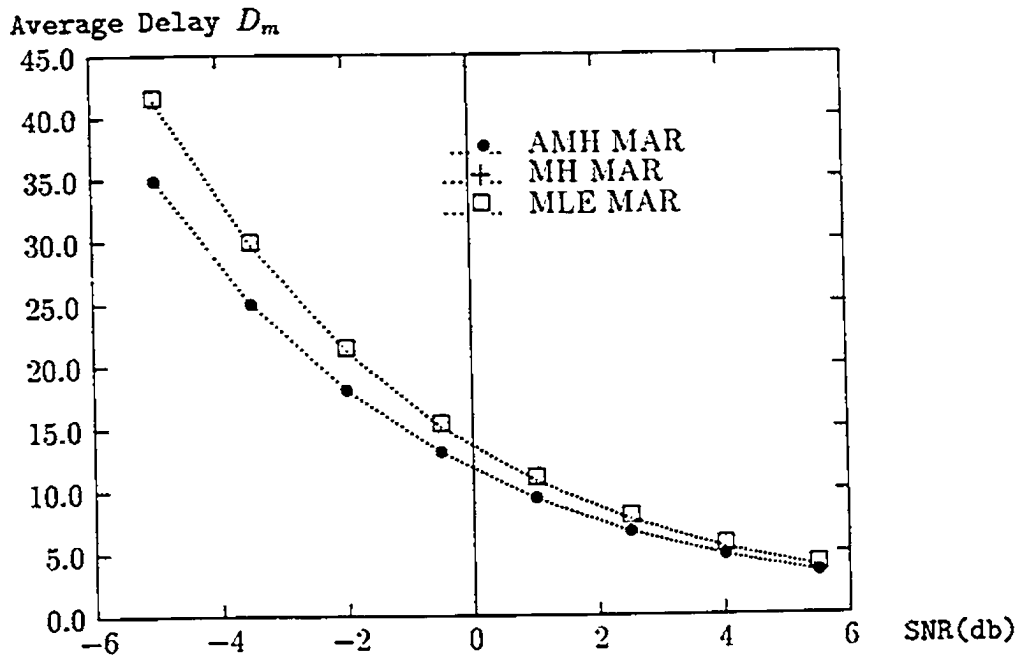
Figure 4.7: *Quickest detection of a change in the mean of Gaussian distribution (with unknown amplitude). The false alarm probability is $\alpha = 0.1$. The change-time is $m = 500$. The average delay D_m v.s. SNR (strong signal).*
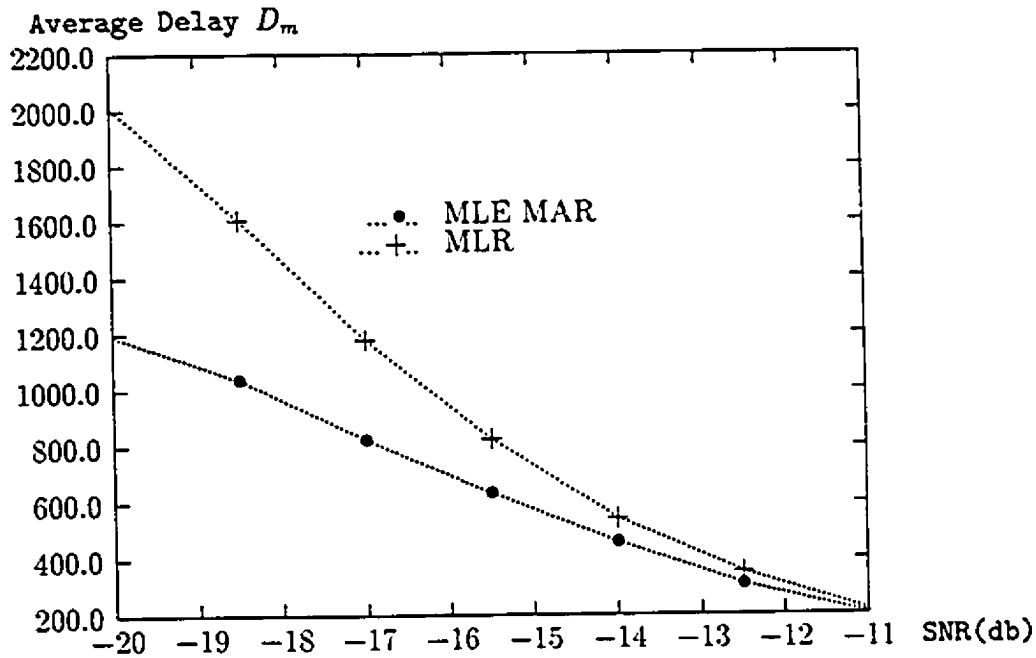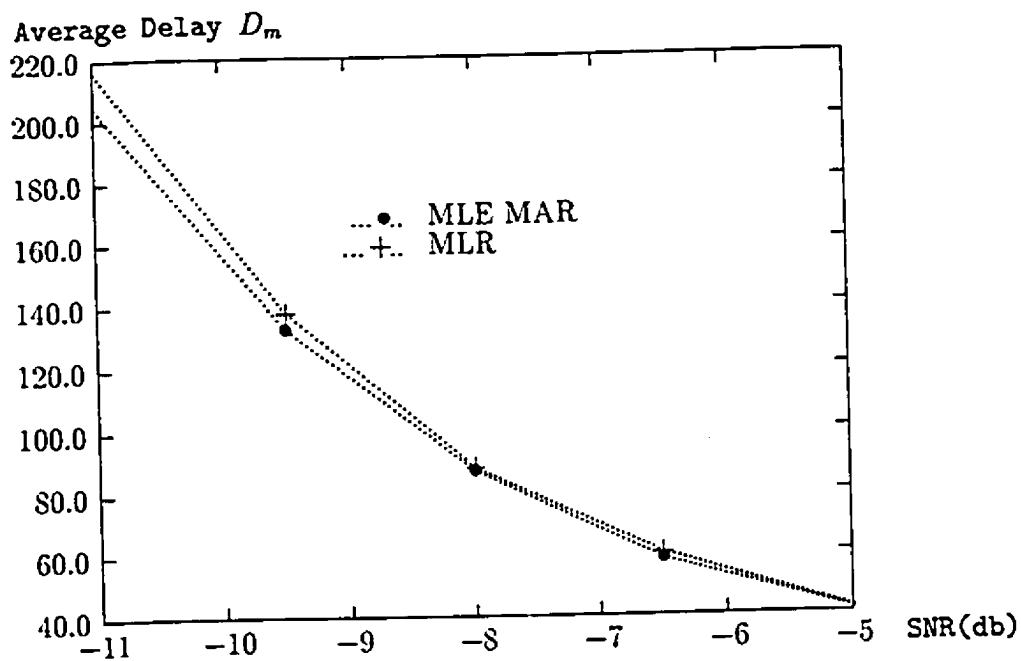
118
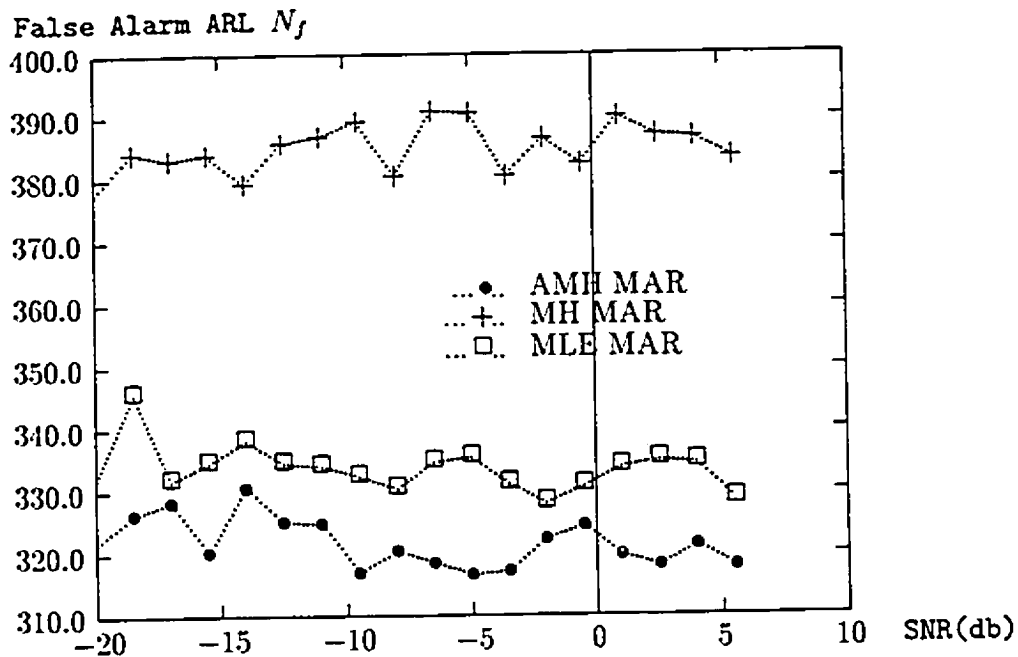
Figure 4.8: *Quickest detection of a change in the mean of Gaussian distribution (with unknown amplitude). The false alarm probability is $\alpha = 0.1$. The change-time is $m = 500$. The false alarm ARL N_f v.s. SNR.*

119

IV. Effect of the Distance between the Actual and Presumed Parameters on the Performance

The performances of the MH and AMH MAR procedures obtained in the previous subsection are based on the assumption that the distance $I_f - I_{max}$ between the actual and presumed parameters is not very large, typically $I_f - I_{max} < 0.01$ (see Table 4.2). However, if the distance $I_f - I_{max}$ between the actual and presumed parameters is large, the results are quite different.

Table 4.4 shows the effect of the distance $I_f - I_{max}$ between the actual and presumed parameters on the performance of the MH and AMH MAR procedures. The presumed parameters are listed in Table 4.1, which are all greater than zero. The actual parameters listed in Table 4.4 are all less than zero. The performance of the MLE MAR procedure is also listed for comparison. The change time is $m = 500$. The false alarm probability $\alpha = 0.1$. The average delay for MLE, MH MAR and AMH MAR procedures are denoted by D_{mMLE}, D_{mMH} and D_{mAMH} respectively. The false alarm ARL for MLE, MH and AMH MAR procedures are denoted by N_{fMLE}, N_{fMH} and N_{fAMH} respectively.

It is seen that the a.s. convergence condition is not satisfied since $I_{max} < 0$. As a result, the performance of the AMH MAR procedure is very poor (it almost diverges), even though it performs very well for moderately weak and strong signals if the distance $I_f - I_{max}$ is small (see Figures 4.5-4.7 and Table 4.2).

The MH MAR procedure appears quite robust. Even when the a.s. convergence condition $I_{max} > 0$ is not satisfied, it still performs well if the distance $I_f - I_{max} < 0.04$. It seems that the MH MAR procedure is convergent with high probability even for fairly large distance $I_f - I_{max}$.

It is also observed that the MLE MAR procedure does not have a convergence problem. Its performance for $\theta < 0$ is the same as that for $\theta > 0$.

4.2 Change Detection for Nonstationary or Dependent Observations

In this section, we first study the sequential probability ratio test (SPRT) for nonstationary observations, which has certain applications to change detection since the CUSUM procedure is essentially a repeated SPRT (see Chapter 2). Then we propose a number of possible solutions to the change-detection problem for nonstationary or dependent observations. The non-i.i.d. change-detection problem is an interesting but difficult topic. We will not spend much effort on this problem and will propose only several possible solutions without much elaboration.

4.2.1 Sequential Probability Ratio Test for Nonstationary Observations

I. Problem Statement

The problem of minimizing the average sample size in a binary H_0 v.s. H_1 test has been studied extensively since Wald published his pioneering work [119] in 1947. For i.i.d. observations, it is well-known that the optimal detection technique is Wald's SPRT [120, 33]. However, for non-i.i.d. situations, the properties of the SPRT are less known and no practically implementable results are available [114]. By using a martingale-theoretic approach, it was shown in [27] that for nonstationary but independent observations the SPRT with varying thresholds is still optimal. It was also shown [27] that for Gaussian and dependent observations the SPRT is equivalent to that for nonstationary but independent observations. Unfortunately, the problem of designing the varying decision thresholds is still unsolved, and this generalized SPRT cannot be practically implemented. In this thesis, we will provide a different approach to generalizing the SPRT to nonstationary situations. As a result of this generalization, several practical properties of the optimum decision thresholds for the generalized SPRT are shown.

121

For nonstationary and independent observations, this problem can be expressed in the following way: Suppose $\{X_i | i = 1, 2, ...\}$ are independent random variables observed sequentially. Consider hypothesis test

$$H_0: \quad X_i \sim f_{0i}(x_i)$$
$$H_1: \quad X_i \sim f_{1i}(x_i) \tag{4.2.1}$$

where $f_{0i} \neq f_{0j}$, $f_{1i} \neq f_{1j}$ for some $i \neq j$ $(i, j = 1, 2, ...)$ and $f_{0i} \neq 0$ for any $i = 1, 2, ...$ It is assumed that X_i are independent under both H_0 and H_1. Since the observations are nonstationary, we modify the definitions of P_i, E_i $(i = 0, 1)$ as follows. Define

$$P_0\{X_i \leq x\} \triangleq \int_{-\infty}^{x} f_{0i}(t) dt \tag{4.2.2}$$

and

$$P_1\{X_i \leq x\} \triangleq \int_{-\infty}^{x} f_{1i}(t) dt \tag{4.2.3}$$

and expectations E_0, E_1 are defined under P_0 and P_1, respectively. The above definitions of P_i, E_i $(i = 0, 1)$ are natural extensions to that in Chapter 2. The one-sample likelihood ratio l_i is defined as

$$l_i \triangleq \frac{f_{1i}(x_i)}{f_{0i}(x_i)} \tag{4.2.4}$$

Notice that

$$S_k^n = \prod_{i=k}^{n} l_i \tag{4.2.5}$$

for $k \leq n$. Particularly, for $k = 1$ S_k^n may be written as

$$S_n \triangleq S_1^n. \tag{4.2.6}$$

The generalized SPRT is defined as the following decision rule:

$$
S_1^n
\begin{cases}
\geq B_n & a_1 \\
\in (A_n, B_n) & \text{no decision and continue sampling} \\
\leq A_n & a_0
\end{cases}
\tag{4.2.7}
$$

122

for varying thresholds A_n and B_n, where $0 \leq A_n \leq B_n$. This generalized SPRT is referred to as NSPRT.

The problem is to minimize

$$E_i(N) \tag{4.2.8a}$$

for given error probabilities

$$\alpha = P_0(a_1) \tag{4.2.8b}$$

and

$$\beta = P_1(a_0), \tag{4.2.8c}$$

where $i = 0, 1$ and a_i denotes "accepting hypothesis H_i" (defined in Chapter 2).

II. A Generalized Wald Lower Bound for NSPRT

For some decision rule we have a stopping variable N ($N \in \{1, 2, ...\}$). Define

$$\overline{I_1} \triangleq E_1(\frac{\sum_{i=1}^{N} \ln l_i}{E_1(N)}) \tag{4.2.9}$$

and

$$\overline{I_0} \triangleq E_0(\frac{\sum_{i=1}^{N} \ln l_i}{E_0(N)}). \tag{4.2.10}$$

It is obvious that the average information terms $\overline{I_1}$ and $\overline{I_0}$ are constants not related to N but to the pdfs of X_i. Using the average information terms, we can generalize Wald's lower bound to non-stationary case:

THEOREM 4.1 *For any procedure to test H_1 versus H_0 as defined above, given the error probability of the first kind α and the second kind β, the lower bound for the expectation of the sample size N is*

$$E_1(N) \geq \overline{I_1}^{-1} \left\{ (1 - \beta) \ln \frac{1 - \beta}{\alpha} + \beta \ln \frac{\beta}{1 - \alpha} \right\}, \tag{4.2.11}$$

123

$$E_0(N) \geq \overline{I_0}^{-1} \left\{ \alpha \ln \frac{1-\beta}{\alpha} + (1-\alpha) \ln \frac{\beta}{1-\alpha} \right\}. \qquad (4.2.12)$$

□

To prove Theorem 4.1, we first present the following lemma:

LEMMA 4.1 *Let Y be a random variable, then*

$$E_1(Y) = E_0(Y S_n). \qquad (4.2.13)$$

Let Θ denote a probability event and define the indicator function

$$I_\Theta \triangleq \begin{cases} 1 & X \in \Theta \\ 0 & X \notin \Theta. \end{cases} \qquad (4.2.14)$$

When $Y = I_\Theta$ then

$$P_1(\Theta) = E_0(S_n; \Theta) \qquad (4.2.15)$$

and when $Y = I_\Theta S_n^{-1}$ we have

$$P_0(\Theta) = E_1(S_n^{-1}; \Theta). \qquad (4.2.16)$$

□

PROOF:

$$
\begin{aligned}
E_1(Y) &= \int_\Omega y f_{11}(x)...f_{1n}(x) dx_1...dx_n \\
&= \int_\Omega y S_n f_{01}(x)...f_{0n}(x) dx_1...dx_n \\
&= E_0(Y S_n).
\end{aligned}
$$

Equations (4.2.15) and (4.2.16) are straightforward results of (4.2.13) using definition (4.2.14). **QED**

PROOF OF THEOREM 4.1: Using the properties of conditional expectation and conditional probability and Lemma 4.1 yields

$$\alpha = P_0(a_1)$$

124

$$= E_1(S_n^{-1}; a_1)$$

$$= E_1(S_n^{-1}|a_1)P_1(a_1)$$

$$= E_1(e^{-\ln S_n}|a_1)(1 - \beta)$$

$$\geq e^{-E_1(\ln S_n|a_1)}(1 - \beta)$$

$$= e^{-E_1(\ln S_n; a_1)/(1-\beta)}(1 - \beta),$$

where Jensen's inequality was used in the second-to-last expression.

Taking log on both sides yields

$$E_1(\ln S_n; a_1) \geq (1 - \beta) \ln \frac{1 - \beta}{\alpha}.$$

Similarly we have

$$E_1(\ln S_n; a_0) \geq \beta \ln \frac{\beta}{1 - \alpha}.$$

So

$$E_1(\ln S_n) = E_1(\ln S_n; a_0) + E_1(\ln S_n; a_1) \geq (1 - \beta) \ln \frac{1 - \beta}{\alpha} + \beta \ln \frac{\beta}{1 - \alpha}.$$

Therefore

$$
\begin{aligned}
E_1(\ln S_N) &= E_1(\ln S_N; N \in \{1, 2, ...\}) \\
&= \sum_{n=1}^{+\infty} E_1(\ln S_N; N = n) \\
&= \sum_{n=1}^{+\infty} E_1(\ln S_N|N = n)P_1(N = n) \\
&= \sum_{n=1}^{+\infty} E_1(\ln S_n)P_1(N = n) \\
&\geq \sum_{n=1}^{+\infty} ((1 - \beta) \ln \frac{1 - \beta}{\alpha} + \beta \ln \frac{\beta}{1 - \alpha})P_1(N = n) \\
&= ((1 - \beta) \ln \frac{1 - \beta}{\alpha} + \beta \ln \frac{\beta}{1 - \alpha}) \sum_{n=1}^{+\infty} P_1(N = n) \\
&= (1 - \beta) \ln \frac{1 - \beta}{\alpha} + \beta \ln \frac{\beta}{1 - \alpha}.
\end{aligned}
\tag{4.2.17}
$$

From the above result and the definitions of $\bar{I}_i$ we have

$$E_1(N) \geq \bar{I}_1^{-1}\{(1 - \beta) \ln \frac{1 - \beta}{\alpha} + \beta \ln \frac{\beta}{1 - \alpha}\}.$$

125

Similarly we can prove the second inequality.

QED

Generally it is difficult to calculate $\overline{T_1}$ and $\overline{T_0}$. However, for periodically varying signals with period T, when $E_i(N) >> T$, $\overline{T_i}$ can be approximated by

$$\overline{T_i} \approx \frac{E_i(\sum_{j=1}^{T} l_j)}{T} \qquad (4.2.18)$$

where $i = 0, 1$.

Theorem 4.1 shows the lower bound of the expected sample size for testing hypotheses about a non-stationary sequence under the given strength. Finally we remark that the relations between the decision thresholds A_n, B_n and the error probabilities formulated by Wald (1947) still hold for the non-i.i.d. situation (see, e.g. Tantaratana (1986)). The relations are:

$$\alpha \leq B_n^{-1}(1 - \beta), \qquad \beta \leq A_n(1 - \alpha). \qquad (4.2.19)$$

III. Optimality of NSPRT

Using Bayesian analysis, we now show that the NSPRT with varying thresholds is optimal among all procedures. Some important properties of the varying thresholds are also shown, which may be helpful for designing the varying thresholds for some practical problems.

In the following we prove the optimality of the NSPRT, i.e., that the NSPRT minimizes the expected sample size among all procedures for given error probabilities (or vice versa). This is formally expressed as Theorem 4.2. The proof of Theorem 4.2 is quite long, but similar to that given by Ferguson [33], which is a modified version of Wald-Wolfowitz's proof [120]. As an overview, this proof uses Bayesian analysis, where the Bayesian risk consists of two terms: (i) the Bayesian loss due to the false terminal decision, and (ii) the random sample size. It will be shown that this Bayesian procedure with varying thresholds is a NSPRT and any NSPRT is Bayesian. This result implies

126

that a linear combination of the expected sample size and error probabilities for the NSPRT is minimal among any procedures. Therefore the expected sample size for the NSPRT is minimized among all procedures for any given error probabilities (or vice versa).

This proof is organized in the following way: we first introduce essential notation for the Bayesian analysis, which is similar to that in [33] except for minor modifications. Then we show that a Bayesian procedure can be expressed as a NSPRT and vice-versa. These results parallel those found in [33] with certain generalizations. Using by-products of the optimality proof, we will derive properties of the NSPRT optimum thresholds for several important special cases.

We now introduce the relevant concepts in Bayesian analysis. Let π denote the prior probability that H_1 is true. The probability associated with event Θ is then $P_\pi(\Theta) = \pi P_1(\Theta) + (1 - \pi)P_0(\Theta)$. Let N denote the (possibly random) stopping time for the test of H_1 versus H_0. Define

$$\psi_n \triangleq P_\pi(N = n | n \geq 1) \tag{4.2.20}$$

It is obvious that ψ_n is the density of N, i.e., the probability of stopping at time n, which is the stopping rule for the sequential test. Define the final decision rule as

$$\delta_n \triangleq \begin{cases} 1 & a_1 \\ 0 & a_0 \\ any & \text{otherwise.} \end{cases} \tag{4.2.21}$$

Let c (> 0) represent a finite constant cost for each observation. Then the total cost for the test due to the length of observation is

$$C_N(\pi, \psi) = cN. \tag{4.2.22}$$

Let $L(\pi, \delta_N)$ denote the loss function due to a false terminal decision. Then the total risk for sequential stopping rule $\psi = (\psi_1, \psi_2, ...)$ and terminal decision rule δ_N is

$$R(\pi, \psi, \delta_N) = L(\pi, \delta_N) + C_N(\pi, \psi) = L(\pi, \delta_N) + cN, \tag{4.2.23}$$

and its expectation is denoted by

$$r(\pi, \psi, \delta) = E_\pi \{ L(\pi, \delta_N) \} + c E_\pi(N).$$ (4.2.24)

The Bayesian rule is to choose stopping rule ψ and terminal decision rule δ to minimize $r(\pi, \psi, \delta)$ for the posterior π after the samples have been observed. Let $\pi_{x_1^j}$ denote the posterior probability that H_1 is true after $X_1^j = x_1^j$ have been observed, i.e..

$$(\pi)_{x_1^j} \triangleq P(H_1 | X_1^j = x_1^j) = \frac{\pi S_1^j(x_1^j)}{\pi S_1^j(x_1^j) + 1 - \pi}.$$ (4.2.25)

By using Bayes' theorem, the above result can be easily verified. We note that the independence assumption is needed here. Similar to Theorem 1 in [33], p.314, and Eq. (7.43), we have

LEMMA 4.2 *Let δ_j^π be a fixed sample size (FSS) Bayes rule for the test of H_1 versus H_0 based on an observation sequence $X_1^j = x_1^j$ with fixed sample size j and given prior knowledge of π. For any fixed stopping rule ψ, $r(\pi, \psi, \delta)$ is minimized by $\delta^\pi = (\delta_0^\pi, \delta_1^\pi, \delta_2^\pi, ...)$. Let the loss due to a correct terminal decision be zero, w_{10} (0 < w_{10} < +\infty) be the loss due to falsely accepting H_0 when H_1 is true, and w_{01} (0 < w_{01} < +\infty) be the loss due to falsely accepting H_1 when H_0 is true. Then*

$$\delta_j^\pi = \begin{cases} 0 & \text{if } w_{10}\pi_{x_1^j} < w_{01}(1 - \pi_{x_1^j}) \\ any & \text{if } w_{10}\pi_{x_1^j} = w_{01}(1 - \pi_{x_1^j}) \\ 1 & \text{if } w_{10}\pi_{x_1^j} > w_{01}(1 - \pi_{x_1^j}) \end{cases}$$ (4.2.26)

and the minimum posterior expected Bayes loss is

$$\rho_j(\pi_{x_1^j}) = \begin{cases} w_{10}\pi_{x_1^j} & \text{if } \pi_{x_1^j} \le \frac{w_{01}}{w_{01}+w_{10}} \triangleq w \\ w_{01}(1 - \pi_{x_1^j}) & \text{if } \pi_{x_1^j} \ge \frac{w_{01}}{w_{01}+w_{10}} = w \end{cases}$$ (4.2.27)

for $j \ge 0$ with $\pi_0 \triangleq \pi$. □

The proof of Lemma 4.2 is the same as that for Theorem 1 in [33], p.314, where the i.i.d. assumption is not required ([33], p.310, line 2-3). We should notice that in [33]

128

the i.i.d. assumption was first introduced on p.350, Section 7.5, and that from p.310 to p.350 the i.i.d. assumption is not required.

The minimum decision loss $\rho_j(\pi_{x_1^j})$ can be also written as

$$\rho_j(\pi_{x_1^j}) = \min\{w_{10}\pi_{x_1^j}, w_{01}(1 - \pi_{x_1^j})\}. \tag{4.2.28}$$

The above result provides the Bayesian terminal decision rule. Now we discuss the sequential stopping rule. Defining a posterior risk associated with the j-sample test as

$$U_j \triangleq \rho_j + cj \tag{4.2.29}$$

Eq. (4.2.24) can be written as

$$
\begin{aligned}
r(\pi, \psi, \delta^\pi) &= E(\rho_N + cN) \\
&= E\{\sum_{j=0}^{+\infty} \psi_j[\rho_j(X_1^j, \pi) + cj]\} \\
&= E\{\sum_{j=0}^{+\infty} \psi_j U_j(X_1^j, \pi)\} = E[U_N(X_1^N, \pi)]. \tag{4.2.30}
\end{aligned}
$$

The above equation shows that a greater sample size j corresponds to a smaller terminal decision loss ρ_j which is traded off by the cost cj due to the greater sample size. The goal is to determine an optimum j such that the above gross risk is minimized. To find this optimum j, we first truncate this sequential test at some finite J, find an optimum j, and then let J go to infinity to get the required result. Define the truncated test as

$$\psi_1^J + ... + \psi_J^J \triangleq 1 \tag{4.2.31}$$

This implies that we must stop before or at the observation J. For $0 < \pi < 1$, let

$$
V_j^J(x_1^j, \pi) \triangleq
\begin{cases}
\min\{U_0(\pi), E[U_1(X_1, \pi)], E[U_2(X_1^2, \pi)], ..., E[U_J(X_1^J, \pi)]\} & \text{if } j = 0 \\
\min\{U_j(x_1^j, \pi), E[U_{j+1}(x_1^j, X_{j+1}, \pi)|X_1^j = x_1^j], \\
\quad E[U_{j+2}(x_1^j, X_{j+1}^{j+2}, \pi)|X_1^j = x_1^j], ..., E[U_J(x_1^j, X_{j+1}^J, \pi)|X_1^j = x_1^j]\} & \text{if } 0 < j < J.
\end{cases}
$$
$$\tag{4.2.32}$$

It can be easily verified that the above definition is the same as that in [33], p.353 or pp.316- 317. It is obvious that $V_j^J(x_1^j, \pi)$ is the minimum expected posterior risk among

129

all tests after observing $X_1^j = x_1^j$ and with the sample size not less than j and not greater than J. Particularly, if $J \rightarrow +\infty$, $V_0^J(\pi)$ represents the minimum expected posterior risk among all stopping rules. It should be noted that $V_0^J(\pi)$ is independent of x and is dependent upon π only, while $V_j^J(x_1^j, \pi)$ depends upon both π and x_1^j.

We now state a result similar to Theorem 2 in [33], p.317, where a formal proof was not supplied:

LEMMA 4.3 *The Bayes decision rule for the sequential decision problem (defined in this paper) truncated at J is (ψ^π, δ^π), where δ^π is defined in Lemma 4.2 and ψ^π is given by $\phi^\pi = (\phi_0^\pi, \phi_1^\pi, ...)$:*

$$\psi_j^\pi = (1 - \phi_0^\pi)...(1 - \phi_{j-1}^\pi)\phi_j^\pi \tag{4.2.33}$$

where ϕ^π is defined by

$$\phi_j^\pi \triangleq \begin{cases} 1 & \text{if } U_j(x_1^j, \pi) \leq E\{V_{j+1}^J(x_1^j, X_{j+1}, \pi)|X_1^j = x_1^j\} \\ 0 & \text{if } U_j(x_1^j, \pi) > E\{V_{j+1}^J(x_1^j, X_{j+1}, \pi)|X_1^j = x_1^j\} \end{cases} \tag{4.2.34}$$

where $0 \leq j < J$ and $\phi_J^\pi \triangleq 1$. The result also holds as $J \rightarrow +\infty$. $\square$

A proof of Lemma 4.3 is presented later. The above functions have some interesting properties, which are listed below:

$$\begin{aligned} (\pi_{x_1^i})_{x_{j+1}^n} &\triangleq P(H_1|X_{j+1}^n = x_{j+1}^n, \pi_{x_1^i}) \\ &= \frac{\pi_{x_1^i} S_{j+1}^n(x_{j+1}^n)}{\pi_{x_1^i} S_{j+1}^n(x_{j+1}^n) + 1 - \pi_{x_1^i}} \\ &= \pi_{x_1^n}; \end{aligned} \tag{4.2.35}$$

$$\rho_j(x_{i+1}^{i+j}, \pi_{x_1^i}) = \rho_0(\pi_{x_1^{i+j}}) = \rho_{j-k}(x_{i+k+1}^{i+j}, \pi_{x_1^{i+k}}) \quad \text{for } -i \leq k \leq j; \tag{4.2.36}$$

$$U_j(x_1^j, \pi) = U_0(\pi_{x_1^j}) + cj; \tag{4.2.37}$$

$$V_j^J(x_{i+1}^{i+j}, \pi_{x_1^i}) = V_0^{J-j}(\pi_{x_1^{i+j}}) + cj; \tag{4.2.38}$$

$$V_0^{J+1}(\pi_{x_1^i}) = \min\{U_0, E\{V_0^J[(\pi_{x_1^i})_{x_{i+1}}|X_1^i = x_1^i]\} + c\} \tag{4.2.39}$$

130

When $J \to +\infty$ the above results are still the case. By using the related definitions, the proof of the above properties is straightforward and ignored here. We note that the property (4.2.35) only requires the independence assumption rather that the i.i.d. assumption used in [33]. Using the above properties it can be easily verified that the Bayes rule in Lemma 4.3 can be expressed as

$$
\phi_j^\pi = \begin{cases} 1 & \text{if } U_0(\pi_{x_1^j}) \le E\{V_0^{+\infty}[(\pi_{x_1^j})x_{j+1}|X_1^j = x_1^j]\} + c \\ 0 & \text{if } U_0(\pi_{x_1^j}) > E\{V_0^{+\infty}[(\pi_{x_1^j})x_{j+1}|X_1^j = x_1^j]\} + c \end{cases} \tag{4.2.40}
$$

where $0 \le j < +\infty$. Define

$$
W_j(\pi_{x_1^j}) \triangleq E\{V_0^{+\infty}[(\pi_{x_1^j})x_{j+1}]|X_1^j = x_1^j\} + c, \tag{4.2.41}
$$

which is a function of $\pi_{x_1^j}$ only, for given j. Since pdfs of X_j are varying with j, $W_j(\pi_{x_1^j})$ is also varying with j. If X_j is i.i.d., then $W_j(\pi_{x_1^j})$ is not varying with j, which becomes $V(\pi_{x_1^j})$ defined in [33]. $W_j(\pi_{x_1^j})$ is a key generalization to $V(\pi_{x_1^j})$ in [33], and results in the varying thresholds. From the above definition, the stopping rule in Lemma 4.3 is equivalent to

$$
\phi_j^\pi = \begin{cases} 1 & \text{if } U_0(\pi_{x_1^j}) \le W_j(\pi_{x_1^j}) \\ 0 & \text{if } U_0(\pi_{x_1^j}) > W_j(\pi_{x_1^j}) \end{cases} \tag{4.2.42}
$$

where $0 \le j < +\infty$. Letting $\pi_{x_1^j}$ take on value π', the properties of $W_j(\pi')$ are given in

LEMMA 4.4 *For* $0 \le \pi' \le 1$, $W_j(\pi')$ *is a continuous function of* π',

$$
W_j(\pi') \ge c, \tag{4.2.43}
$$

$$
\lim_{\pi' \to 0} W_j(\pi') = \lim_{\pi' \to 1} W_j(\pi') = c. \tag{4.2.44}
$$

In addition, when $U_0(\pi') \ge W_j(\pi')$, $W_j(\pi')$ *is concave.* $\square$

The proof of Lemma 4.4 can be found in [33], p. 362.

$U_0(\pi')$ and $W_j(\pi')$ are shown in Figure 4.9. Recalling the definition

$$
w \triangleq \frac{w_{01}}{w_{10} + w_{01}}, \tag{4.2.45}
$$

131

let π_{Lj} and π_{Uj} denote two solutions of the equation

$$U_0(\pi') = W_j(\pi'), \tag{4.2.46}$$

where $\pi_{Lj} \leq w \leq \pi_{Uj}$. If there exists no π_{Lj} or π_{Uj}, let $\pi_{Lj} = w = \pi_{Uj}$.

It should be noted that there exists at most one solution π_{Lj} and at most one solution π_{Uj} to the equation $U_0(\pi') = W_j(\pi')$. The argument is as follows: If there exists more than one π_{Lj}, say π_{Lj} and π'_{Lj} ($\pi_{Lj} \neq \pi'_{Lj}$ and both $\leq w$), then there exists some $\pi' \leq w$ such that $W_j(\pi')$ is convex when $U_0(\pi') \geq W_j(\pi')$, which contradicts Lemma 4.4. A similar argument holds for π_{Uj}.

From the decision rules defined previously, the above results imply

LEMMA 4.5 *The Bayes decision rule for the sequential decision problem defined in (4) is (ψ^π, δ^π), i.e. (ϕ^π, δ^π), where δ^π is given by*

$$\delta_j^\pi = \begin{cases} 0 & \text{if } \pi_{x_1^j} \leq \pi_{Lj} \\ no\ decision & \text{if } \pi_{Lj} < \pi_{x_1^j} < \pi_{Uj} \\ 1 & \text{if } \pi_{x_1^j} \geq \pi_{Uj} \end{cases} \tag{4.2.47}$$

and ϕ^π is given by

$$\phi_j^\pi = \begin{cases} 1 & \text{if } \pi_{x_1^j} \leq \pi_{Lj} \text{ or } \pi_{x_1^j} \geq \pi_{Uj} \\ 0 & \text{if } \pi_{Lj} < \pi_{x_1^j} < \pi_{Uj} \end{cases} \tag{4.2.48}$$

where $0 \leq j < +\infty$. Or equivalently

$$\delta_j^\pi = \begin{cases} 0 & \text{if } S_1^j \leq A_j \\ no\ decision & \text{if } A_j < S_1^j < B_j \\ 1 & \text{if } S_1^j \geq B_j \end{cases} \tag{4.2.49}$$

and ϕ^π is

$$\phi_j^\pi = \begin{cases} 1 & \text{if } S_1^j \leq A_j \text{ or } S_1^j \geq B_j \\ 0 & \text{if } A_j < S_1^j < B_j \end{cases} \tag{4.2.50}$$

132

for $0 \leq j < +\infty$, which is a $NSPRT(A_j, B_j)$ with

$$A_j = \frac{1-\pi}{\pi} \cdot \frac{\pi_{Lj}}{1 - \pi_{Lj}} \tag{4.2.51}$$

$$B_j = \frac{1-\pi}{\pi} \cdot \frac{\pi_{Uj}}{1 - \pi_{Uj}}. \tag{4.2.52}$$

$\square$

By using the Lemmas 4.2-4.4, the proof of Lemma 4.5 is straightforward.

133
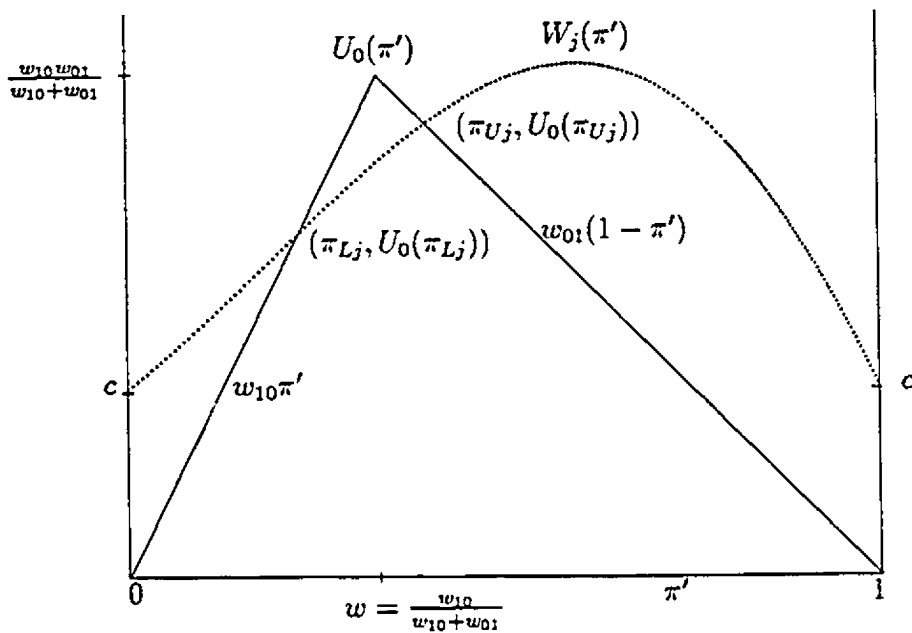
Figure 4.9: $U_0(\pi')$ is the Bayes cost of a FSS test and $W_j(\pi')$ is the minimum cost among all stopping rules. Both of them are functions of π'. One stops as soon as $U_0(\pi') \leq W_j(\pi')$.

134

So far we have showed that a Bayes decision rule is equivalent to a $NSPRT(A_j, B_j)$, which has the minimum gross risk, and can be used to prove

THEOREM 4.2 *Let N denote the sample size for a NSPRT and N_a denote the sample size for any alternative procedure. Let α and β denote the error probabilities of the first kind and the second kind, respectively, for a NSPRT and α_a, β_a the corresponding error probabilities for any alternative procedure. If*

$$\alpha_a \leq \alpha, \qquad \beta_a \leq \beta, \qquad (4.2.53)$$

then

$$E_1(N) \leq E_1(N_a), \qquad E_0(N) \leq E_0(N_a). \qquad (4.2.54)$$

On the other hand, if

$$E_1(N_a) \leq E_1(N), \qquad E_0(N_a) \leq E_0(N), \qquad (4.2.55)$$

then

$$\alpha \leq \alpha_a, \qquad \beta \leq \beta_a. \qquad (4.2.56)$$

$\square$

PROOF: Starting from Eq.(4.2.24), since the NSPRT rule N minimizes the total expected risk, then we have

$$E_\pi\{L(\pi, \delta_N)\} + cE_\pi\{N\} \leq E_\pi\{L(\pi, \delta_{N_a})\} + cE_\pi\{N_a\} \qquad (4.2.57)$$

Notice that

$$
\begin{aligned}
E_\pi\{L(\pi, \delta_N)\} &= E\{L(\pi, \delta_N)|H_1\}P\{H_1\} + E\{L(\pi, \delta_N)|H_0\}P\{H_0\} \\
&= w_{10}P\{a_0, NSPRT|H_1\}\pi + w_{01}P\{a_1, NSPRT|H_0\}(1 - \pi) \\
&= w_{10}\beta\pi + w_{01}\alpha(1 - \pi) \qquad (4.2.58)
\end{aligned}
$$

$$
\begin{aligned}
E_\pi\{N\} &= E\{N|H_1\}P\{H_1\} + E\{N|H_0\}P\{H_0\} \\
&= E_1\{N\}\pi + E_0\{N\}(1 - \pi) \qquad (4.2.59)
\end{aligned}
$$

$$E_\pi\{L(\pi, \delta_{N_a})\} = E\{L(\pi, \delta_{N_a})|H_1\}P\{H_1\} + E\{L(\pi, \delta_{N_a})|H_0\}P\{H_0\}$$

$$= w_{10}P\{a_0, Alternative|H_1\}\pi + w_{01}P\{a_1, Alternative|H_0\}(1 - \pi)$$

$$= w_{10}\beta_a\pi + w_{01}\alpha_a(1 - \pi) \tag{4.2.60}$$

and

$$E_\pi\{N_a\} = E\{N_a|H_1\}P\{H_1\} + E\{N_a|H_0\}P\{H_0\}$$

$$= E_1\{N_a\}\pi + E_0\{N_a\}(1 - \pi) \tag{4.2.61}$$

Summarizing the above equations yields that

$$w_{10}\beta\pi + w_{01}\alpha(1 - \pi) + cE_1\{N\}\pi + cE_0\{N\}(1 - \pi)$$

$$\leq w_{10}\beta_a\pi + w_{01}\alpha_a(1 - \pi) + cE_1\{N_a\}\pi + cE_0\{N_a\}(1 - \pi) \tag{4.2.62}$$

We finally conclude the theorem using the classical argument in [120, 33]. From (4.2.53) and (4.2.62), it is seen that

$$E_1(N)\pi + E_0(N)(1 - \pi) \leq E_1(N_a)\pi + E_0(N_a)(1 - \pi).$$

Since π is the prior probability that H_1 is true and can be made arbitrarily close to 0 or close to 1, then Eq. (4.2.54) is concluded from the above equations. **QED** .

PROOF OF LEMMA 4.3: To prove the stopping rule in Lemma 4.3 is Bayesian, we only need to show that it leads to the minimum Bayesian risk. Let

$$n = \min\{j|U_j(x_1^j, \pi) \leq E[V_{j+1}^J(x_1^j, X_{j+1}, \pi)|X_1^j = x_1^j]\} \tag{4.2.63}$$

then

$$\phi_j^\pi = \begin{cases} 0 & \text{if } j < n \\ 1 & \text{if } j = n \\ 0 \text{ or } 1 & \text{if } j > n \end{cases} \tag{4.2.64}$$

and

$$\psi_j^\pi = \begin{cases} 1 & \text{if } j = n \\ 0 & \text{if } j \neq n \end{cases} \tag{4.2.65}$$

136

for $0 \leq j \leq J$. Therefore

$$
\begin{aligned}
r^J(\pi, \psi^\tau, \delta^\tau) &= E\{\sum_{j=0}^{J} \psi_j^\tau U_j(X_1^j, \pi)\} = E\{U_n(X_1^n, \pi)\} = E\{\sum_{j=0}^{J} \psi_j U_n(X_1^n, \pi)\} \\
&= E\{\sum_{j=0}^{n-1} \psi_j U_n(X_1^n, \pi)\} + E\{\psi_n U_n(X_1^n, \pi)\} + E\{\sum_{j=n+1}^{J} \psi_j U_n(X_1^n, \pi)\},
\end{aligned}
$$

where the fact $\sum_{j=0}^{J} \psi_j = 1$ was used. For $j \geq n+1$ we have

$$
E\{U_n(X_1^n, \pi)\} \leq E\{E[V_{n+1}^J(X_1^n, X_{n+1}, \pi)|X_1^n]\} = E\{V_{n+1}^J(X_1^{n+1}, \pi)\} \leq E\{U_j(X_1^j, \pi)\}.
$$
$$(4.2.66)$$

For $0 \leq j \leq n-1$ we have

$$
\begin{aligned}
& E\{U_j(X_1^j, \pi)\} \\
> \ & E\{E[V_{j+1}^J(X_1^j, X_{j+1}, \pi)|X_1^j]\} \\
= \ & E\{V_{j+1}^J(X_1^{j+1}, \pi)\} \\
= \ & E[\min\{U_{j+1}(X_1^{j+1}, \pi), E[U_{j+2}(X_1^{j+2}, \pi)|X_1^{j+1}], E[U_{j+3}(X_1^{j+3}, \pi)|X_1^{j+1}], \\
& ..., E[U_J(X_1^J, \pi)|X_1^{j+1}]\}] \\
\geq \ & E[\min\{E[V_{j+2}^J(X_1^{j+2}, \pi)|X_1^{j+1}], E[U_{j+2}(X_1^{j+2}, \pi)|X_1^{j+1}], \\
& E[U_{j+3}(X_1^{j+3}, \pi)|X_1^{j+1}], ..., E[U_J(X_1^J, \pi)|X_1^{j+1}]\}] \\
= \ & E\{V_{j+2}^J(X_1^{j+2}, \pi)\} \\
\geq \ & E\{V_{j+3}^J(X_1^{j+3}, \pi)\} \\
\geq \ & ... \geq E\{V_n^J(X_1^n, \pi)\} \\
= \ & E\{U_n(X_1^n, \pi)\}.
\end{aligned}
$$
$$(4.2.67)$$

Summarizing the above equations yields

$$
\begin{aligned}
r^J(\pi, \psi^\tau, \delta^\tau) &= E\{\sum_{j=0}^{n-1} \psi_j U_n(X_1^n, \pi)\} + E\{\psi_n U_n(X_1^n, \pi)\} + E\{\sum_{j=n+1}^{J} \psi_j U_n(X_1^n, \pi)\} \\
&\leq E\{\sum_{j=0}^{J} \psi_j U_j(X_1^j, \pi)\} = r^J(\pi, \psi, \delta^\tau) \leq r^J(\pi, \psi, \delta).
\end{aligned}
$$
$$(4.2.68)$$

As we pointed out previously, the convergence results in [33] do not require the i.i.d. assumption. Then the proof is completed. **QED.**

137

IV. Properties of Optimal NSPRT Thresholds

We now discuss three properties of optimal NSPRT test thresholds, which can be applied to the NSPRT threshold design problem. We first present the results. The proofs are presented later.

The first property implies that if the signal parameters are periodically varying with time and the noise is stationary, then the thresholds for the NSPRT are also periodically varying:

PROPERTY 4.1 *If the pdfs $f_{1j}(x_j)$ and $f_{0j}(x_j)$ are periodically varying with j, then the thresholds A_j and B_j for the NSPRT are also periodically varying with j, i.e. $A_j = A_{j+T}$ and $B_j = B_{j+T}$ for all $j = 1, 2, ...,$ where T is the period.* $\Box$

The second property implies that if the signal in noise decays to zero at time instant, m, then the optimal hypothesis testing procedure is a truncated NSPRT with sample size not greater than m:

PROPERTY 4.2 *If there exists an $m(\geq 1)$ such that $l_j(x_j) = 1$ for all $j \geq m$, then the optimal hypothesis testing procedure becomes a truncated NSPRT with sample size not greater than m.* $\Box$

The third property implies that if the signal in stationary noise is monotonically increasing or decreasing, then the upper thresholds B_j are increasing or decreasing with j, respectively, under the conditions listed below:

PROPERTY 4.3 *(a). If the following conditions are satisfied: (1) $S_j^n(x) \geq 1$ implies $S_{j-i}^{n-i}(x) \geq 1$ for all $i, j, n = 1, 2, ...$ with $j \leq n$ and $i < j$, (2) $S_j^n(x) \leq 1$ implies $S_j^n(x) \leq S_{j-i}^{n-i}(x)$, and (3) $f_{0j}(x) = f_{0(j+1)}(x)$ for all $j = 1, 2, ...,$ then the upper thresholds B_j are increasing with j, i.e. $B_j \leq B_{j+1}$.*

(b). On the other hand, if the following conditions are satisfied: (1) $S_j^n(x) \geq 1$ implies $S_{j+i}^{n+i}(x) \geq 1$ for all $i, j = 1, 2, ...$ with $j \leq n$, (2) $S_j^n(x) \leq 1$ implies $S_j^n(x) \leq S_{j+i}^{n+i}(x)$, and

(3) $f_{0j}(x) = f_{0(j+1)}(x)$ *for all* $j = 1, 2, ...,$ *then the upper thresholds* B_j *are decreasing with* j, *i.e.* $B_j \geq B_{j+1}$. $\square$

For the case of a monotonically varying signal plus stationary noise, the conditions in Property 4.3 are generally satisfied. Notice that only the upper thresholds are monotonically varying. The properties of the lower thresholds are less straightforward. Since the lower thresholds are required for accepting f_{0j}, which is stationary, we conjecture that the lower thresholds may be constant.

PROOF OF PROPERTY 4.1: We only need to show that $\pi_{Lj} = \pi_{L\,j+T}$ and $\pi_{Uj} = \pi_{U\,j+T}$, in other words that $W_j(\pi') = W_{j+T}(\pi')$. Note that from (4.2.41) and (4.2.32),

$$
\begin{aligned}
W_j(\pi') &= E\{V_0^{+\infty}(\pi'_{X_{j+1}})\} + c \\
&= E[\min\{U_0(\pi'_{X_{j+1}}), E[U_1(X_{j+2}, \pi'_{X_{j+1}})|X_{j+1}], \\
&\quad E[U_2(X_{j+2}^{j+3}, \pi'_{X_{j+1}})|X_{j+1}], ...\}] + c
\end{aligned}
\tag{4.2.69}
$$

Now we show that

$$
W_j(\pi') = \int \min[G_0(x_{j+1}, \pi'), G_1(x_{j+1}, \pi'), G_2(x_{j+1}, \pi'), ..., G_i(x_{j+1}, \pi'), ...] f_{0(j+1)}(x_{j+1}) dx_{j+1} + c,
\tag{4.2.70}
$$

where

$$
G_0(x_{j+1}, \pi') \triangleq \min[w_{10}\pi' l_{j+1}(x_{j+1}), w_{01}(1 - \pi')]
\tag{4.2.71}
$$

$$
G_i(x_{j+1}, \pi') \triangleq \int \min[w_{10}\pi' S_{j+1}^{j+1+i}(x_{j+1}^{j+1+i}), w_{01}(1-\pi')] f_{0(j+2)}(x_{j+2})...f_{0(j+1+i)}(x_{j+1+i}) dx_{j+2}^{j+1+i}
$$
$$
+ic
\tag{4.2.72}
$$

for $i \geq 1$. Since

$$
E[U_i(X_{j+2}^{j+1+i}, \pi'_{x_{j+1}})] = \int U_i(x_{j+2}^{j+1+i}, \pi'_{x_{j+1}})[\pi'_{x_{j+1}} f_{1(j+2)}(x_{j+2})...f_{1(j+1+i)}(x_{j+1+i}) +
$$
$$
(1 - \pi'_{x_{j+1}}) f_{0(j+2)}(x_{j+2})...f_{0(j+1+i)}(x_{j+1+i})] dx_{j+2}^{j+1+i}
$$
$$
= \int \{\min[w_{10}(\pi'_{x_{j+1}})_{x_{j+2}^{j+1+i}}, w_{01}(1 - (\pi'_{x_{j+1}})_{x_{j+2}^{j+1+i}})] + ic\}[\pi'_{x_{j+1}} f_{1(j+2)}(x_{j+2})...f_{1(j+1+i)}(x_{j+1+i}) +
$$

$$(1 - \pi'_{x_{j+1}})f_{0(j+2)}(x_{j+2})...f_{0(j+1+i)}(x_{j+1+i})]dx_{j+2}^{j+1+i}$$

$$= \int \min[w_{10}(\pi'_{x_{j+1}})_{x_{j+2}^{j+1+i}}, w_{01}(1 - (\pi'_{x_{j+1}})_{x_{j+2}^{j+1+i}})][\pi'_{x_{j+1}}f_{1(j+2)}(x_{j+2})...f_{1(j+1+i)}(x_{j+1+i}) +$$

$$(1 - \pi'_{x_{j+1}})f_{0(j+2)}(x_{j+2})...f_{0(j+1+i)}(x_{j+1+i})]dx_{j+2}^{j+1+i} + ic$$

$$= \int \min[w_{10}\frac{\pi'_{x_{j+1}}S_{j+2}^{j+1+i}}{\pi'_{x_{j+1}}S_{j+2}^{j+1+i} + 1 - \pi'_{x_{j+1}}}, w_{01}\frac{1 - \pi'_{x_{j+1}}}{\pi'_{x_{j+1}}S_{j+2}^{j+1+i} + 1 - \pi'_{x_{j+1}}}]$$

$$[\pi'_{x_{j+1}}f_{1(j+2)}(x_{j+2})...f_{1(j+1+i)}(x_{j+1+i}) + (1 - \pi'_{x_{j+1}})f_{0(j+2)}(x_{j+2})...f_{0(j+1+i)}(x_{j+1+i})]dx_{j+2}^{j+1+i}$$

$$+ic$$

$$= \int \min[w_{10}\pi'_{x_{j+1}}S_{j+2}^{j+1+i}, w_{01}(1 - \pi'_{x_{j+1}})]f_{0(j+2)}(x_{j+2})...f_{0(j+1+i)}(x_{j+1+i})dx_{j+2}^{j+1+i} + ic$$

$$= \frac{1}{\pi'l_{j+1} + 1 - \pi'}G_i(x_{j+1}, \pi') \tag{4.2.73}$$

for $i \geq 0$, therefore

$$W_j(\pi') = \int \min[\frac{1}{\pi'l_{j+1} + 1 - \pi'}G_0(x_{j+1}, \pi'), \frac{1}{\pi'l_{j+1} + 1 - \pi'}G_1(x_{j+1}, \pi'),$$

$$..., \frac{1}{\pi'l_{j+1} + 1 - \pi'}G_i(x_{j+1}, \pi'), ...][\pi'f_{1(j+1)}(x_{j+1}) + (1 - \pi')f_{0(j+1)}(x_{j+1})]dx_{j+1}$$

$$= \int \min[G_0(x_{j+1}, \pi'), G_1(x_{j+1}, \pi'), ..., G_i(x_{j+1}, \pi'), ...]f_{0(j+1)}(x_{j+1})dx_{j+1} + c. \tag{4.2.74}$$

Since $f_{1j}(x) = f_{1(j+T)}(x)$ and $f_{0j}(x) = f_{0(j+T)}(x)$ for all j, it can be shown straight-forwardly that $G_i(x_{j+1}, \pi') = G_i(x_{j+1+T}, \pi')$. Substituting these values of G_i into $W_j(\pi')$ completes the proof. **QED.**

PROOF OF PROPERTY 4.2: Using results from proving Property 4.1, when $j \geq m$, $l_j(x_j) = 1$, and therefore

$$G_i(x_{m+1}, \pi_{x_1^m}) = \min[w_{10}\pi_{x_1^m}, w_{01}(1 - \pi_{x_1^m})] + ic, \tag{4.2.75}$$

which implies that

$$W_j(\pi_{x_1^m}) = \min[w_{10}\pi_{x_1^m}, w_{01}(1 - \pi_{x_1^m})] + c \tag{4.2.76}$$

Then

$$U_0(\pi_{x_1^m}) = \min[w_{10}\pi_{x_1^m}, w_{01}(1 - \pi_{x_1^m})] < W_j(\pi_{x_1^m}). \tag{4.2.77}$$

140

According to the stopping rule given by Equation (4.2.42), the optimal Bayesian stopping rule is to stop before or at m. The proof is complete. **QED.**

PROOF OF PROPERTY 4.3: We give the proof of part (a).

From (4.2.52), B_j is monotonically varying with π_{Uj}. Therefore it is enough to show that π_{Uj} is increasing with j. Lemma 4.4 and the definition of $U_0(\pi')$ reveal that it is sufficient to show that $W_j(\pi') \le W_{j-1}(\pi')$ for $\pi' \ge w$. Using the results in the proof of Property 4.1 and the assumption that $f_{0j}(x) = f_{0(j+1)}(x)$ for all $j = 1, 2, \dots$ reveals that we only need to show that $H_{i,j+1}(\pi') \le H_{ij}(\pi')$ for $\pi' \ge w$, where

$$H_{i,j+1}(\pi') \triangleq \min[w_{10}\pi' S_{j+1}^{j+1+i}(x), w_{01}(1 - \pi')] \qquad (4.2.78)$$

We consider two situations:

(1) When $S_{j+1}^{j+1+i}(x) \ge 1$, then $S_j^{j+i}(x) \ge 1$ according to the assumptions. Notice that if $\pi' \ge w = \frac{w_{01}}{w_{01}+w_{10}}$, then $w_{10}\pi' \ge w_{01}(1 - \pi')$. For this case we have

$$H_{i,j+1}(\pi') = w_{01}(1 - \pi') = H_{ij}(\pi'). \qquad (4.2.79)$$

(2) When $S_{j+1}^{j+1+i}(x) \le 1$, then $S_{j+1}^{j+1+i}(x) \le S_j^{j+i}(x)$ according to the assumptions, which implies that

$$H_{i,j+1}(\pi') \le H_{ij}(\pi'). \qquad (4.2.80)$$

Summarizing the above results yields the first part of Property 3. Part (b) can be proved by a similar argument. **QED.**

4.2.2   Change Detection for Non-i.i.d. Observations

I. Problem Statement

In this section we present a number of possible solutions to the change detection problem for non-i.i.d. observations. We are mainly concerned with two problems: (i) Suppose there exists a time-varying change in a sequence of i.i.d. observations at unknown time m, i.e. that the pdfs of $X_1, \dots, X_{m-1}, X_m, X_{m+1}, \dots$ are $f_0(x_1), \dots, f_0(x_{m-1}), f_1(x_m), f_2(x_{m+1})$,

141

..., respectively and all observations $X_1, ..., X_{m-1}, X_m, X_{m+1}, ...$ are independent. One is required to detect the change-time m as quickly as possible subject to a given false alarm rate. (ii) Suppose the observations are i.i.d. before a change-time m but dependent after m, i.e. that the pdfs of $X_1, ..., X_{m-1}$ are $f_0(x_1), ..., f_0(x_{m-1})$ respectively, and the joint pdf of $X_m, X_{m+1}, ..., X_n$ is $f_{n-m+1}(x_m, x_{m+1}, ..., x_n)$. One is required to detect the change-time m as quickly as possible subject to a given false alarm rate.

The above problems were studied in [17, 21]. As pointed out in [17], for non-i.i.d. situations Page's CUSUM statistic can not be written in the recursive form Eq. (2.2.3). Since both Page's CUSUM procedure and the MAR procedure depend upon the CUSUM statistic (recall that MAR statistic is a rational function of the CUSUM statistic), then directly substituting CUSUM statistic (2.2.3) for non-i.i.d. observations into the decision statistics for the CUSUM or MAR procedure may result in too much computational complexity for practical implementation. The solution to this problem is to approximate the CUSUM statistic in a reasonable way. Using a moving window FSS procedure to solve this problem is an alternative.

In [17], a modified version of the CUSUM procedure for detecting a time-varying change was proposed. Recall that the CUSUM procedure for i.i.d. observations is a repeated SPRT with lower threshold $A = 1$ and upper threshold $B > 1$. For time-varying and independent observations, a natural generalization of the CUSUM procedure is to modify CUSUM statistic as

$$\hat{T}_n = \max\{\hat{T}_{n-1}, 1\} l_{jn} \qquad (4.2.81)$$

for $0 \leq j \leq n$ and $n \geq 0$, where $\hat{T}_0 \stackrel{\triangle}{=} 1$,

$$l_{jn} \stackrel{\triangle}{=} \frac{f_{n-j}(x_n)}{f_0(x_n)}, \qquad (4.2.82)$$

and

$$j \stackrel{\triangle}{=} \sup_{1 \leq i \leq n} \{i | \hat{T}_{i-1} \leq 1\}. \qquad (4.2.83)$$

Then the modified CUSUM procedure (M-CUSUM procedure) is a repeated NSPRT. According to the properties of the optimum thresholds for the NSPRT given in the last

142

section and in [70], the lower threshold for the M-CUSUM procedure may be 1 and the upper decision thresholds should be time-varying and have certain properties shown in the last section and in [70].

A moving window Bayesian procedure was proposed in [21], which is quite similar to moving window FSS procedure but has extra computational complexity introduced to handle the time-varying problem. A similar structure, called moving window maximum likelihood ratio (MLR) procedure, will be presented in detail later.

II. Change Detection for Time-Varying and Independent Observations

Since the MAR statistic is a rational function of CUSUM statistic, the complexity of computing CUSUM statistic for time-varying and independent observations may result in difficulty of exactly implementing the MAR procedure for this non-i.i.d. case. A straightforward approach to circumvent this problem is to substitute the M-CUSUM statistic $\hat{T}_n$ for exact CUSUM statistic T_n in (3.2.1). The modified MAR (M-MAR) statistic is then written as

$$\hat{r}(n) = \frac{\hat{T}_n \pi_0 c_0 - (1 - \pi_0)}{\hat{T}_n \pi_0 + 1 - \pi_0} n \qquad (4.2.84)$$

and the modified MAR (M-MAR) procedure is that one stops as soon as

$$\hat{r}(n) > \hat{r}(n - 1) \qquad (4.2.85)$$

where $\hat{T}_n$ is given by (4.2.81)-(4.2.83). Obviously, the computational complexity of the M-MAR procedure is quite low and is comparable to the M-CUSUM procedure proposed in [17].

Combining the idea of moving window FSS procedure and moving window Bayesian procedure proposed in [21], we now propose a moving window maximum likelihood ratio (MLR) procedure to solve this problem for time-varying and independent observations. The fixed window size is w. Beginning from $n = 1$, $X_1, ..., X_w$ are first taken. An M-ary maximum likelihood ratio test is performed with hypotheses $H_0, H_1, ..., H_w$, respectively,

143

i.e., the maximum likelihood ratio is

$$S_{\hat{m}}^w = \max_{1 \le j \le w} \{S_j^w\} \qquad (4.2.86)$$

where

$$S_j^n \triangleq \prod_{i=j}^n \frac{f_{i-j+1}(x_i)}{f_0(x_i)}, \qquad (4.2.87)$$

if $S_{\hat{m}}^w$ is greater than a threshold d, then $H_{\hat{m}}$ is accepted and the procedure is terminated. If H_0 is accepted, one more sample X_{w+1} is taken and a similar M-ary maximum likelihood ratio test is performeded with hypotheses $H_0, H_2, ..., H_{w+1}$, respectively. Again, if $H_{\hat{m}}$ is accepted ($\hat{m} = 2, ..., w+1$), the procedure is terminated. Otherwise the procedure is repeated by taking X_{w+2} and performing an M-ary maximum likelihood ratio test over the new w samples with hypotheses $H_0, H_3, ..., H_{w+2}$, respectively. This procedure is repeated until $H_{\hat{m}}$ is first accepted ($\hat{m} \ge 1$). Generally, the decision rule is defined as follows:

$$S_{\hat{m}}^n = \max_{n-w+1 \le j \le n} \{S_j^n\} \qquad (4.2.88)$$

and

$$S_{\hat{m}}^n \begin{cases} \ge d & a_{\hat{m}} \\ < d & a_0 \end{cases} \qquad (4.2.89)$$

for all n.

Clearly, the computational complexity of this moving window MLR procedure is higher than both the M-CUSUM and M-MAR procedures.

III. Change Detection for Dependent Observations

For dependent observations, the complexity of computing CUSUM statistic is very high. To circumvent this problem, we propose a modified CUSUM statistic (M-CUSUM statistic) similar to the time-varying and independent situation. This M-CUSUM statistic is

$$\hat{T}_n = \max\{S_j^n, 1\} \qquad (4.2.90)$$

144

for $n \geq 0$, where $\hat{T}_0 = 1$,

$$S_j^n \triangleq \frac{f_{n-j+1}(x_j, x_{j+1}, ..., x_n)}{f_0(x_i) f_0(x_{i+1}) \cdots f_0(x_n)} \qquad (4.2.91)$$

and

$$j \triangleq \sup_{1 \leq i \leq n} \{i | \hat{T}_{i-1} = 1\}. \qquad (4.2.92)$$

Substituting the above M-CUSUM statistic into CUSUM or MAR procedures results in a modified CUSUM or MAR procedure for dependent observations with low computational complexity. The modified MAR (M-MAR) statistic is

$$\hat{r}(n) = \frac{\hat{T}_n \pi_0 c_0 - (1 - \pi_0)}{\hat{T}_n \pi_0 + 1 - \pi_0} n \qquad (4.2.93)$$

and the modified MAR (M-MAR) procedure is that one stops as soon as

$$\hat{r}(n) > \hat{r}(n - 1) \qquad (4.2.94)$$

where $\hat{T}_n$ is given by (4.2.90)-(4.2.92).

In addition, by following exactly the same approach described in the previous subsection, a moving window maximum likelihood ratio procedure can be also designed. But the latter approach has much higher computational complexity.

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
.0248	.026481	.031473	.037406	.044194	.052556	.0625	.074325	.088388	.105112

θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}	θ_{16}	θ_{17}	θ_{18}	θ_{19}	
.125	.148651	.176777	.210224	.25	.297302	.353553	.420448	.5	

Table 4.1: *Presumed parameters for the AMH and MII MAR procedures.*

SNR (db)	-20	-18.5	$-17.$	-15.5	$-14.$	-12.5	$-11.$
θ	.025	.029713	.035313	.04197	.049882	.059284	.07046
I_f	.005	.007063	.009976	.014092	.019906	.028117	.039717
I_{max}	.0049997	.0070379	.009941	.014052	.019848	.028034	.039597
$I_f - I_{max}$	$3. \times 10^{-7}$	.0000248	.000035	.0000396	.0000572	.0000827	.00012
SNR (db)	-9.5	$-8.$	-6.5	$-5.$	-3.5	$-2.$	-0.5
θ	.083741	.099527	.118288	.140585	.167086	.198582	.236015
I_f	.056100	.079245	.111936	.158114	.223342	.315479	.445625
I_{max}	.055928	.078995	.111576	.157593	.22259	.314394	.444061
$I_f - I_{max}$	.000173	.00025	.00036	.00052	.000751	.00108	.00156
SNR(db)	1.	2.5	4.0	5.5			
θ	.280505	.33338	.396223	.470912			
I_f	.629463	.88914	1.25594	1.77407			
I_{max}	.627205	.885884	1.25125	1.7673			
$I_f - I_{max}$	.00226	.00326	.00469	.00677			

Table 4.2: *Real parameters for the Gaussian distributions and the distance between the actual parameter and the presumed parameters $I_f - I_{max}$.*

146

k	1	2	3	4	5	6	7
π_{0k}	.09	.0216057	.00926628	.00504963	.00314101	.00212569	.00152538
c_{0k}	.1	.0240063	.0102959	.0056107	.00349001	.00236187	.00169486
k	8	9	10	11	12	13	14
π_{0k}	.00114288	.000885155	.00070381	.000571702	.00047269	.000396704	.000337207
c_{0k}	.00126986	.00098350G	.000782011	.000635225	.000525211	.000440782	.000374674
k	15	16	17	18	19		
π_{0k}	.000289811	.000251486	.00022008S	.000194065	.000172274		
c_{0k}	.000322012	.000279429	.000244542	.000215628	.000191415		

Table 4.3: *Design parameters for the MH MAR procedure.*

147

SNR (db)	−20	−18.5	−17.	−15.5	−14.	−12.5	−11.
θ	−.025	−.029713	−.035313	−.04197	−.049882	−.059284	−.07046
I_{max}	−.01484	−.01671	−.018933	−.021574	−.024713	−.028444	−.032879
$I_f - I_{max}$	.01984	.023773	.028909	.035666	.044619	.056561	.072595
D_{mMLE}	1192.5	1036.2	823.1	634.3	458.7	308.3	204.4
D_{mMH}	427.4	435.1	447.0	455.9	468.2	479.3	487.3
D_{mAMH}	3457.9	3835.6	4167.7	4628.8	5093.3		
N_{fMLE}	332.1	345.6	331.7	334.5	338.1	334.4	334.
N_{fMH}	383.4	385.8	383.4	384.8	383.0	384.3	387.3
N_{fAMH}	318.3	317.7	311.5	317.6	324.3		
SNR(db)	−9.5	−8.	−6.5	−5.			
θ	−.083741	−.099527	−.118288	−.140585			
I_{max}	−.038149	−.044413	−.051857	−.060705			
$I_f - I_{max}$	.09425	.123657	.163793	.218818			
D_{mMLE}	132.3	87.1	58.3	41.3			
D_{mMH}	497.1	503.0	508.6	511.7			
N_{fMLE}	332.3	330.4	334.6	335.3			
N_{fMH}	386.1	382.0	384.3	383.7			

Table 4.4: *Effect of the distance $I_f - I_{max}$ between the actual and presumed parameters on the performance of the MH and AMH MAR procedures. The presumed parameters are listed in Table 3.1, which are all greater than zero. The actual parameters are all less than zero. The performance of the MLE MAR procedure is also listed for comparison. The change time is $m = 500$. The false alarm probability $\alpha \approx 0.1$. The average delay for MLE, MH and AMH MAR procedures are denoted by D_{mMLE}, D_{mMH} and D_{mAMH} respectively. The false alarm ARL for MLE, MH and AMH MAR procedures are denoted by N_{fMLE}, N_{fMH} and N_{fAMH} respectively.*

148

Chapter 5

SIGNAL DETECTION OVER FADING CHANNELS

In Chapters 2-4, we have investigated the signal change detection problem. In Chapters 5 and 6, we apply change detection to the problem of fading channel identification encountered in mobile communications applications. In Chapter 5, we provide an overview of digital communications over a fading channel and motivate the need for improved on-line channel estimation algorithms. Chapter 6 then proposes a new approach to channel estimation based on combining decision-feedback and adaptive linear prediction (DFALP). In Section 6.3, it is demonstrated that the change detection algorithm proposed in Section 3.2 is able to improve DFALP fading channel identification in certain nonstationary situations.

In this chapter, we will first extend an optimum detection scheme of Lodge and Moher [76] to address a general class of fading channel problems. It is shown that the globally optimum detection scheme for constant-envelope modulation over frequency selective/nonselective fading channels is approximately equivalent to a combination of an MMSE (minimum-mean-square-error) channel gain estimator and a coherent signal detector. Motivated by this result, we will briefly review a number of existing fading channel tracking algorithms. The shortcomings of existing tracking algorithms lead us

149

to investigate more efficient approaches to this important problem. To further motivate our investigation into an advanced fading channel tracking algorithm, we also present a number of well-known results concerning differential phase shift keying (DPSK) and coherent PSK (CPSK) with perfect or imperfect channel state information (CSI). The results in this chapter justify that tracking a time varying fading channel is important for improving the performance of digital transmission over fading channels.

5.1 Fading Channel Problems

In this section, we introduce the fading channel problem. The well-known CPSK and DPSK modulation schemes for the fading channels are briefly described.

5.1.1 Frequency Nonselective Fading Channel

Let I_k denote a binary information sequence, and x_k the q-ary output of a low-pass equivalent discrete-time encoder/modulator. The complex signal x_k is transmitted over a frequency-nonselective Rayleigh or Rician fading channel. The received low-pass equivalent discrete-time signal is then [98]

$$y_k = x_k c_k + n_k \qquad (5.1.1)$$

where c_k is channel gain, a complex Gaussian process with memory. The mean of c_k is

$$a = E\{c_k\}. \qquad (5.1.2)$$

When $a = 0$, the fading channel is Rayleigh. Otherwise it is Rician. The covariance function of c_k is

$$r_{k,k-n} = r_n = E\{(c_k - a)(c_{k-n} - a)^*\}. \qquad (5.1.3)$$

A special case of the above model is the Jakes-Reudink fading channel [59, 99] with r_n given by

$$r_n = r_0 J_0(2\pi f_m nT) = r_0 \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j2\pi f_m nT \sin\theta} d\theta \qquad (5.1.4)$$

150

where $J_0()$ is the zeroth order Bessel function, T is the symbol period and f_m is the maximum Doppler frequency given by

$$f_m = \frac{v}{\lambda} \qquad (5.1.5)$$

with v and λ defined as mobile vehicle speed and transmission wavelength, respectively.

The power of the channel gain c_k can be divided into two parts: the line-of-sight (LOS) part a^2 and the random scattering part r_0. The well-known K factor is defined as the ratio [82]:

$$K \triangleq \frac{a^2}{r_0}. \qquad (5.1.6)$$

If r_0 is normalized to 1, then $K = a^2$. The K factor is equal to zero for Rayleigh fading channels and greater than zero for Rician fading channels. The signal-to-noise (SNR) ratio per symbol is then

$$\gamma_s = \frac{|a|^2 + r_0}{\sigma_n^2} \qquad (5.1.7)$$

where σ_n^2 is the variance of the additive white Gaussian noise (AWGN) n_k. The signal-to-noise (SNR) ratio per bit for q-ary constant-envelope modulation is

$$\gamma = \frac{|a|^2 + r_0}{\sigma_n^2 \log_2 q} \qquad (5.1.8)$$

The phase and amplitude of a typical Rayleigh fading channel are plotted in Figure 5.1, which is typical of a cellular mobile telephone channel. Figures 5.2 and 5.3 show the phase and amplitude of two Rician fading channels which are typical cellular mobile telephone and satellite channels. The data for plotting the figures are generated by a fading channel simulator developed in this thesis research using the language C++. The simulator passes a sequence of i.i.d. Gaussian random variables through a finite impulse response (FIR) filter whose frequency response is approximately the square-root of the Fourier transform of the zeroth order Bessel function.

From Figures 5.1 and 5.2 it is seen that under the same conditions, the phase of a Rayleigh fading channel changes more significantly than a Rician fading channel, and the

deep fade is more serious for a Rayleigh fading channel than a Rician fading channel. Therefore it appears more difficult to track a Rayleigh fading channel than a Rician channel. It is also observed that a deep fade often corresponds to a significant phase change, and may most likely result in burst errors.

Amplitude $|c_k|$ and Phase $\arg(c_k)$ $(\times\pi)$
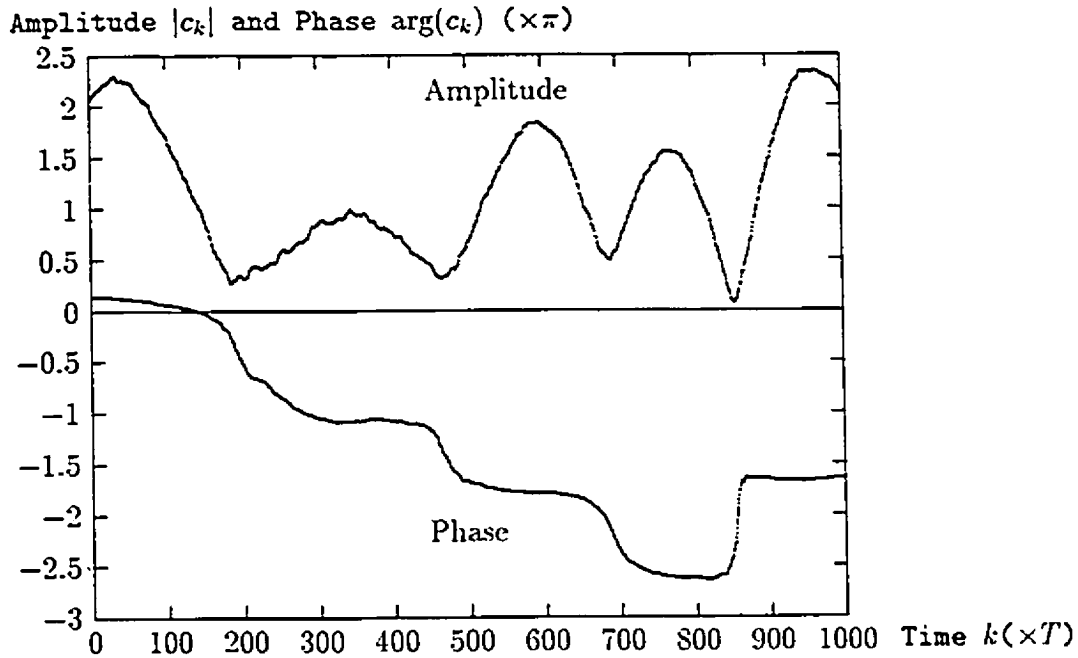
Figure 5.1: *The amplitude and phase of a Jakes-Reudink Rayleigh fading channel gain c_k for a typical cellular telephone channel. The phase equals the plotted value multiplied by π. The mobile speed is 100 km/hour. The symbol rate is 24000 symbols/second $(T = 4.17 \times 10^{-5}$ seconds). The wave-frequency is 800 mhz. The maximum Doppler frequency is $f_m = 74.07$ hz. The normalized fading bandwidth is $f_m T = 0.00309$.*

152

Amplitude $|c_k|$ and Phase $\arg(c_k)$ $(\times\pi)$
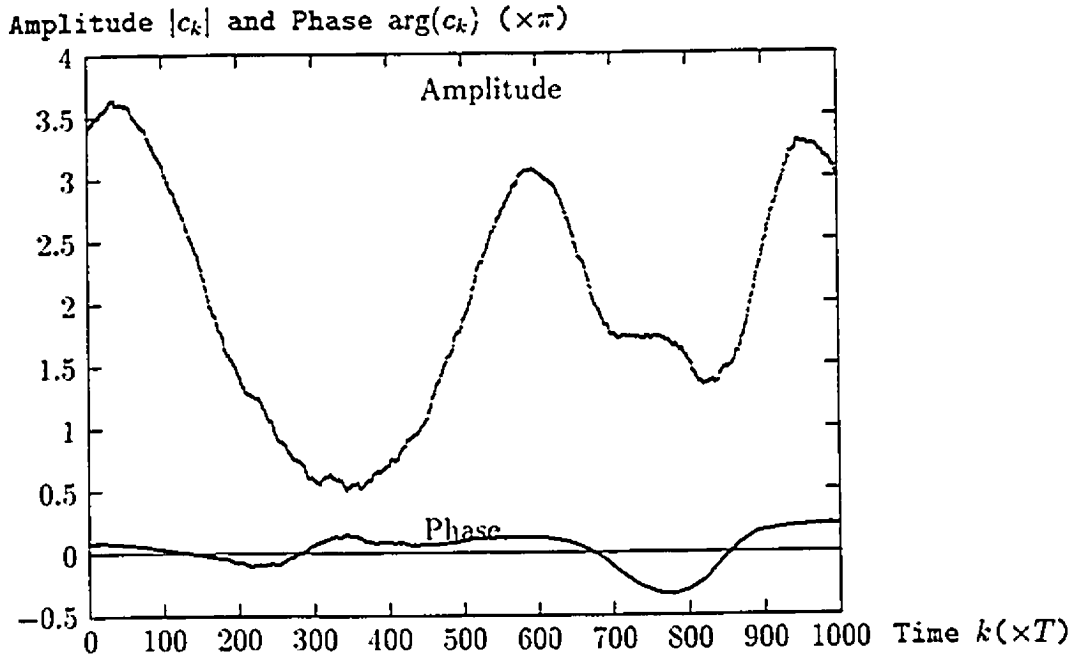
Figure 5.2: *The amplitude and phase of a Jakes-Reudink Rician fading channel gain c_k for a typical cellular telephone channel. The phase equals the plotted value multiplied by π. The mobile speed is 100 km/hour. The symbol rate is 24000 symbols/second $(T = 4.17 \times 10^{-5}$ seconds). The wave-frequency is 800 mhz. K factor equals 1. The maximum Doppler frequency is $f_m = 74.07$ hz. The normalized fading bandwidth is $f_m T = 0.00309$.*
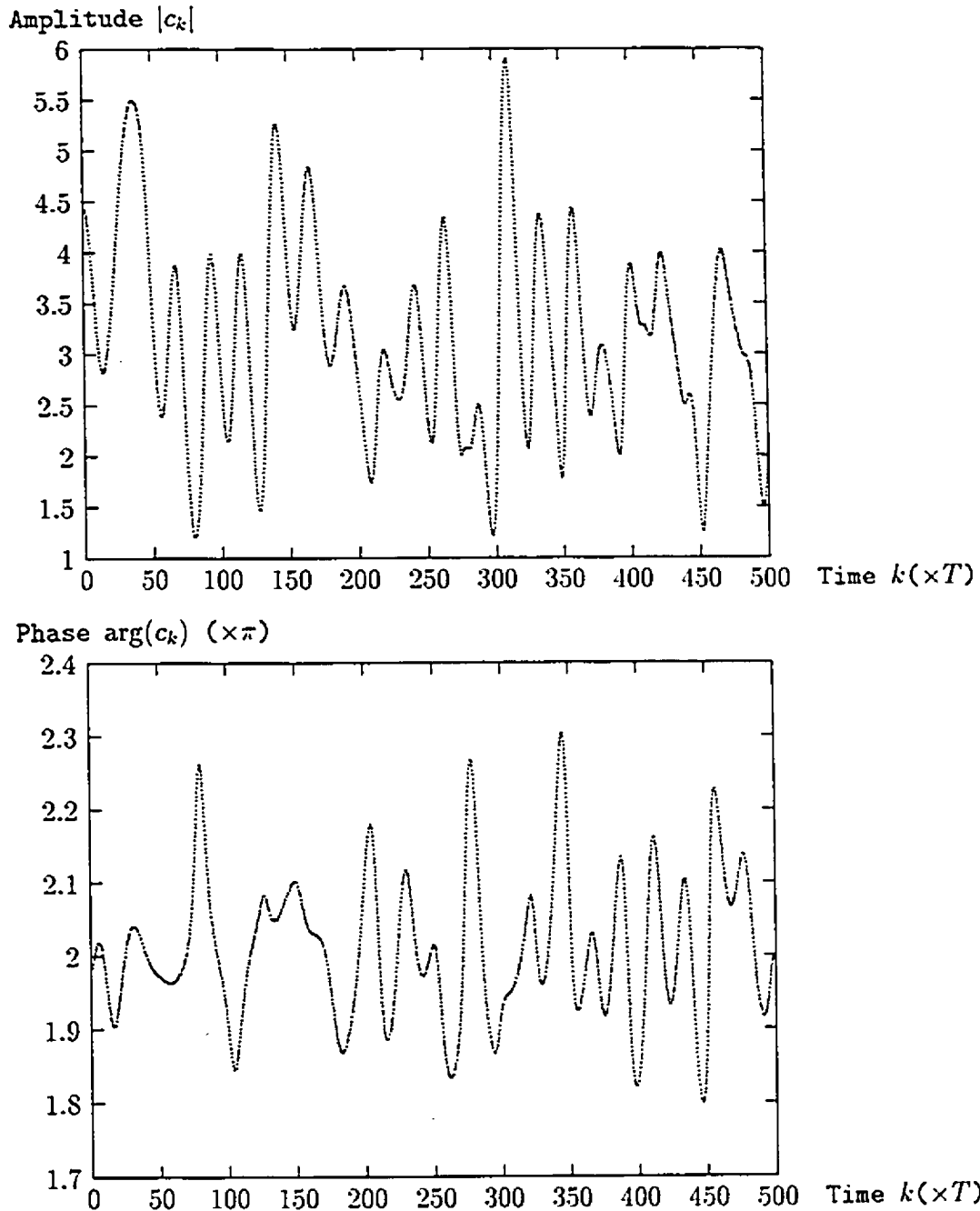
Figure 5.3: *The amplitude and phase of a Jakes-Reudink Rician fading channel gain c_k for a typical satellite channel. The phase equals the plotted value multiplied by π. The mobile speed is 60 km/hour. The symbol rate is 6000 symbols/second ($T = 1.67 \times 10^{-4}$ seconds). The wave-frequency is 5 ghz. K factor equals 4.5 (6.5 dB). The maximum Doppler frequency is $f_m = 277.78$ hz. The normalized fading bandwidth is $f_m T = 0.0462$.*

154

5.1.2 Frequency Selective Fading Channel

The frequency nonselective fading channel presented in the last section may be generalized to a class of frequency selective/nonselective fading channels as is depicted by Figure 5.4. This model is a generalization of the well-known one-beam and two-beam fading channels presented in [59, 62, 68, 76, 79]. The output y_k equals the convolution of the input modulated/encoded q-ary signal x_k with a time-varying discrete-time finite impulse response (FIR) filter $c_{i,k}$ plus additive white Gaussian noise (AWGN) n_k: [68, 98]

$$y_k = \sum_{i=0}^{l} x_{k-i} c_{i,k} + n_k \tag{5.1.9}$$

for $k = 1, 2, ...$ where $c_{i,k}$ is a complex zero-mean wide-sense stationary Gaussian process with memory. It should be noted that c_{ik} is the gain of an *equivalent beam* with a delay iT ($i = 0, ..., l$). By an equivalent beam we mean that it is the output of the actual beams passed through a linear (matched) filter plus a symbol rate sampler. In fact c_{ik} is a linear combination of a number of parameters, such as the actual inter-beam delay τ_i, the power-split-ratio between different real beams, etc.[68].

It can be seen that the cases $l = 0$ and 1 correspond to flat and two-beam fading channels, respectively [76, 68]. Notice that for a deterministic time-invariant channel the output is simply $y_k = \sum_{i=0}^{l} x_{k-i} c_i + n_k$, i.e., that the channel is modeled as a deterministic time-invariant FIR filter [98]. Therefore we also observe that the model (5.1.9) is a generalization of the conventional deterministic time-invariant channel model.

Consider a block of N received symbols $y_1, ..., y_N$. Let

$$\vec{y}_k \triangleq (y_1, ..., y_k)^T. \tag{5.1.10}$$

The mth transmitted sequence is

$$\vec{x}_N(m) = (x_1(m), ..., x_N(m))^T \tag{5.1.11}$$

where $m \in \{1, ..., q^N\}$. Defining

$$\vec{c}_{i,k} \triangleq (c_{i,1}, ..., c_{i,k})^T \tag{5.1.12}$$

155
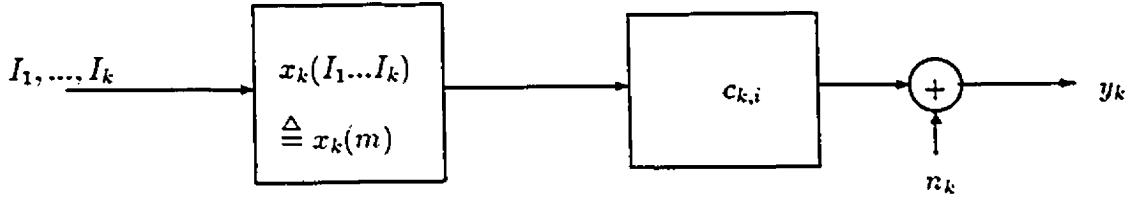
Figure 5.4: *An equivalent discrete-time model of frequency selective/nonselective Rayleigh fading channels*

and

$$\vec{n}_k \triangleq (n_1, ... n_k)^T \tag{5.1.13}$$

It is easily seen that

$$\vec{y}_N = \sum_{i=0}^{l} diag(\vec{x}_{N-i})\vec{c}_{i,N} + \vec{n}_N \tag{5.1.14}$$

where $diag()$ stands for transforming a vector into a diagonal matrix. It is noted that when $i' < 0$ for $x_{i'}$, the notation $x_{i'}$ denotes the $N - |i'|$th symbol transmitted from the last block.

The covariance matrix of the received signal $\vec{y}_N$ is

$$R_N(m) = cov\{\vec{y}_N\} = \sum_{i=0}^{l} diag(\vec{x}_{N-i})R_{c_i} diag(\vec{x}_{N-i})^* + R_\sigma \tag{5.1.15}$$

where $R_{c_i} = \{r_{c_i}(j,k)\}_{N \times N}$ is the covariance matrix of $\vec{c}_{i,k}$ and is Toeplitz, and

$$R_\sigma = diag(\sigma_n^2, ..., \sigma_n^2) \tag{5.1.16}$$

is the covariance matrix of the noise n_k.

For the Jakes-Reudink fading model, $r_{c_i}(j,k)$ is a linear function of Bessel function $J_0(2\pi f_d(j-k)T)$, inter-beam delay and power-split-ratio [22, 68], T is the symbol period, $f_d = \frac{v}{\lambda}$ is the maximum Doppler frequency, while v and λ are vehicle speed and wavelength, respectively.

Letting r_{jk} denote the (j,k)th element of $R_N(m)$, it is easily seen that

$$r_{jk}(m) = [x_j(m)x_k^*(m)r_{c_0}(j,k) + x_{j-1}(m)x_{k-1}^*(m)r_{c_1}(j,k) + ... + x_{j-l}(m)x_{k-l}^*(m)r_{c_l}(j,k)] + \sigma_n^2 \delta_{j,k} \tag{5.1.17}$$

156

for $1 \leq j, k \leq N$, where $\delta_{j,k}$ is the Kronecker delta function. For analytical purposes [76, 53], we assume that R_{c_i} and R_σ are known. We will drop this assumption after an MMSE-channel-estimator/coherent-detector structure is derived.

5.1.3   DPSK for Nonselective Fading Channels

From Figures 5.1 and 5.2, it is seen that the phase and amplitude of the fading channel gain c_k change significantly with time, particularly for Rayleigh fading channels. Without estimating the channel gain c_k reliably, one may have to use differential phase shift keying (DPSK) to transmit data [59, 62]. The DPSK transmits symbol x'_k which is equal to

$$x'_k = x_k x'_{k-1} \tag{5.1.18}$$

where x_k is a q-ary PSK modulated data symbol taking values on

$$D = \{ e^{2n\pi/q} : n = 0, 1, ..., q - 1 \}. \tag{5.1.19}$$

The received DPSK signal is then

$$y_k = c_k x'_k + n_k. \tag{5.1.20}$$

To detect the DPSK signal, one needs to formulate

$$z_k = y_k y^*_{k-1}. \tag{5.1.21}$$

Using Eqs.(5.1.18)-(5.1.21) yields that

$$
\begin{aligned}
z_k &= c_k c^*_{k-1} x'_k x'^*_{k-1} + c^*_{k-1} x'^*_{k-1} n_k + c_k x'_k n^*_{k-1} + n_k n^*_{k-1} \\
&= c_k c^*_{k-1} x_k x'_{k-1} x'^*_{k-1} + n'_k \\
&= c_k c^*_{k-1} x_k + n'_k
\end{aligned}
\tag{5.1.22}
$$

where

$$n'_k = c^*_{k-1} x'^*_{k-1} n_k + c_k x'_k n^*_{k-1} + n_k n^*_{k-1} \tag{5.1.23}$$

157

is an equivalent noise term. Using a minimum distance decision rule yields that the detected data symbol $\bar{x}_k$ is

$$|\arg(\bar{x}_k) - \arg(z_k)| = \min_{x_k \in D} |\arg(x_k) - \arg(z_k)| \tag{5.1.24}$$

where $\arg(x)$ stands for taking the phase of the complex number x.

Using (5.1.22) and (5.1.23), one can show [59] that the bit-error-rate (BER) of symbol-by-symbol DBPSK (differential binary PSK) signal detection for Rayleigh fading channels ($a = 0$) is

$$P_c = \frac{1}{2(\gamma + 1)} + \frac{\gamma(1 - r_1)}{2(\gamma + 1)} = P_1 + P_2 \tag{5.1.25}$$

where γ is the signal-to-noise ratio (SNR) per bit and r_1 is the correlation between c_k and c_{k-1} given by (5.1.3). It should be noted that the actual BER of the DPSK would be greater than the above equation since the equivalent noise term n_k' is dependent upon n_{k-1}' according to Eq.(5.1.23).

When the SNR γ approaches ∞, the BER becomes

$$P_I = P_c|_{\gamma - \infty} = P_2|_{\gamma - \infty} = \frac{1 - r_1}{2} \tag{5.1.26}$$

which is an irreducible BER.

For Jakes-Reudink Rayleigh fading channel, r_1 is given by (5.1.4), i.e.,

$$r_1 = J_0(2\pi f_m T). \tag{5.1.27}$$

For example, consider a channel with 800 mhz wave frequency, mobile speed 100 km/hour and symbol rate 24000 symbols/second. The normalized fading bandwidth is $f_m T$ 0.003086 and $r_1 = 0.999906$. The irreducible BER is then

$$P_I = 9.4 \times 10^{-5} \tag{5.1.28}$$

which may not be significant compared to the first part P_1 of P_c in (5.1.25). However, if the wave frequency is 30 ghz, as is the case for some satellite communications, then $r_1 = 0.8721$ and the irreducible BER is

$$P_I = 6.395 \times 10^{-2} \tag{5.1.29}$$

158

and may not be acceptable.

For differential quaternary PSK (DQPSK) and Differential q-ary PSK ($q > 4$), the BER performance becomes worse and worse for larger and larger q [98]. The derivation of the closed-form solution of BER for q-ary DPSK is involved.

5.1.4   Coherent PSK for Nonselective Fading Channels

It is well-known that the coherent PSK (CPSK) is superior to DPSK if the channel information is completely known [98]. For this modulation scheme, the transmitted symbols are q-ary PSK modulated symbols x_k taking values on

$$D = \{e^{2n\pi/q} : n = 0, 1, ..., q - 1\}.$$

The received signal is then

$$y_k = c_k x_k + n_k. \tag{5.1.30}$$

Suppose c_k is known, then x_k is estimated by

$$\hat{x}_k = \frac{y_k}{c_k} = x_k + n_k' \tag{5.1.31}$$

where n_k' is a transformed noise term. Then using the minimum distance decision rule yields that the detected symbol $\bar{x}_k$ satisfies

$$|\arg(\bar{x}_k) - \arg(\hat{x}_k)| = \min_{x_k \in D} |\arg(x_k) - \arg(\hat{x}_k)| \tag{5.1.32}$$

or equivalently

$$|\arg(c_k \bar{x}_k) - \arg(y_k)| = \min_{x_k \in D} |\arg(c_k x_k) - \arg(y_k)|. \tag{5.1.33}$$

The BER of symbol-by-symbol CPSK signal detection over a Rayleigh fading channel is [98]

$$P_e = \frac{1}{2}(1 - \sqrt{\frac{\gamma}{1+\gamma}}) \tag{5.1.34}$$

159

where γ is the SNR per bit. For large SNR the above equation is approximated by

$$P_e = \frac{1}{4(1+\gamma)}. \qquad (5.1.35)$$

Comparing (5.1.25) to (5.1.35), we can see that the CPSK scheme is over 3 dB better than DPSK and the gap is even greater between the CPSK and q–ary DPSK for larger q provided that the channel information c_k is completely known. It is also interesting to observe that the BER of CPSK does not have any irreducible error rate if the channel state information is reliable. It should be noted that the above BER is derived assuming independent channel gains. For a real channel the actual BER would be higher than the theoretical result.

When the channel gain c_k is estimated inaccurately, the BER performance of the CPSK scheme degenerates. This problem was analyzed in [46]. A particularly interesting result in [46] is an analysis of the BER of symbol-by-symbol CPSK signal detection for frequency nonselective fading channels with inaccurately estimated channel gains. The BER for BPSK and QPSK are

$$P_e = \frac{1}{2}(1 - \frac{1}{\sqrt{1 + \frac{\sigma_{\hat{c}}^2 \log_2 q + \gamma^{-1}}{1-\sigma_{\hat{c}}^2}}}) \qquad (5.1.36)$$

where $\sigma_{\hat{c}}^2$ is the mean-square-error of the estimator $\hat{c}_k$ of c_k, and $q = 2$ for BPSK and $q = 4$ for QPSK. For $q \geq 8$ the closed-form solution appears involved. When the SNR γ approaches infinity, an irreducible error rate results and is equal to

$$P_I = P_e|_{\gamma \to \infty} = \frac{1}{2}[1 - \sqrt{\frac{1 - \sigma_{\hat{c}}^2}{1 + (\log_2 q - 1)\sigma_{\hat{c}}^2}}] \qquad (5.1.37)$$

It is seen that if the estimator is accurate enough the irreducible BER P_I can be made very small.

160

5.2 Optimum and Suboptimum Algorithms for Detecting Constant-Envelope Signals over Fading Channels

In this section, we first introduce a globally optimum algorithm of Lodge and Moher [76] for detecting constant-envelope modulated signals over a fading channel. Then we will show that the optimum receiver is approximately equivalent to a combination of an MMSE channel estimator and a coherent signal detector. This result strengthens the report in [60] that an MMSE-estimator/coherent-detector structure is locally optimal in the sense that it approximately minimizes the symbol error probability. It should be noted that a locally optimum detector is not necessarily globally optimal when the detected signal is correlated [97]. The fading channel model studied in this thesis is correlated and therefore the well-known optimal signal detection property applies.

The result in this section implies that combining a good fading channel tracker with coherent signal detection is a promising alternative to the conventional differential detection scheme for fading channels and provides a motivation for finding a good fading channel tracking algorithm.

5.2.1 An Optimum Algorithm for Detecting Modulated Signals Over Fading Channels

In [76], Lodge and Moher proposed an optimum receiver which minimizes the error probability for detecting a *sequence* of constant-envelope modulated signals over frequency nonselective fading channels. The same optimum receiver is generalized in [74] to approach a class of frequency selective/nonselective fading channel problems with constant/non-constant envelope modulation schemes. We now present this result and show an interesting property of the optimum detection algorithm.

161

Following [76], an MLSE scheme chooses a data sequence $\vec{x}_N(m)$ which achieves

$$\max_{m \in \{1,...,q^N\}} p(y_1,....,y_N | \vec{x}_N(m)) \tag{5.2.1}$$

where q denotes that the data symbol x_k is q-ary modulated (for QPSK $q = 4$). Using the properties of conditional probability, it is easily shown that

$$p(y_1,...,y_N | \vec{x}_N(m)) = p_1(y_1 | \vec{x}_1(m)) p_2(y_2 | y_1, \vec{x}_2(m))...p_k(y_k | \vec{y}_{k-1}, \vec{x}_k(m))...p_N(y_N | \vec{y}_{N-1}, \vec{x}_N(m)) \tag{5.2.2}$$

Therefore maximizing Eq.(5.2.1) is equivalent to maximizing Eq.(5.2.2) over all $m \in \{1,...,q^N\}$. Notice that $p_k(y_k | \vec{y}_{k-1}, \vec{x}_k(m))$ is Gaussian with mean

$$\bar{y}_k(m) = E\{y_k | \vec{y}_{k-1}, \vec{x}_k(m)\} \tag{5.2.3}$$

and variance

$$\sigma_k^2(m) = Var\{y_k | \vec{y}_{k-1}, \vec{x}_k(m)\}. \tag{5.2.4}$$

Then Eq.(5.2.2) is equivalent to

$$\max_{m \in \{1,...,q^N\}} : f_1(y_1 | m)...f_k(y_k | m)...f_N(y_N | m) \tag{5.2.5}$$

where

$$f_k(y_k | m) = \frac{1}{\sqrt{2\pi\sigma_k^2(m)}} e^{-[y_k - \bar{y}_k(m)]^2 / [2\sigma_k^2(m)]}. \tag{5.2.6}$$

It should be noted that under the density f_k the random variables y_k are whitened and form a statistically independent sequence.

Using the properties of conditional mean and conditional variance of a multivariate Gaussian distribution ([97], p.216), it follows that

$$\bar{y}_k(m) = [\vec{r}_k(m)]^T [R_{k-1}(m)]^{-1} \vec{y}_{k-1} \tag{5.2.7}$$

and

$$\sigma_k^2(m) = r_{kk}(m) - [\vec{r}_k(m)]^T [R_{k-1}(m)]^{-1} \vec{r}_k(m). \tag{5.2.8}$$

where $\vec{r}_k(m) \triangleq (r_{k,1}(m), r_{k,2}(m), ..., r_{k,k-1}(m))^T$ and $k = 1,...,N$.

162

Theoretically, the above quantities may be computed using the Viterbi algorithm, an effective solution to a maximum likelihood sequence estimation (MLSE) scheme [76, 74]. Unfortunately, its large computational complexity makes it very difficult to implement. However, the MLSE scheme has some interesting theoretical properties and a practical MMSE-estimator/coherent-detector structure can be derived from it.

In [76], it is pointed out that the optimum MLSE scheme for a flat Rayleigh fading channel and constant-envelope modulation does not exhibit an irreducible error rate if (i) the SNR approaches infinity, (ii) the prediction order N is sufficiently large, and (iii) the fading channel is bandlimited. In this section, we arrive at a similar conclusion for the generalized receiver discussed earlier, but without requiring condition (i). The following result also suggests that condition (i) may not be necessary provided a suitable modulation scheme is found.

LEMMA 5.1 *If the following conditions are satisfied for all $k \geq 1$ and $m, m_0 \in \{1, 2, ..., M\}$:*

C1: $\bar{y}_k(m) \neq \bar{y}_k(m_0)$ or $\sigma_k^2(m) \neq \sigma_k^2(m_0)$ for all $m \neq m_0$,

C2: $\bar{y}_k(m)$ and $\sigma_k^2(m)$ are both finite,

where m_0 denotes the transmitted information sequence and M is a finite integer, then the error probability of maximum likelihood sequence estimation scheme (5.2.1) or (5.2.5) approaches zero as $N \to +\infty$, i.e.

$$p_e(N) = P\{f_1(y_1|m_0)...f_k(y_k|m_0)...f_N(y_N|m_0) \leq f_1(y_1|m)...f_k(y_k|m)...f_N(y_N|m)\} \to 0$$

$$as\ N \to +\infty \qquad (5.2.9)$$

PROOF: Since f_k is Gaussian with mean $\bar{y}_k(m)$ and variance $\sigma_k^2(m)$, the condition C1 assures that

$$f_k(y_k|m_0) \neq f_k(y_k|m) \qquad (5.2.10)$$

163

for all $m \neq m_0$ and $1 \leq k \leq N$.

Notice that inequality (5.2.9) is equivalent to

$$P\{[\ln \frac{f_1(y_1|m)}{f_1(y_1|m_0)} + \ln \frac{f_2(y_2|m)}{f_2(y_2|m_0)} + ... + \ln \frac{f_N(y_N|m)}{f_N(y_N|m_0)}]/N < 0\} = 1$$

(5.2.11)

as $N \to +\infty$. Under the density $f_k(y_k|m_0)$ (k=1,2,...), the log-likelihood ratios $\eta_k \triangleq \ln \frac{f_k(y_k|m)}{f_k(y_k|m_0)}$ are independent for all $k \geq 1$. Using the condition C2 it can be shown that the variance of η_k, $Var_{km_0}(\eta_k)$, is finite for all $k \geq 1$, where $Var_{km_0}()$ is defined under the density $f_k(y_k|m_0)$. The proof that $Var_{km_0}(\eta_k)$ is finite is presented later. According to the strong law of large numbers (see e.g. [112], p.364, Theorem 2), the term

$$[\ln \frac{f_1(y_1|m)}{f_1(y_1|m_0)} + \ln \frac{f_2(y_2|m)}{f_2(y_2|m_0)} + ... + \ln \frac{f_N(y_N|m)}{f_N(y_N|m_0)}]/N$$

in Eq.(5.2.11) almost-surely approaches

$$E_0\{\ln \frac{f_1(y_1|m)}{f_1(y_1|m_0)} + \ln \frac{f_2(y_2|m)}{f_2(y_2|m_0)} + ... + \ln \frac{f_N(y_N|m)}{f_N(y_N|m_0)}\}/N$$

(5.2.12)

where E_0 denotes taking expectation under the true hypothesis m_0. Using Jensen's inequality, it is easily shown that

$$E_0\{\ln \frac{f_k(y_k|m)}{f_k(y_k|m_0)}\} < \ln E_0\{\frac{f_k(y_k|m)}{f_k(y_k|m_0)}\}$$
$$= \ln \int_{-\infty}^{+\infty} \frac{f_k(y_k|m)}{f_k(y_k|m_0)} \cdot f_k(y_k|m_0)dt = 0$$

(5.2.13)

for all $1 \leq k \leq N$, where (5.2.10) is used and implies that a strict inequality holds in the above derivation. Therefore the inequality

$$E_0\{\ln \frac{f_1(y_1|m)}{f_1(y_1|m_0)} + \ln \frac{f_2(y_2|m)}{f_2(y_2|m_0)} + ... + \ln \frac{f_N(y_N|m)}{f_N(y_N|m_0)}\}/N < 0$$

(5.2.14)

holds almost-surely, i.e., (5.2.11) holds as $N \to +\infty$, which is equivalent to (5.2.9).

To complete the proof, we now show that under the density $f_k(y_k|m_0)$ ($k \geq 1$), the variance of $\eta_k = \ln \frac{f_k(y_k|m)}{f_k(y_k|m_0)}$ is finite provided that the condition C2 is satisfied.

164

It is easily seen that

$$\eta_k = \ln \frac{\sigma_k(m_0)}{\sigma_k(m)} + \frac{(y_k - \bar{y}_k(m_0))^2}{2\sigma_k^2(m_0)} - \frac{(y_k - \bar{y}_k(m))^2}{2\sigma_k^2(m)} \tag{5.2.15}$$

Under the density $f_k(y_k|m_0)$, the variance of η_k is

$$Var_{km_0}(\eta_k) = \frac{Var_{km_0}(y_k^2) + 4\bar{y}_k^2(m_0)Var_{km_0}(y_k)}{4\sigma_k^4(m_0)} + \frac{Var_{km_0}(y_k^2) + 4\bar{y}_k^2(m)Var_{km_0}(y_k)}{4\sigma_k^4(m)} \tag{5.2.16}$$

Since $Var_{km_0}(y_k) = \sigma_k^2(m_0)$ and

$$
\begin{aligned}
Var_{km_0}(y_k^2) &= E_{km_0}(y_k^4) - E_{km_0}^2(y_k^2) \\
&= [\bar{y}_k^4(m_0) + 6\bar{y}_k^2(m_0)\sigma_k^2(m_0) + 3\sigma_k^4(m_0)] - [\bar{y}_k^2(m_0) + \sigma_k^2(m_0)]^2 \\
&= 4\bar{y}_k^2(m_0)\sigma_k^2(m_0) + 2\sigma_k^4(m_0)
\end{aligned}
$$

therefore

$$Var_{km_0}(\eta_k) = \frac{1}{2} + \frac{2\bar{y}_k^2(m_0)}{\sigma_k^2(m_0)} + \frac{[2\bar{y}_k^2(m_0) + 2\bar{y}_k^2(m) + \sigma_k^2(m_0)]\sigma_k^2(m_0)}{2\sigma_k^4(m)} \tag{5.2.17}$$

The finiteness of $Var_{km_0}(\eta_k)$ is easily seen from the above equation and the condition C2.

QED .

We should note that the condition C2 can easily be satisfied in general. The condition C1 has several interesting implications that require further elaboration:

Remark 1. Condition C1 does not require the SNR to approach $+\infty$. On the other hand, Theorem 3 in [76] requires an infinitely-large SNR.

Remark 2. It is commonly assumed that the random fading channel is bandlimited, or equivalently, that the fading channel has infinitely long memory [59, 76]. In this case, it is possible to find a signal constellation sufficient for C1 to hold. Let us consider any two sequences, $\vec{x}_N(m_0)$ and $\vec{x}_N(m)$ with only *one* different symbol at time $k = 1$,

165

$x_1(m_0) \neq x_1(m)$. Therefore for all $k > 1$, we have $x_k(m)x_1^*(m) \neq x_k(m_0)x_1^*(m_0)$ since $x_1(m_0) \neq x_1(m)$ and $x_k(m_0) = x_k(m)$. From (5.1.17) and (5.2.7)(5.2.8) it is seen that $\vec{r}_k(m) \neq \vec{r}_k(m_0)$ for all $k > 1$, which in turn implies that C1 is satisfied. As a result, for bandlimited channels sequence estimation with infinitesimal error rate is achievable if an infinitely long delay and infinitely large complexity can be tolerated, even though the SNR does not approach $+\infty$.

5.2.2 Combined MMSE-Estimator/Coherent-Detector for Fading Channels

We now show that the optimum MLSE receiver studied in the last subsection is approximately equivalent to a combination of MMSE-estimator and coherent-detector for constant-envelope modulations.

Starting with Eq. (5.2.3), (5.2.4) and (5.1.9), it is easily seen that

$$
\begin{aligned}
\bar{y}_k(m) &= E\{y_k|\vec{y}_{k-1}, \vec{x}_k(m)\} \\
&= E\{\sum_{i=0}^{l} c_{i,k}x_{k-i} + n_k|\vec{y}_{k-1}, \vec{x}_k(m)\} \\
&= \sum_{i=0}^{l} E\{c_{i,k}|\vec{y}_{k-1}, \vec{x}_k(m)\}x_{k-i}(m) \\
&= \sum_{i=0}^{l} \hat{c}_{i,k}x_{k-i}(m)
\end{aligned}
\tag{5.2.18}
$$

where

$$
\hat{c}_{i,k} = E\{c_{i,k}|\vec{y}_{k-1}, \vec{x}_k(m)\}.
\tag{5.2.19}
$$

Since $c_{i,k}$ is a Gaussian process, the conditional mean $\hat{c}_{i,k}$ is identical to an MMSE estimator [97]. Similarly, it is seen that the conditional variance

$$
\begin{aligned}
\sigma_k^2(m) &= Var\{y_k|\vec{y}_{k-1}, \vec{x}_k(m)\} \\
&= Var\{\sum_{i=0}^{l} c_{i,k}x_{k-i} + n_k|\vec{y}_{k-1}, \vec{x}_k(m)\}
\end{aligned}
$$

166

$$= \sum_{i=0}^{l} \sigma_{i,k}^2 + \sigma_n^2 \qquad (5.2.20)$$

where the fact that $x_k x_k^* = 1$ is used and

$$\sigma_{i,k}^2 = Var\{c_{i,k}|\vec{y}_{k-1}, \vec{x}_k(m)\} \qquad (5.2.21)$$

is the conditional variance of $c_{i,k}$, which is identical to the variance of the MMSE estimator $\hat{c}_{i,k}$. It is noted that we assumed that $c_{i,k}$ is independent of $c_{j,k}$ for $i \neq j$ (independence between different beams).

Notice that the optimum MLSE algorithm is a growing sample-size estimation-detection procedure, and is difficult to implement. An approximation of the optimum algorithm is to use a moving-window (with window-size N) estimation-detection scheme. When the window size N and SNR are both large, the variance of the moving-window MMSE estimator $\hat{c}_{i,k}$ may be independent of the data sequence $x_k(m)$. In this case, the optimum detector (5.2.5) is approximately equivalent to searching for $\bar{x}_k$ such that

$$\sum_{k=1,2,\dots} |y_k - \sum_{i=0}^{l} \hat{c}_{i,k} \bar{x}_{k-i}|^2 = \min_{m \in D} \sum_{k=1,2,\dots} |y_k - \sum_{i=0}^{l} \hat{c}_{i,k} x_{k-i}(m)|^2 \qquad (5.2.22)$$

which is exactly a combined MMSE-estimator/coherent-detector and can be easily implemented.

The above result justifies that combining a good channel estimator with coherent signal detection is a promising alternative to the conventional DPSK scheme for fading channels.

5.3 Existing Algorithms for Tracking Fading Channels

5.3.1 Phase Lock Loop

The classical algorithm for tracking the phase of a communication channel is the phase lock loop (PLL) [98]. A PLL consists of a phase error detector followed by a voltage

167

controlled oscillator (VCO).

The major problems of PLL are false lock (i.e. handup) and slow convergence. A false lock occurs when the phase error is close to $\pm\pi$ [37]. During a false lock, the output voltage of the PLL is very small and the PLL remains in the incorrect state for a long time. It is difficult to predict the false lock and drive the PLL out of the undesirable state [37, 46]. The false lock problem is particularly disastrous in a fading environment where the phase of the channel varies fast during deep fade (see Figures 5.1-5.3). Another major problem of PLL is that it does not converge to the correct state fast enough. This is undesirable in a fading environment since the phase of the channel varies significantly during a deep fade. In addition, a PLL does not provide any information about the amplitude of the channel, which has been proven [57] to be able to provide as much as 1 dB gain using a soft decision Viterbi algorithm.

5.3.2 Kalman Filtering Approach

A number of researchers have proposed using a Kalman filter to track fading channels [55, 25, 46]. The basic idea is to formulate an auto-regressive moving average (ARMA) model, where perfect detection of the past data symbols is assumed[55, 25].

However, as is pointed out in [25, 26, 55], the currently used Kalman filtering approach has some serious problems: (i) a fading channel cannot be modeled as a low order ARMA model, and as a result, the computational complexity must be very high; (ii) Kalman filtering approaches are very sensitive to computer roundoff errors which may result in complete failure of the filter; and (iii) it is difficult to determine the Kalman filter coefficients and adjust these coefficients according to varying channel statistics.

It is well known that many fading channels can be modeled as a Jakes-Reudink model [59, 99], where the correlation function of the channel gain $c_{i,k}$ is a zeroth order Bessel function. Since the resulting model is not a rational function, a low order ARMA process may not be able to model the fading channel properly [25, 26]. This problem has also

168

been observed in our simulations. In order to model a fading channel properly, a high order model is needed, which in turn results in very high computational complexity since the complexity of Kalman filtering is proportional to the square of filter order.

The computer roundoff error and stability problems of the Kalman filtering approaches have been reported in [55, 25, 26]. In [55], a square-root Kalman filtering approach is used to circumvent this problem. However, the result in [55] only applies to a very slow fading model where the maximum Doppler frequency is $f_m = 1$ hz, which is much less than the current cellular and satellite channels (see Figures 5.1-5.3).

The Kalman filter coefficients are determined by the fading channel statistics, which have to be measured accurately. This on-line measurement brings about significantly more computational complexity. When the channel experiences changes of channel statistics, such as vehicle speed and random channel distributions (Rayleigh and Rician), it appears difficult to adjust the Kalman filter coefficients efficiently.

Due to the above reasons, the Kalman filtering approaches may have difficulties to solve the cellular and satellite fading channel problems.

5.3.3 Irvine-McLane's Fading channel Tracking Algorithm

An interesting algorithm was proposed by Irvine and McLane [57, 58] to track the phase and amplitude of frequency nonselective fading channels. This algorithm combines Moher-Lodge's [82] periodic re-training approach with the information extracted from a detected data symbol, where one training symbol is sent periodically after transmitting every $K_t - 1$ data symbols. The channel tracker in [57] is called symbol-aided plus decision-directed (SADD) algorithm.

A block diagram of the SADD algorithm is shown in [57], Figure 7. The channel gain estimate derived from the training symbols is first passed through a finite impulse response (FIR) low pass filter (LPF) to remove the noise. Since the training symbols are spaced K_t symbols (baud) apart, an order $2D_f + 1$ LPF results in a delay of $K_t D_f$ baud

(see [57], p.1292). Then the smoothed channel estimates are linearly interpolated (see [57], p.1292), which form an initial estimate of the channel gain. The linear interpolation scheme is shown in Figure 5.5. The initial channel estimate is used to make an initial decision on the received data symbols, which are in turn used to derive a decision-feedback channel estimate. A thresholding approach is then used to decide whether the decision-feedback estimate or interpolated estimate from the training symbols should be used (see [57], p.1293). Finally, the estimated channel gain is passed through the second LPF to reduce the noise. A smoothed channel estimate is then obtained for soft Viterbi decoding. The computational complexity of the algorithm is low compared to the Kalman filtering approach but is higher than the DPSK algorithm.

It has been shown by Irvine and McLane [57] that combining trellis-coded modulation (TCM), interleaving and soft Viterbi decoding with their fading channel tracker results in a 3 dB BER performance gain of CPSK signal detection over the same coded DPSK scheme, even though their tracking algorithm without coding does not achieve such a gain over its DPSK counterpart.

The major disadvantages of Irvine-McLane's algorithm are (i) extra decision delay is added due to the two LPFs used; (ii) more complexity is introduced than its DPSK counterpart, and (iii) more bandwidth is required to accommodate the training symbols.

In [57], it is pointed out that if an order $2D_f + 1$ LPF is used, then the total delay of the first LPF is $K_t D_f$ since the training symbols are spaced K_t symbols apart. The second LPF used in [57] also results in a delay of D_f. Then the total delay of the SADD algorithm is $(K_t+1)D_f$. Typically, assuming a training period $K_t = 5$ and an LPF order $2D_f + 1 = 21$ [57], the total delay would be 60 symbols, which may not be practical for certain applications, particularly for voice communications. The complexity of SADD involves two LPFs, interpolation processing and threshold switching. The bandwidth required by the SADD is $K_t/(K_t - 1)$ times that of DPSK detection. When the training period is $K_t = 5$ the bandwidth expansion of the SADD is 25%.

It is noted that [57] in spite of the extra complexity, decision delay and bandwidth used, the uncoded SADD algorithm with CPSK detection does not always outperform its DPSK counterpart. This problem is particularly serious for large SNRs (see simulation results and theoretical analysis in the next chapter). From Figure 5.5 it can be seen that the estimation error of the SADD sometimes may be very large. In Figure 5.5 the training symbols may be available at the time k_1, $k_1 - K_{t1}$ or k_2 and $k_2 - K_{t2}$ with K_{t1} or K_{t2} being the training period. A linear interpolation of the SADD algorithm is implemented along the lines drawn in the figure. It is seen that during the period between $k_2 - K_{t2}$ and k_2 the interpolation error is very large due to the large variation of the channel phase, while the error is not very serious during the period between $k_1 - K_{t1}$ and k_1. To circumvent this large error, the training period has to be very small and results in a large bandwidth expansion, which is not desirable in high speed data transmission. A more precise analysis of the MSE of the SADD presented in the next chapter will also confirm this observation. This fact leads us to further investigate the channel tracking problem and hopefully discover a better approach to solving this problem.
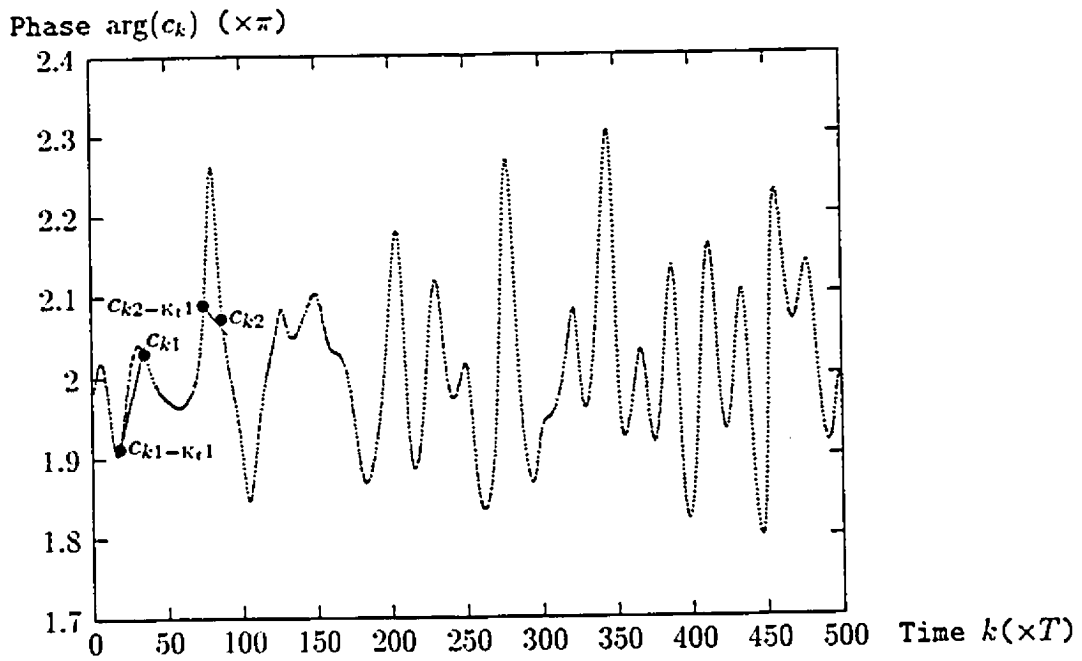
Figure 5.5: Linear interpolation of the SADD algorithm.

172

Chapter 6

DFALP: a New Algorithm for Adaptive Identification of Fading Channels

From the last chapter we conclude that developing a good fading channel tracker is both important and feasible. In this chapter we will follow the idea of combining periodical re-training with decision-feedback proposed by Irvine and McLane [57], which is an important improvement of the symbol aided (SA) tracking algorithm proposed by Moher and Lodge [82]. As is observed in Chapter 5, low-pass-filtering and interpolating large-spaced training-symbols result in the large delay and the large interpolation error involved with the SADD. To circumvent these drawbacks, the new algorithm uses N tentatively detected data symbols as inputs to a linear predictor adapted by the well-known LMS (least-mean-square) algorithm of Widrow and Hoff [121, 122] to predict the current channel gain (phase and amplitude). The predicted channel estimate is passed through a low pass filter (LPF) with much smaller delay than the SADD to reduce the noise from the noisy channel estimate and smooth the estimate.

Simulation results and theoretical analysis show that the new algorithm improves the performance of existing algorithms significantly in both BER (bit-error-rate) and decision delay. It is also shown that combining the proposed fading channel tracker with the uncoded coherent quaternary phase shift keying (CQPSK) achieves significant

173

gain over its differential QPSK (DQPSK) counterpart. The complexity of the new algorithm involves about 30 to 60 complex number multiplications depending upon the fading bandwidth, which is larger than DPSK detection and is similar to the SADD. The new algorithm is herein referred to as the *decision-feedback adaptive linear predictive* (DFALP) fading channel tracker. A generalization of the DFALP algorithm to frequency selective fading channels will also be addressed.

The DFALP algorithm proposed in this thesis is able to track certain nonstationarity of fading channels because of the robustness of the DFALP algorithm and the adaptivity of the LMS algorithm used. The nonstationarity may be a small change of certain channel statistics, such as change of vehicle speed, change of channel gain distributions (Rayleigh or Rician distributions), and slow change of the line-of-sight (LOS) part of Rician channels (shadowing). However, if the channel experiences abrupt change of channel statistics, the existing channel tracking algorithms may not be able to react to the change quickly. In the latter situation, the quickest change-detectors are used to improve the performance of the DFALP algorithm.

It has been shown by Irvine and McLane [57] that combining trellis coded modulation (TCM), interleaving and soft Viterbi decoding with the SADD tracker results in 3 dB BER performance improvement of CPSK signal detection over the same coded DPSK scheme, even though the SADD with uncoded CPSK does not always achieve such a gain over its DPSK counterpart. It was found in [57] that 1 dB gain of the CPSK detection comes from soft Viterbi decoding using the channel estimate while the DPSK has no channel information available. If the TCM, interleaving and Viterbi decoding are combined with the DFALP channel tracker as is done in [57], the resulting scheme would perform much more than 3 dB better than the same coded DQPSK according to the result in [57] and the fact that the uncoded DFALP algorithm outperforms the SADD algorithm of [57]. A simulation study combining the DFALP tracker with TCM and Viterbi algorithm is left for future research.

174

6.1 DFALP Algorithm for Frequency Nonselective Fading Channels

6.1.1 DFALP Algorithm

We first derive the new DFALP algorithm for nonselective (flat) fading channels. If x_k is a known training symbol, c_k can be easily estimated using

$$c_k \approx y_k / x_k \stackrel{\triangle}{=} \tilde{c}_k \qquad (6.1.1)$$

according to (5.1.1), where y_k is the received signal, x_k is known training symbol. It is assumed that the signal-to-noise ratio (SNR) is sufficiently large. However, in most situations training symbols are not available. In these cases the only available information is the past detected symbols $\bar{x}_i$ ($i < k$). Since a slow fading channel is highly correlated [59, 62], which is also clearly seen from Figures 5.1-5.3, it is possible to use an adaptive linear filter (a standard one-step adaptive predictor [124]) to estimate the current complex channel gain c_k using the past detected symbols $\bar{x}_i$ ($i < k$) and the current observed signal y_k. The idea of decision-feedback has long been used in adaptive equalization of deterministic telephone channels [98]. The same idea was also used by Irvine and McLane [57]. However, Irvine and McLane mainly used training-symbol information and did not consider the high correlation of both Rayleigh and Rician fading channels.

We now derive the DFALP algorithm. Let the past detected data symbols be $\bar{x}_{k-1}, \bar{x}_{k-2}, \dots \bar{x}_{k-N}$. The past received signals are $y_{k-1}, y_{k-2}, \dots y_{k-N}$. Notice that $\bar{x}_i$ and y_i are complex, and that $\bar{x}_i$ takes on a finite number of values (for QPSK it takes on only four values). Then the past decision feedback complex channel gains are

$$\frac{y_{k-i}}{\bar{x}_{k-i}} \quad i = 1, 2, \dots, N. \qquad (6.1.2)$$

If the fedback data symbol happens to be correct, that is, $\bar{x}_{k-i} = x_{k-i}$, the estimate $y_{k-i}/\bar{x}_{k-i}$ would be reliable. However, this may not be always true. Therefore the

175

estimate $y_{k-i}/\tilde{x}_{k-i}$ should be corrected. Let $\tilde{c}_{k-i}$ denote the corrected fading channel estimate from $y_{k-i}/\tilde{x}_{k-i}$, or the fading channel estimate using training symbols if they are available (Eq.(6.4.1)). Using the standard linear prediction approach [49, 124], we formulate the predicted fading channel gain at time k as [49]

$$\hat{c}_k = \sum_{i=1}^{N} b_i^* \tilde{c}_{k-i} \triangleq \vec{b}(k)^H \vec{\tilde{c}}(k) \tag{6.1.3}$$

where

$$\vec{\tilde{c}}(k) = (\tilde{c}_{k-1}, \tilde{c}_{k-2}, ..., \tilde{c}_{k-N})^T \tag{6.1.4}$$

is a vector of past corrected channel gain estimates and

$$\vec{b}(k) = (b_1, b_2, ..., b_N)^T \tag{6.1.5}$$

are the filter (linear predictor) coefficients at time k. The superscript T stands for transpose and H stands for Hermitian (conjugate and transpose). N is the order of the linear predictor. Our objective is to minimize the estimation error adaptively, i.e.,

$$\min \xi_k = |c_k - \hat{c}_k|^2 \tag{6.1.6}$$

where c_k is the actual complex channel gain. A well-known solution to the problem is Widrow-Hoff's least-mean-square (LMS) algorithm [121, 122]. The LMS algorithm computes the filter coefficients $\vec{b}(k+1)$ of the next time-step using the current filter coefficients $\vec{b}(k)$ and the estimation error $c_k - \tilde{c}_k$. Formally, the algorithm is

$$\vec{b}(k+1) = \vec{b}(k) + \mu (c_k - \hat{c}_k)^* \vec{\tilde{c}}(k) \tag{6.1.7}$$

where μ is a step-size controlling the convergence-rate and steady-state error of the algorithm, and $\vec{\tilde{c}}(k) = (\tilde{c}_{k-1}, \tilde{c}_{k-2}, ..., \tilde{c}_{k-N})^T$ is the vector of past (corrected) complex channel gain estimates. Since the actual channel gain c_k is not available, we substitute the current corrected channel gain estimate $\tilde{c}_k$ and obtain the following practical algorithm

$$\vec{b}(k+1) = \vec{b}(k) + \mu (\tilde{c}_k - \hat{c}_k)^* \vec{\tilde{c}}(k). \tag{6.1.8}$$

176

Now we address the problem of using decision feedback and correction to obtain the corrected channel gain estimate $\tilde{c}_k$ from the predicted channel gain $\hat{c}_k$. First, we estimate the data symbol using the predicted channel gain $\hat{c}_k$:

$$\tilde{x}_k = \frac{y_k}{\hat{c}_k} \qquad (6.1.9)$$

where y_k is the current received signal plus noise, and $\hat{c}_k$ is a channel estimate given by the linear predictor (6.1.3). Second, we use the minimum distance decision rule

$$\min_{x_k \in D} |\tilde{x}_k - x_k| \qquad (6.1.10)$$

where D is the signal constellation of the modulated complex low-pass equivalent signal x_k. For QPSK, $D = \{e^{jn\pi/2}, n = 0, 1, 2, 3\}$. Let $\bar{x}_k$ denote the detected data symbol, i.e.

$$|\tilde{x}_k - \bar{x}_k| = \min_{x_k \in D} |\tilde{x}_k - x_k|. \qquad (6.1.11)$$

Using the detected data symbol $\bar{x}_k$, we formulate a new estimate of the channel gain

$$\frac{y_k}{\bar{x}_k}. \qquad (6.1.12)$$

There exist two possibilities for the detection rule (6.1.11). One possibility is that the decision is correct, i.e., $\bar{x}_k = x_k$. Then the estimate $y_k/\bar{x}_k$ would be reliable. On the other hand, if the decision is wrong, i.e., $\bar{x}_k \neq x_k$, the estimate $y_k/\bar{x}_k$ will certainly be very poor. To solve this problem, we use a thresholding idea proposed by Irvine and McLane [57]. In most cases, if the decision is correct, the distance between the predicted channel gain $\hat{c}_k$ and the decision feedback estimate $y_k/\bar{x}_k$ would not be large, i.e., the probability that $|\hat{c}_k - y_k/\bar{x}_k| < \beta$ would be high, where β is a chosen threshold. On the other hand, if the decision is wrong, the distance between the predicted channel gain $\hat{c}_k$ and the decision-feedback estimate $y_k/\bar{x}_k$ would be large, i.e. the probability that $|\hat{c}_k - y_k/\bar{x}_k| \geq \beta$ would be high. Therefore the corrected channel estimate may be expressed as

$$\tilde{c}_k = \begin{cases} y_k/\bar{x}_k & \text{if } |\hat{c}_k - y_k/\bar{x}_k| < \beta \\ \hat{c}_k & \text{if } |\hat{c}_k - y_k/\bar{x}_k| \geq \beta \end{cases} \qquad (6.1.13)$$

177

There exists no analytical approach to choosing the threshold β. In [57], it was proposed that for q-ary PSK β is chosen to be $e^{j\pi/(2q)}$. This threshold also works well in our simulation.

The corrected channel estimate $\tilde{c}_k$ is then low-pass-filtered using a low pass filter (LPF) of order $2D_f + 1$ to reduce the noise. That is, the final channel gain estimate is

$$\bar{c}_{k-D_f} = \sum_{i=0}^{2D_f} h_i \tilde{c}_{k-i} \tag{6.1.14}$$

where h_i is the impulse response of an LPF of order $2D_f + 1$. It is noted that a delay D_f is incurred after the above final low-pass filtering.

We now summarize the decision-feedback adaptive-linear-prediction (DFALP) algorithm:

DFALP Fading-Channel Tracker

Repeat Steps 1-4:

- Step 1: If a training symbol x_k is available, set $\tilde{c}_k = y_k/x_k$.

    Otherwise,

    1.1 Predict channel gain: $\hat{c}_k = \sum_{i=1}^{N} b_i^* \tilde{c}_{k-i} = \vec{b}(k)^H \vec{\tilde{c}}(k)$;

    1.2 Estimate data symbol: $\hat{x}_k = y_k/\hat{c}_k$;

    1.3 Tentative decision: find $\bar{x}_k$ such that $|\hat{x}_k - \bar{x}_k| = \min_{r_k \in D} |\hat{x}_k - x_k|$;

    1.4 Correction:

    if predicted gain agrees with decision feedback estimate, use fedback estimate,

    i.e., if $|\hat{c}_k - y_k/\bar{x}_k| < \beta$, then $\tilde{c}_k = y_k/\bar{x}_k$,

    else use predicted gain only, i.e., if $|\hat{c}_k - y_k/\bar{x}_k| \geq \beta$, then $\tilde{c}_k = \hat{c}_k$;

- Step 2: Update linear predictor coefficients: $\vec{b}(k+1) = \vec{b}(k) + \mu(\tilde{c}_k - \hat{c}_k)^* \vec{\tilde{c}}(k)$;

- Step 3: Low pass filter $\tilde{c}_k$: $\bar{c}_{k-D_f} = \sum_{i=0}^{2D_f} h_i \tilde{c}_{k-i}$.

- Step 4: Output $\bar{c}_{k-D_f}$ and feedback $\tilde{c}_k$.

The notations are:

$\bar{c}_{k-D_f}$ = final (delayed) channel gain estimate;

$\vec{b}(k) = (b_1, b_2, ..., b_N)^T$ = filter coefficients at time kT;

$\vec{\tilde{c}}(k) = (\tilde{c}_{k-1}, \tilde{c}_{k-1}, ..., \tilde{c}_{k-N})^T$ = past corrected channel gain estimate;

y_k = received signal plus noise at time kT;

μ = step-size;

$\bar{x}_k$ = detected data symbol at time kT.

One training symbol is sent for transmitting every $K_t - 1$ data symbols.
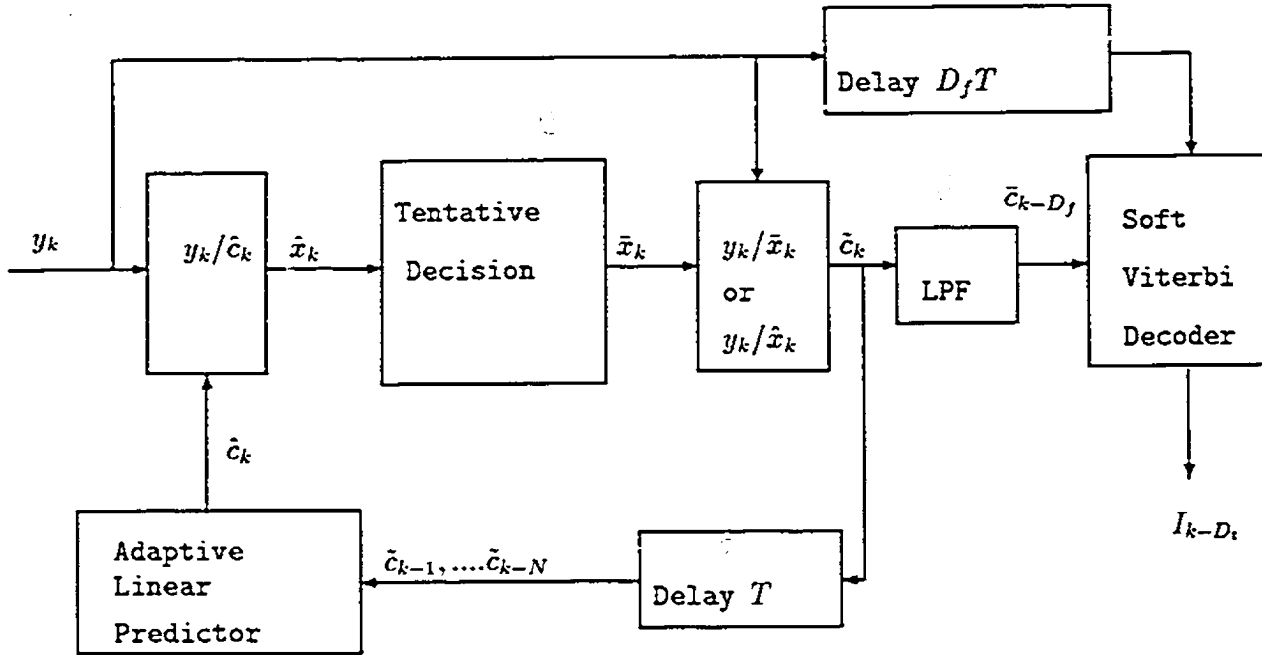
Figure 6.1: *The DFALP algorithm for tracking phase and amplitude of frequency nonselective fading channels.*

The initial conditions for the filter coefficients $\bar{b}$ are chosen to be

$$\vec{b}(0) = (1, 0, ..., 0)^T. \tag{6.1.15}$$

A block diagram of the DFALP algorithm is shown in Figure 6.1. The received signal y_k is divided by a channel estimate $\hat{c}_k$ from the adaptive linear predictor. The estimated data symbol is detected using a minimum distance decision rule. Then the corrected channel estimate $\tilde{c}_k$ is obtained. The corrected channel estimate $\tilde{c}_k$ is then passed through a low pass filter of order $2D_f + 1$ to reduce the noise, resulting in a decision delay D_f. The final smoothed channel estimate $\bar{c}_{k-D_f}$ is sent out for soft Viterbi decoding. Note that only $\tilde{c}_k$ can be fedback and the smoothed estimate $\bar{c}_{k-D_f}$ cannot be fedback since it is delayed by D_f.

Comparing Figure 6.1 to Figure 7 of Irvine-McLane [57] yields that the major difference between the DFALP algorithm and the SADD is that the former mainly utilizes the information extracted from the adaptive linear predictor while the latter uses filtering and interpolation of the channel estimate derived from well-spaced training symbols and results in a large decision delay. Since the DFALP does not require the first filter of the SADD, the total decision delay of the DFALP is significantly shorter than the SADD.

6.1.2 Choosing Parameters for the DFALP Algorithm

The major parts of the DFALP algorithm are the adaptive linear predictor and low pass filter (LPF). The threshold β and the training period K_t also have to be determined.

Design parameters of the adaptive linear predictor include the filter order (number of taps) and the step-size. These parameters are controlled by the convergence rate, stability, steady-state misadjustment (normalized steady-state mean-square error) and channel statistics. According to [124], a condition for stability of the LMS algorithm is

$$0 < \mu < \frac{1}{N\sigma_{\tilde{c}}^2} \tag{6.1.16}$$

where N is the number of taps of the linear predictor and $\sigma_{\tilde{c}}^2$ is the variance of the corrected channel estimate. The misadjustment is given by [124]

$$M_c \approx \mu N \sigma_{\tilde{c}}^2 \tag{6.1.17}$$

and the convergence rate is [124]

$$\frac{1}{\tau} \approx 4\mu\sigma_{\tilde{c}}^2. \tag{6.1.18}$$

To have a small misadjustment M_c and stable algorithm, the step-size μ should be small. Note that

$$\sigma_{\tilde{c}}^2 = Var\{\tilde{c}_k\} = Var\{c_k + n_{\tilde{c}}\} = r_0 + Var\{n_{\tilde{c}}\} \tag{6.1.19}$$

where r_0 is the average channel power and $n_{\tilde{c}}$ is noise plus estimation error. It should be noted that the term $Var\{n_{\tilde{c}}\}$ may be very large sometimes due to occasional large

181

estimation errors. Thus, it is safer to choose small μ to guarantee stability and small misadjustment, particularly for slow fading channels. For fast fading channels a slightly larger step-size may be chosen for adapting the filter to fast varying channels. Typically, it is observed through simulations that choosing the step-size

$$\mu = 10^{-3} f_m T \qquad (6.1.20)$$

achieves a satisfactory performance for a normalized fading bandwidth between $f_m T = 0.0001$ and .08.

The BER performance of the DFALP algorithm is affected by both the linear predictor order N and the low pass filter order $2D_f + 1$.

Theorem 5.1 shows that under certain conditions an optimal detector approaches error-free detection for infinitely large prediction order. In section 5.2.2 it is shown that the optimal detector is approximately equivalent to a combined channel-estimator/coherent-detector. The DFALP algorithm with coherent detection is an approximately optimal detector and therefore would behave in a similar way to the optimal detector. Therefore it is expected that choosing a larger predictor order would result in a better performance. This appears to be true from Figure 6.2, where the BER performance becomes better and better with an increasing predictor order N. However, as shown in Figure 6.2, when the predictor order N is greater than a certain value (50), the improvement of the BER is very small. Therefore a finite predictor order is suitable considering both the performance and the complexity.

We now discuss design parameters for the low-pass filter. The parameters of the LPF include its cut-off frequency and filter order $2D_f + 1$. Notice that the major purpose of the LPF is to reduce the noise from the channel estimate $\tilde{c}_k$ since

$$\tilde{c}_k = c_k + n_{\tilde{c}k} \qquad (6.1.21)$$

where $n_{\tilde{c}k}$ is a combination of additive Gaussian noise and estimation error. For the Jakes-Reudink fading model, the spectrum of the fading channel is the Fourier transform

Bit-Error-Rate P_e
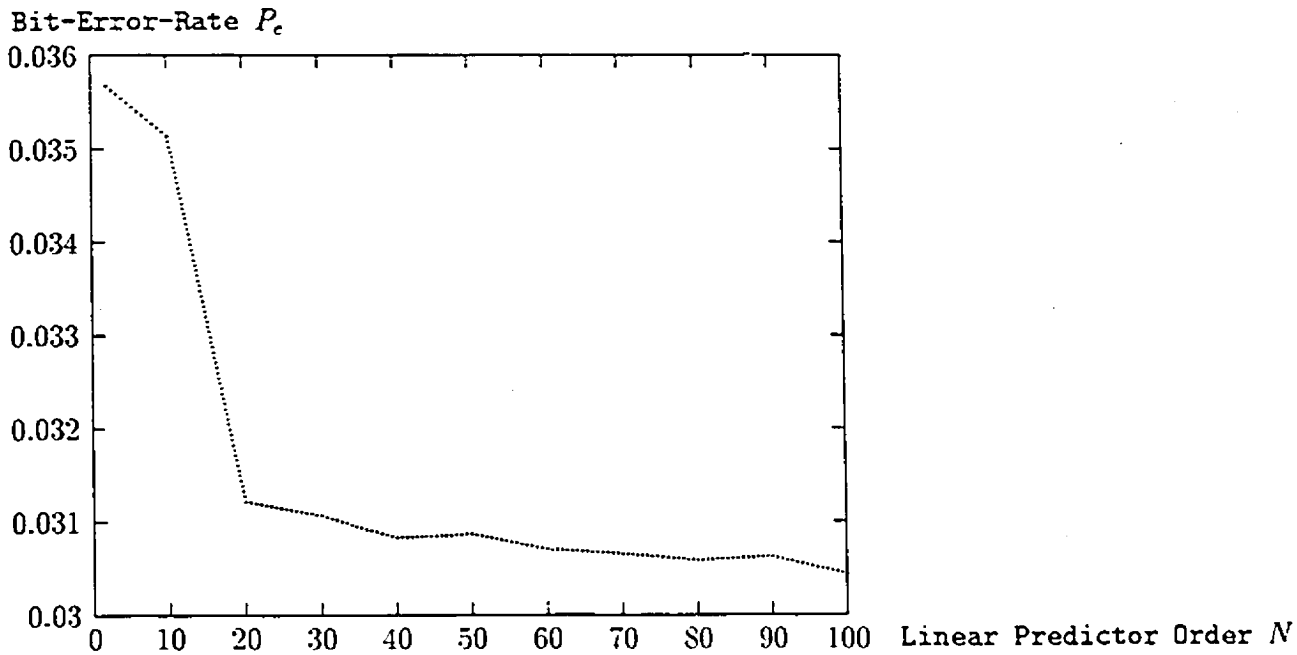
Figure 6.2: *The BER of the DFALP for different linear predictor orders. The low-pass filter order is $2D_f + 1 = 9$ (the decision delay is $D_f = 4$). The CQPSK is used. $v = 60$ km/hour. $f_s = 6000$ symbols/second. $f = 5$ ghz. $K = 4$ dB. $f_m T = 0.0463$. The SNR equals 10 dB. $K_t = 5$.*

of the Bessel function $J_0(2\pi f_m T k)$, i.e.

$$S_c(f) = \begin{cases} 1/\sqrt{1 - \frac{f}{f_m}} & |f| < f_m \\ 0 & |f| \geq f_m \end{cases} \qquad (6.1.22)$$

The channel bandwidth is equal to f_m, the maximum Doppler frequency. Equivalently, the normalized channel bandwidth is $f_m T$. Other fading channel models, for example the exponential spectrum models, have a similar property [80]. On the other hand, the spectrum of the noise n_{ck} is flat and wide. Therefore it is reasonable to let the cut-off frequency of the LPF equal the bandwidth of c_k, i.e., f_m (or its normalized value $f_m T$).

The filter order $2D_f + 1$ of the LPF is constrained by two factors: (i) the BER performance and (ii) the decision delay.

The decision delay of the LPF with order $2D_f + 1$ is D_f [57]. For some applications a large decision delay cannot be tolerated. For example, in voice communications via

183
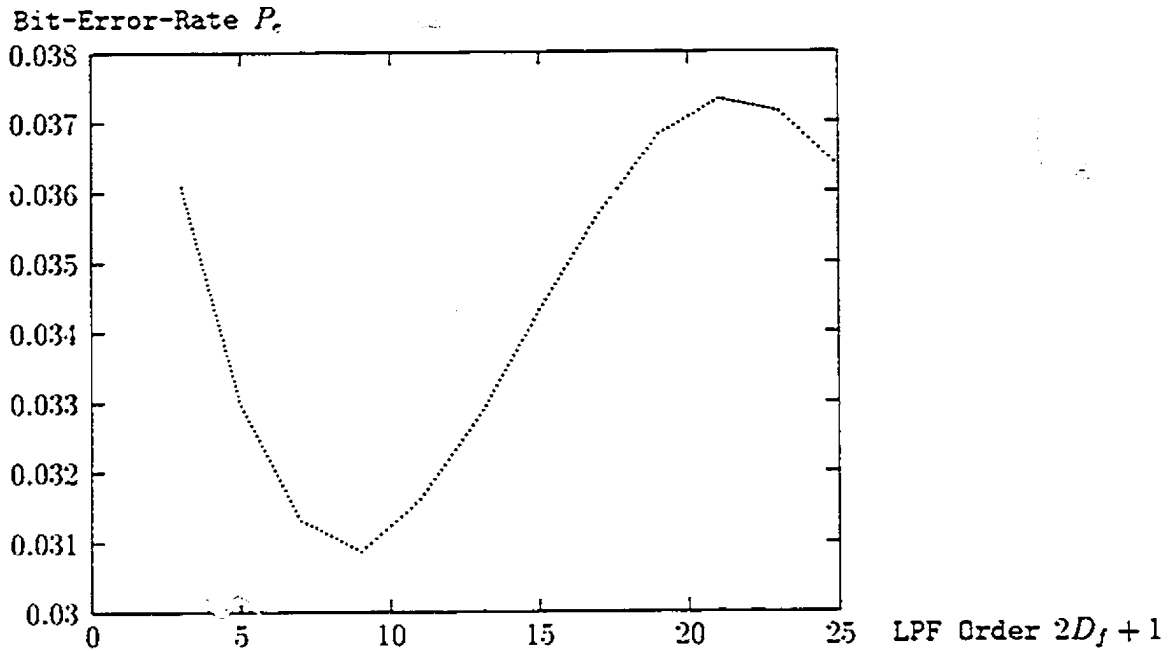
Figure 6.3: *The BER of the DFALP algorithm for different low pass filter orders $2D_f + 1$. The linear predictor order is $N = 50$. The CQPSK is used. $v = 60$ km/hour. $f_s = 6000$ symbols/second. $f = 5$ ghz. $K = 4$ dB. $f_m T = 0.0462$. $K_t = 5$.*

satellite, the transmission delay takes typically 250 ms. and only 50 ms is left for inter-leaving. Viterbi decoding and low-pass filtering. For this reason, we recommend that the LPF order is set to less than $2D_f + 1 = 31$ with decision delay $D_f = 15$ symbols. Notice that Irvine-McLane's SADD algorithm has a decision delay $(K_t + 1)D_f$, which is K_t times larger than the DFALP algorithm, where K_t is the training period. Another important problem in choosing the low pass filter order is that the BER performance of the DFALP algorithm depends upon the LPF order and the normalized fading band-width $f_m T$. This can be observed from Figure 6.3: too large or too small a LPF order results in worse performance.

Considering both the BER performance and complexity, we recommend a number of LPF and linear predictor orders according to extensive simulation results, which are shown in Figure 6.4. The data shown in Figure 6.4 are obtained by simulating a number of linear predictors of different order N and a number of LPFs of different order $2D_f + 1$

184

Predictor Order N and LPF order $2D_f + 1$
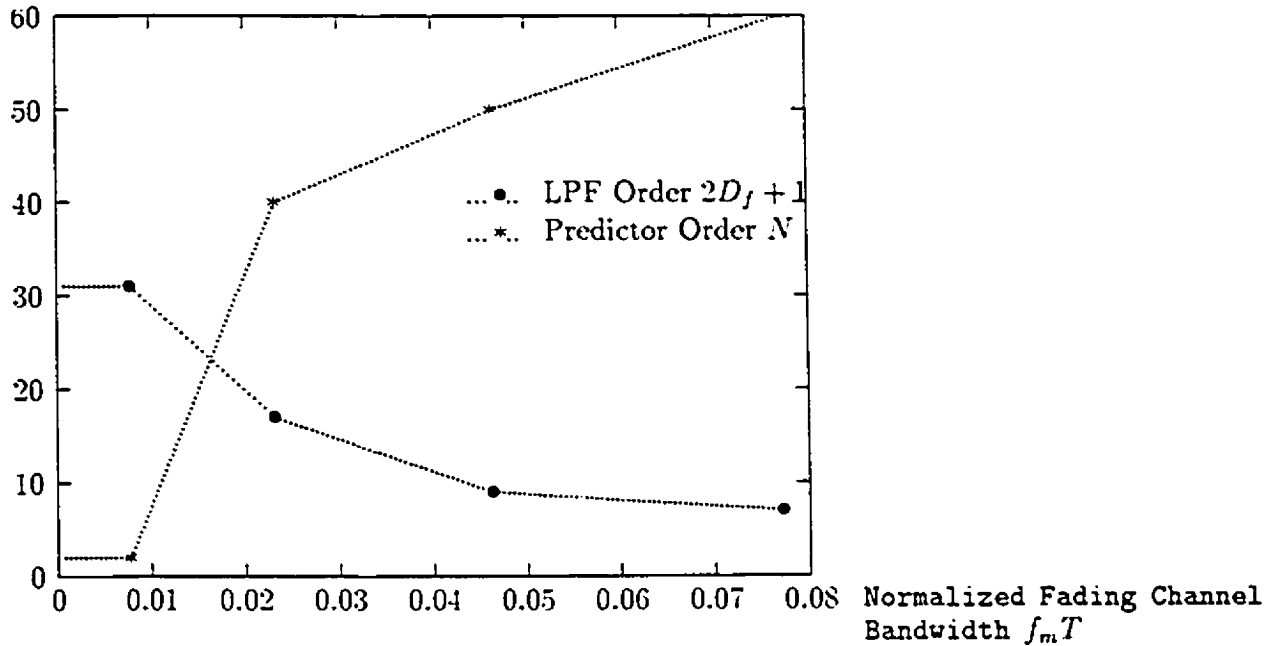
Figure 6.4: *Recommended linear predictor order N and low pass filter order $2D_f + 1$ for the DFALP algorithm.*

for each value of $f_m T$ and choosing the best one. From Figure 6.4 it is seen that for slow fading the LPF order should be large and the predictor order should be small, and the opposite is true for fast fading. Using a least-square-error-fit approach, two empirical equations are derived for determining the linear predictor order and low pass filter order:

$$N = 10^5(3.4241(f_m T)^3 - .5211(f_m T)^2 + .0276(f_m T)) \qquad (6.1.23)$$

and

$$D_f = 10^3(-7.526(f_m T)^3 + 3.6729(f_m T)^2 - 0.3981(f_m T) + 0.0153) \qquad (6.1.24)$$

As is mentioned before, there exists no analytical approach to choosing threshold β for the DFALP algorithm to switch between the decision feedback estimate $y_k/\tilde{x}_k$ and predicted estimate $\hat{c}_k$. Intuitively, this depends upon our confidence in the decision $\tilde{x}_k$. As is suggested in [57], choosing $\beta = e^{j\pi/(pq)}$ for q-ary PSK appears reasonable according to simulations, where $p = 2, 3, \ldots$ depending on different applications. We chose $p = 2$ in our simulations.

185

It is easily seen that choosing a small training period K_t would improve the BER performance and also results in a bandwidth expansion of factor $1/(K_t - 1)$. Therefore there exists a trade-off between the BER performance and bandwidth expansion for choosing the training period K_t. In our simulations we chose $K_t = 5$, which results in 20% bandwidth expansion and a fairly good BER performance.

6.2 Performance of the DFALP Algorithm

6.2.1 Simulation Results

We have simulated the DFALP algorithm with uncoded CQPSK to compare its performance to Irvine-McLane's SADD algorithm, DQPSK, and CQPSK with perfect channel state information (CSI). The results are plotted in Figures 6.5-6.10.

In Figure 6.5 we simulate a typical cellular telephone channel, where the carrier frequency is 800 mhz, the symbol rate is 24000 symbols/second, the mobile speed is 100 km/hour, and the fading channel is Rayleigh. The normalized fading bandwidth is then $f_m T = 0.00309$. The low pass filter order of the DFALP is set to $2D_f + 1 = 31$ and the linear predictor order is $N = 2$. The low pass filter order of the SADD is set to $2D'_f + 1 = 21$. The training period of both SADD and DFALP is $K_t = 5$. Then the decision delay for the DFALP is $D_f = 15$ symbols and is $(K_t + 1)D'_f = 60$ for the SADD. It is clear that the decision delay of the SADD is much larger than the DFALP. The BER performance of the DFALP with uncoded CQPSK is about 2 dB better than the uncoded DQPSK and is about 1 dB worse than the uncoded CQPSK with perfect channel state information. The SADD with uncoded CQPSK performs well for small SNR but reaches a large error floor for large SNR. The large error floor of the SADD is mainly due to its irreducible interpolation error which is independent of the additive noise. However, it should be noted that the SADD with TCM coded and interleaved CPSK performs 3 dB better than its DPSK counterpart as reported in [57].

186

Similar results are also observed from Figure 6.6 for a typical satellite channel, where the carrier frequency is 5 ghz, the mobile speed is 60 km/hour, the symbol rate is 6000 symbols/second, and the channel is Rician with K factor equal to 4 dB. The normalized fading bandwidth is then $f_m T = 0.0463$. The training period is chosen $K_t = 5$. The LPF order for the DFALP is $2D_f + 1 = 9$ and is $2D'_f + 1 = 15$ for the SADD. Then the decision delay for the DFALP is $D_f = 4$ symbols and is $(K_t + 1)D'_f = 42$ for the SADD.

The performance improvement of the DFALP over the SADD is also evident from Figures 6.7 and 6.8, where the mean square error (MSE) of the estimate $\hat{c}_k$ of the complex channel gain is plotted for different fading channels.

It is observed that the performance of the DFALP is much better than the SADD for medium and large SNRs and the gap diminishes as the SNR becomes smaller and smaller. This fact is due to an irreducible interpolation error of the SADD algorithm, which cannot be decreased for even large SNRs. The DFALP does not have this problem since it uses adaptive linear prediction and low pass filtering for estimating the channel gain and does not involve interpolation of the training symbols. However, when the SNR becomes small and the additive white Gaussian noise becomes dominant, the interpolation error of the SADD algorithm becomes less important and its relative performance improves.

Figures 6.9 and 6.10 show the BER performance of the DFALP compared to the SADD for various training periods K_t, where one training symbol is sent to transmit every $K_t - 1$ data symbols. It is seen that for a large range of training symbol rates, the DFALP performs well while the SADD degenerates fast for large training intervals (i.e. fewer training symbols). It is also observed that the DFALP approaches the ideal performance of the CPSK with perfect channel state information when the training period K_t equals 2, i.e. when one training symbol is sent for transmitting every one data symbol. From Figures 6.9 and 6.10 it is seen that there exists a trade-off between the BER performance and the bandwidth efficiency (the value of the training period), i.e. that a larger number of training symbols corresponds to a better BER performance

187
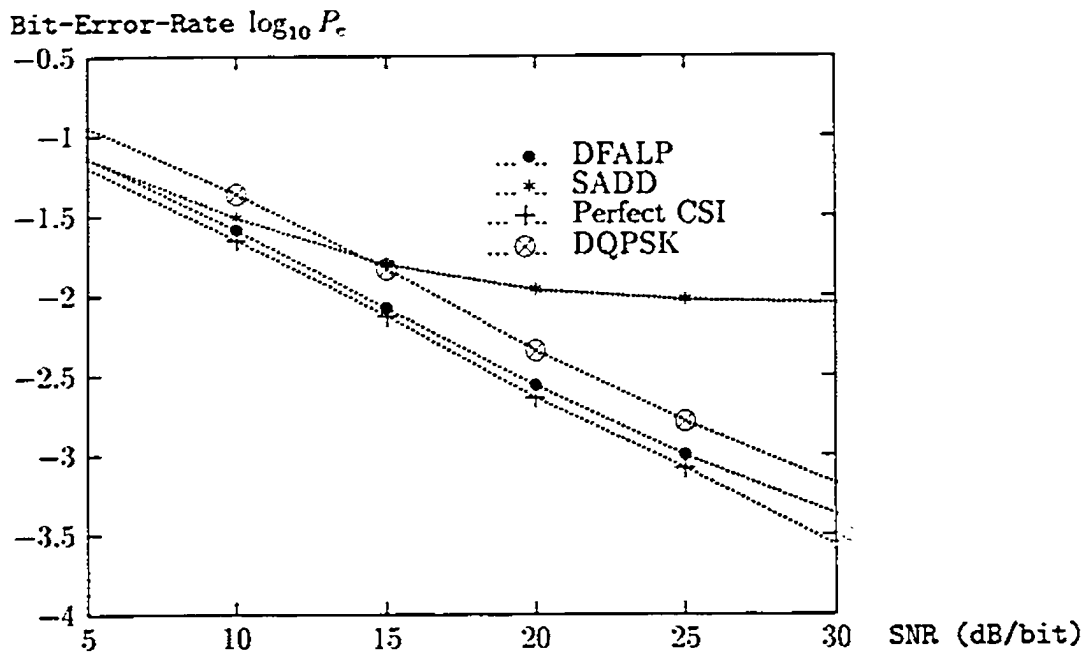
Figure 6.5: *The BER of the DFALP with uncoded CQPSK, Irvine-McLane's SADD with uncoded CQPSK, CQPSK with perfect channel state information, and DQPSK. The decision delay is $D_f = 15$ symbols for the DFALP and 60 symbols for the SADD ($D'_f = 10$). The training period for both DFALP and SADD is $K_t = 5$ (the bandwidth expansion is 20%). The mobile speed is 100 km/hour. The symbol rate is 24000 symbols/second. The wave-frequency is 800 mhz. The normalized fading bandwidth is $f_m T = 0.00309$. The fading channel is Rayleigh.*

but also occupies more bandwidth.

It should be noted that both the phase and amplitude information are obtained by estimating the complex channel gain c_k. In [57] it is shown that combining trellis coded PSK, interleaving and soft Viterbi decoding with the SADD yields 3 dB gain over the same coded DPSK without channel estimation, where 1 dB gain is obtained by utilizing the amplitude estimate in a soft Viterbi decoding. From Figures 6.5-6.10 it appears that combining trellis coded PSK, interleaving and soft Viterbi decoding with the DFALP may potentially yield much more than 3 dB gain over the same coded DPSK without channel estimation.

188
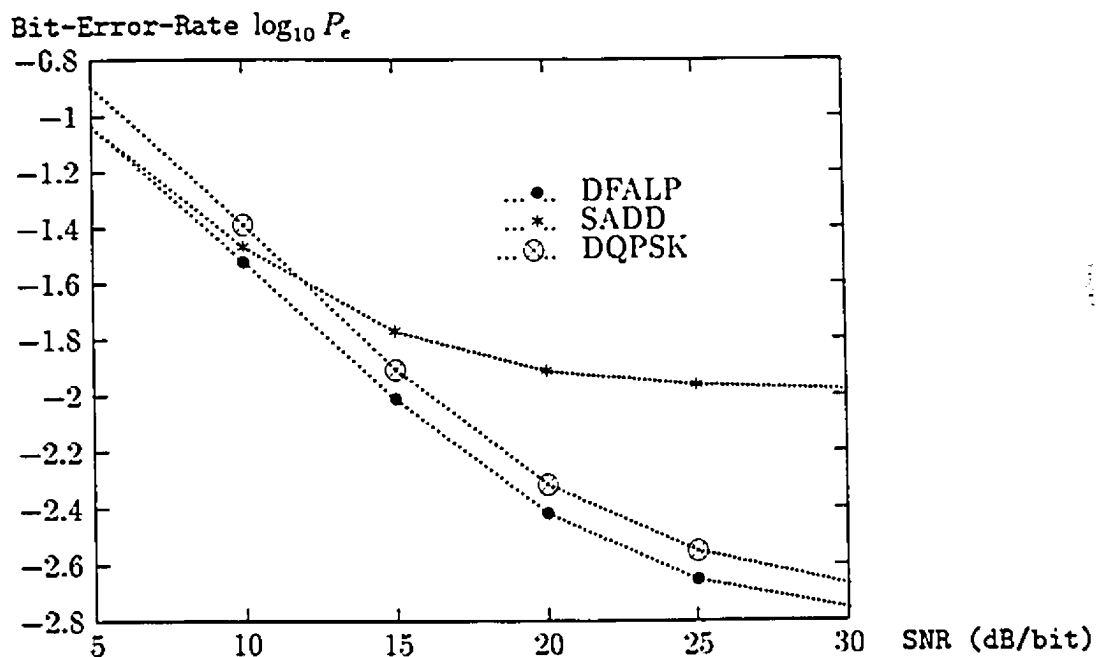
Figure 6.6: *The BER performance of the DFALP with uncoded CQPSK, Irvine-McLane's SADD with uncoded CQPSK, and DQPSK. The decision delay is $D_f = 4$ symbols for the DFALP and 42 symbols for the SADD ($D'_f = 7$). $K_t = 5$. $v = 60$ km/hour. $f_s = 6000$ symbols/second. $f = 5$ ghz. $f_m T = 0.04630$. The fading channel is Rician ($K = 4$ dB).*
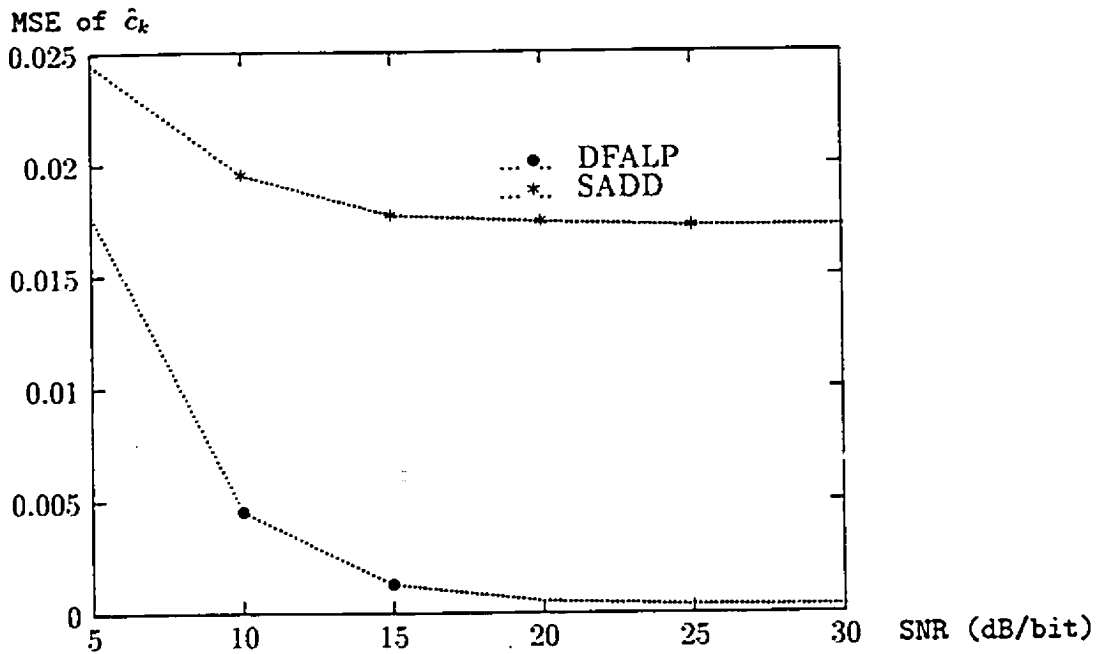
Figure 6.7: *The mean-square-error (MSE) of the estimated complex channel gain $\hat{c}_k$ using the DFALP and Irvine-McLane's SADD. The decision delay is $D_f = 15$ symbols for the DFALP and 60 symbols for the SADD ($D'_f = 10$). $K_t = 5$. $v = 100$ km/hour. $f_s = 24000$ symbols/second. $f = 800$ mhz. $f_m T = 0.00309$. The fading channel is Rayleigh.*

190
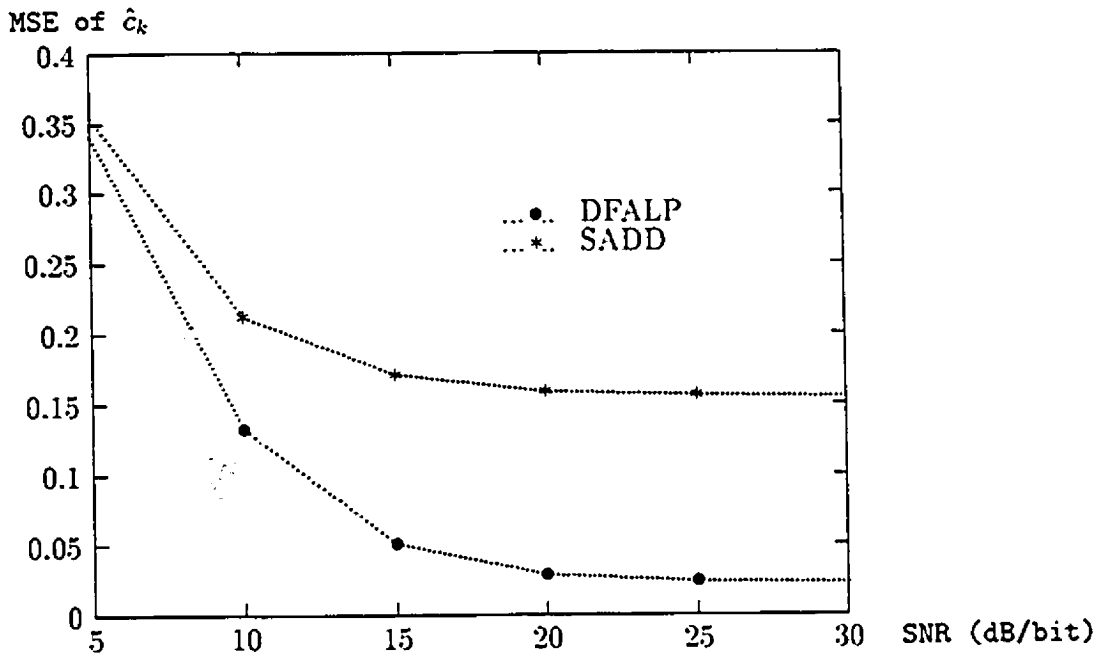
Figure 6.8: *The mean-square-error (MSE) of the estimated complex channel gain $\hat{c}_k$ using the DFALP and Irvine-McLane's SADD. The decision delay is $D_f = 4$ symbols for the DFALP and 42 symbols for the SADD ($D'_f = 7$). $K_z = 5$. $v = 60$ km/hour. $f_s = 6000$ symbols/second. $f = 5$ ghz. $f_m T = 0.04630$. The fading channel is Rician ($K = 4$ dB).*

191
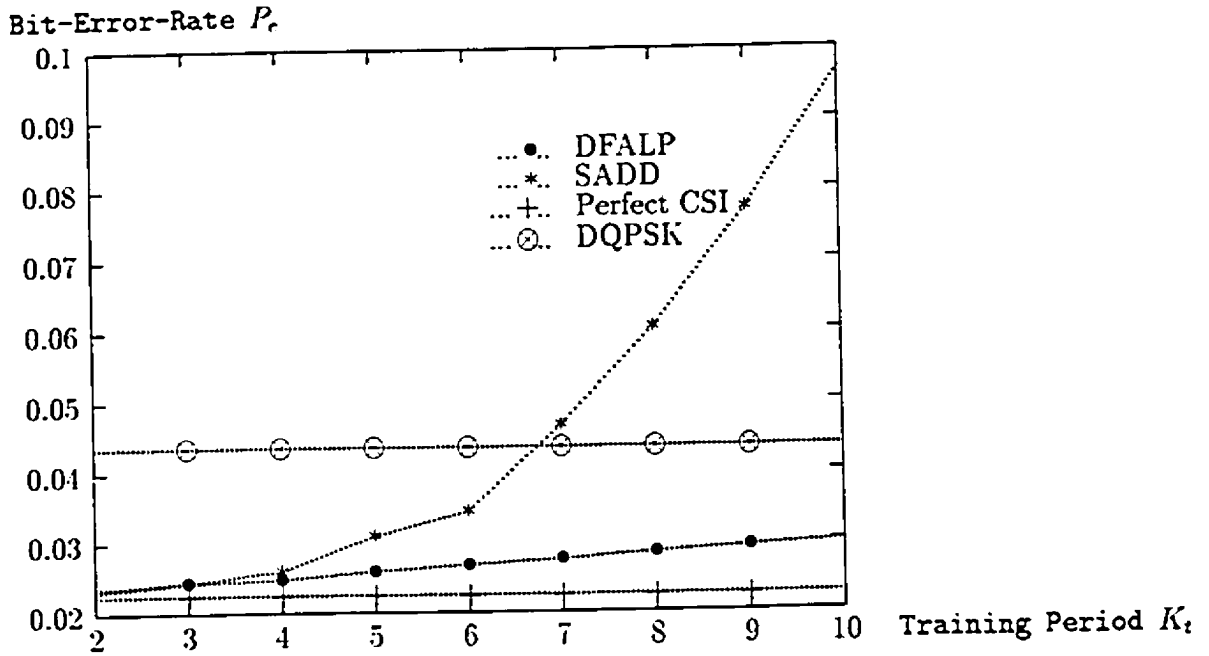
Figure 6.9: *The BER of the DFALP and SADD for different training periods. The decision delay is $D_f = 15$ symbols for the DFALP and 60 symbols for the SADD ($D'_f = 10$). $v = 100$ km/hour. $f_s = 24000$ symbols/second. $f = 800$ mhz. $f_m T = 0.00309$. SNR= 10 dB. The fading channel is Rayleigh.*

192
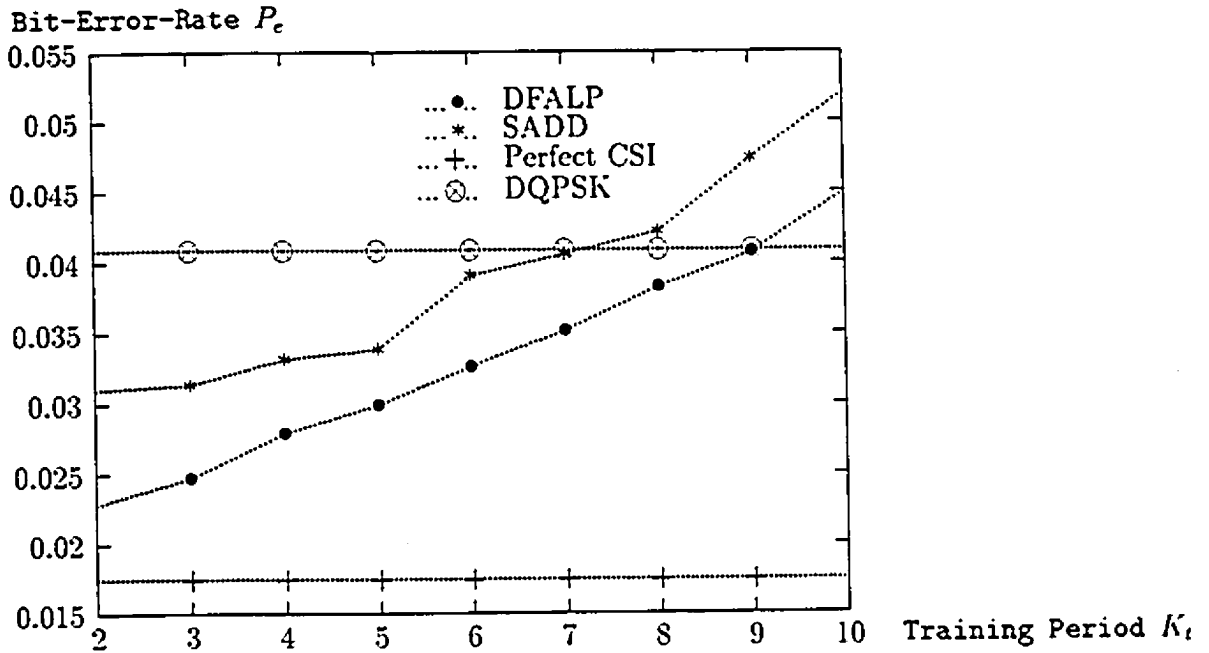
Figure 6.10: *The BER of the DFALP and SADD for different training periods. The decision delay is $D_f = 4$ symbols for the DFALP and 42 symbols for the SADD ($D'_f = 7$). $v = 60$ km/hour. $f_s = 6000$ symbols/second. $f = 5$ ghz. $f_m T = 0.04630$. SNR= 10 dB. The fading channel is Rician ($K = 4$ dB).*

193

6.2.2   Analytical Estimation of Performance of DFALP

According to [46], the BER of symbol-by-symbol CPSK signal detection is

$$P_e = \frac{1}{2}\left(1 - \frac{1}{\sqrt{1 + \frac{\sigma_{\hat{c}}^2 \log_2 q + \gamma^{-1}}{1 - \sigma_{\hat{c}}^2}}}\right) \tag{6.2.1}$$

where $\sigma_{\hat{c}}^2$ is the mean-square-error of the estimator $\hat{c}_k$ of c_k and $q = 2$ for BPSK and $q = 4$ for QPSK. It is seen that the BER performance is proportional to the MSE of the channel gain estimator $\hat{c}_k$. We now derive a close form solution of the MSE for the DFALP algorithm, which, combined with the above equation, may be used to estimate the performance of the DFALP algorithm. The MSE of the estimator $\hat{c}_k$ equals

$$\xi_d = E\{|c_k - \hat{c}_k|^2\}. \tag{6.2.2}$$

Since the MMSE estimator $\hat{c}_k$ is unbiased, without loss of generality it is then sufficient to study the case for $E(c_k) = a = 0$. Let the optimum linear predictor coefficients be denoted by the vector $\vec{b}$. The covariance matrix of $\vec{c}(k) = (c_{k-1}, ..., c_{k-N})^T$ is

$$R_c = E\{\vec{c} \cdot \vec{c}^H\} = \begin{pmatrix} r_0 & r_1 & \cdots & r_{N-1} \\ r_1 & r_0 & \cdots & r_{N-2} \\ \ddots & \ddots & \ddots & \ddots \\ r_{N-1} & r_{N-2} & \cdots & r_0 \end{pmatrix} \tag{6.2.3}$$

and the cross-covariance between c_k and $\vec{c}(k)$ is

$$\vec{r} = E\{c_k \vec{c}\} = (r_1, ..., r_N)^T \tag{6.2.4}$$

Then the MSE is

$$\begin{aligned}
\xi_d &= E\{|c_k - \vec{b}^H \vec{c}|^2\} \\
&= E\{|c_k - \vec{b}^H (\vec{c} + \vec{n}')|^2\}
\end{aligned} \tag{6.2.5}$$

where $\vec{\hat{c}}(k) = (\hat{c}_{k-1}, ..., \hat{c}_{k-N})^T$ is the feedback channel estimate, $\vec{n}'(k) = (\frac{n_{k-1}}{x_{k-1}}, ..., \frac{n_{k-N}}{x_{k-N}})^T$ with n_i the AWGN and x_i the training or decision-feedback data symbols (assuming

194
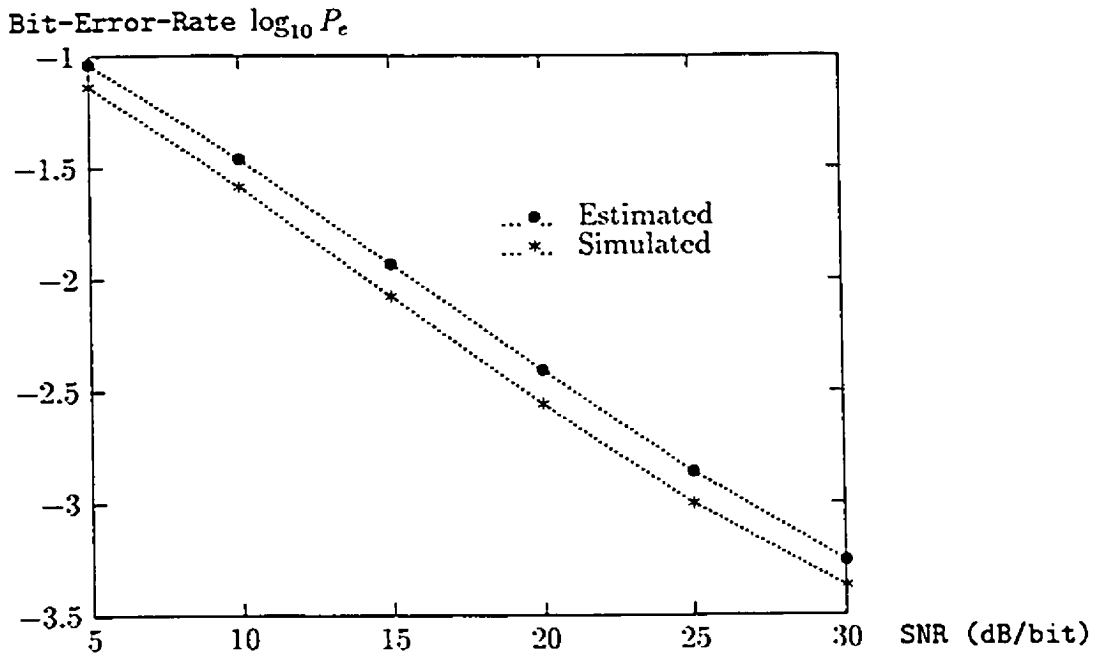
Figure 6.11: *An analytical estimate of the BER of the DFALP with uncoded CQPSK. $N = 2$. $v = 100$ km/hour. $f_s = 24000$ symbols/second. $f = 800$ mhz. $f_m T = 0.00309$. The fading channel is Rayleigh.*

perfect detection). It is noted that n_i has zero-mean and is i.i.d. and independent of c_i. The covariance matrix of $\vec{n}'$ is

$$R_{n'} = \frac{1}{\gamma_s} I_{N \times N} \qquad (6.2.6)$$

where $I_{N \times N}$ is an identity matrix and γ_s is the SNR per symbol. Then

$$
\begin{aligned}
\xi_d &= E\{|c_k|^2\} - 2E\{c_k \hat{c}_k\} + E\{|\hat{c}_k|^2\} + E\{|\vec{b}^H \vec{n}'|^2\} \\
&= r_0 - 2E\{c_k \vec{b}^H \vec{c}\} + E\{\vec{b}^H \vec{c}\vec{c}^H \vec{b}\} + E\{\vec{b}^H \vec{n}' \vec{n}'^H \vec{b}\} \\
&= r_0 - 2\vec{b}^H \vec{r} + \vec{b}^H R_c \vec{b} + \frac{1}{\gamma_s} \vec{b}^H \vec{b} \qquad (6.2.7)
\end{aligned}
$$

It is well-known that the steady-state solution of MMSE filter coefficients $\vec{b}$ satisfies the normal equation.[124]

$$(R_c + R_{n'})\vec{b} = \vec{r}. \qquad (6.2.8)$$

195

Solving (6.2.8) for $\bar{b}$ and substituting into (6.2.7) yields

$$\xi_d = r_0 - \bar{r}^T(R_c + \frac{1}{\gamma_s}I_{N\times N})^{-1}\bar{r} \qquad (6.2.9)$$

where the fact that R_c is symmetric and positive definite is used. It is also noted that $r_k > 0$ for all k. Then an estimated MSE of the channel gain estimator $\hat{c}_k$ is

$$\sigma_{\hat{c}}^2 \approx \xi_d = r_0 - \bar{r}^T(R_c + \frac{1}{\gamma_s}I_{N\times N})^{-1}\bar{r}. \qquad (6.2.10)$$

For the Jakes-Reudink fading model, r_k is given by

$$r_k = J_0(2\pi f_m T k) \qquad (6.2.11)$$

where

$$f_m T = \frac{v}{\lambda}T = \frac{vf}{v_l f_s} \qquad (6.2.12)$$

is the normalized fading bandwidth with v the vehicle speed, f the wave frequency, f_s the symbol rate and v_l the speed of light. Eqs.(6.2.1) and (6.2.10) can be used to estimate the BER performance of the DFALP algorithm. It should be noted that the effect of the low pass filter is not considered in the above derivation, which reduces the effective noise energy of the final channel estimate $\bar{c}_k$. Therefore the BER may be sometimes over-estimated (i.e. worse than the true value).

Figure 6.11 plots the analytical estimate of the BER performance of the DFALP algorithm with CQPSK. The simulated values are also plotted for comparison. It is seen that the estimated BER is slightly over-estimated due to the ignorance of the LPF.

6.2.3   Analytical Estimation of Performance of SADD

To compare the performance of the DFALP to the SADD of [57, 58], we now present an analytical estimate of the performance of the SADD. The linear interpolation of the training symbols of the SADD is plotted in Figure 5.5. Since it is found that interpolation of the complex channel-gain actually yields slightly better performance of

196

Estimated log-MSE $\log_{10} \xi$
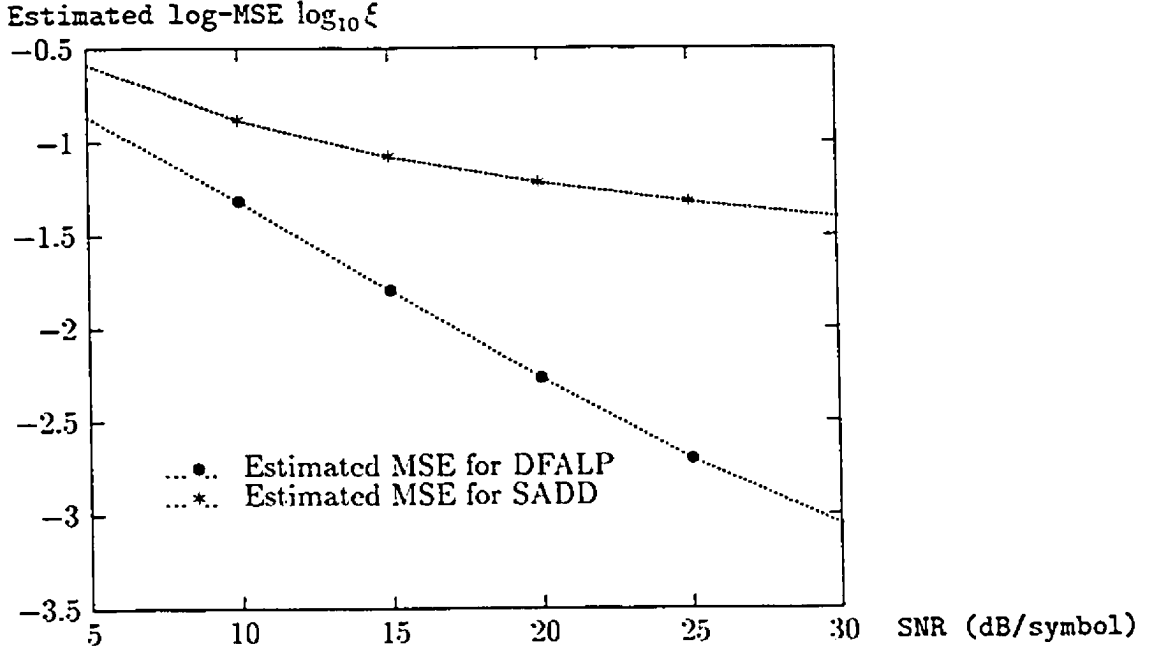
Figure 6.12: *Analytical estimates of the MSE of the DFALP and SADD. $N = 2$. $K_t = 5$. $v = 100$ km/hour. $f_s = 24000$ symbols/second. $f = 800$ mhz. $f_m T = 0.00309$. The fading channel is Rayleigh.*

the SADD than interpolation of the phase as proposed in [57, 58], we will analyze the interpolation of the complex channel-gain to measure the performance of the SADD.

We first ignore the effect of the decision-feedback of the SADD and will later add this operation to the analysis. The step-size of the interpolation may be written as

$$\delta \triangleq \frac{c_k + n'_k - (c_{k-\kappa_t} + n'_{k-\kappa_t})}{K_t} \qquad (6.2.13)$$

where $n'_k = n_k/x_k$. Let the channel estimate of the SADD be denoted by $\hat{c}_{sk}$. The MSE of the symbol-aided part of the SADD algorithm may be determined:

$$\xi_s \triangleq E\{|c_k - \hat{c}_{sk}|^2\}$$

$$= E\{\frac{1}{K_t} \sum_{i=0}^{K_t-1} |c_{k-i} - (c_k + n'_k - i\delta)|^2\}$$

$$= \frac{1}{K_t} E\{\sum_{i=0}^{K_t-1} |c_{k-i} - (1 - \frac{i}{K_t})c_k - \frac{i}{K_t}c_{k-\kappa_t} - (1 - \frac{i}{K_t})n'_k - \frac{i}{K_t}n'_{k-\kappa_t}|^2\}$$

$$= \frac{1}{K_t} \sum_{i=0}^{K_t-1} \{r_0[1 + (1 - \frac{i}{K_t}) + \frac{i}{K_t}] + (1 - \frac{i}{K_t})\frac{1}{\gamma_s} + \frac{i}{K_t}\frac{1}{\gamma_s} -$$

197

$$2(1 - \frac{i}{K_t})r_i - 2\frac{i}{K_t}r_{K_t-i} + 2(1 - \frac{i}{K_t})\frac{i}{K_t}r_{K_t}\}$$

$$= \frac{2}{K_t} \sum_{i=1}^{K_t-1} \{r_0 - (1 - \frac{i}{K_t})r_i - \frac{i}{K_t}r_{K_t-i} + (1 - \frac{i}{K_t})\frac{i}{K_t}r_{K_t}\} + \frac{1}{\gamma_s}$$

$$= \varepsilon_i^2 + \frac{1}{\gamma_s} \qquad (6.2.14)$$

where we have used (6.2.13), and γ_s is the SNR per symbol. We identify the first term

$$\varepsilon_i^2 = \frac{2}{K_t} \sum_{i=1}^{K_t-1} \{r_0 - (1 - \frac{i}{K_t})r_i - \frac{i}{K_t}r_{K_t-i} + (1 - \frac{i}{K_t})\frac{i}{K_t}r_{K_t}\} \qquad (6.2.15)$$

as the MSE due to linear interpolation and the second term $1/\gamma_s$ as the MSE due to the additive noise.

When the decision feedback of the SADD is considered, the interpolation error ε_i^2 would be decreased significantly since a correct tentative decision would eliminate the interpolation error and only the falsely detected feedback would contribute to the interpolation error. Therefore the MSE estimate is modified by

$$\xi_s = \varepsilon_i^2 P_e(\xi_s) + \frac{1}{\gamma_s} \qquad (6.2.16)$$

where ε_i^2 is given by Eq.(6.2.15) and $P_e(\xi_s)$ is given by Eq.(6.2.1) with $\sigma_c^2 = \xi_s$. Eq.(6.2.16) has to be solved recursively. It is found that using less than 10 recursions usually yields reasonable estimate of the MSE of the SADD.

Figure 6.12 plots the estimated MSE of the DFALP and SADD. From both Eq.(6.2.16) and Figure 6.12, it is seen that the irreducible interpolation error ε_i^2 of the SADD is large enough to prevent performance improvement without coding. This also confirms the observation from Figure 5.5 that the interpolation error would be large. By avoiding interpolation, the DFALP algorithm effectively has solved this problem.

198

6.3 Adaptive Identification of Nonselective Fading Channels Experiencing Changes of Vehicle Speed

In the last section it is shown that the performance of the channel tracking algorithm is affected by the channel statistics, particularly the vehicle speed, since the normalized fading channel bandwidth equals

$$f_m T = \frac{vT}{\lambda} \tag{6.3.1}$$

where λ is wavelength and v is vehicle speed. It is easily seen that the vehicle speed is the most important factor affecting the fading channel. Changes of the vehicle speed occur almost surely in a congested urban area. Therefore it is of practical interest to monitor the vehicle speed and adjust the parameters of channel tracking algorithms to improve the performance, particularly for base stations of the cellular telephone network or large mobile units which can afford higher complexity and require better performance.

6.3.1 Monitoring Vehicle Speed by Measuring Level Crossing Rate

It has been shown that an important factor affecting the fading channels is the vehicle speed. It is shown in [62] that there exists a close relationship between the vehicle speed and the level crossing rate. Following [62], it can be shown that the level crossing rate

$$\bar{n}(|c_k - a| = A) = \bar{n}(R) = \frac{\beta v}{\sqrt{2\pi}} R e^{-R^2} = \sqrt{2\pi} f_m R e^{-R^2} \tag{6.3.2}$$

and the average fade duration

$$\bar{t}(|c_k - a| = A) = \bar{t}(R) = \frac{\sqrt{2\pi}}{\beta v} \frac{1}{R} (e^{R^2} - 1) = \frac{1}{\sqrt{2\pi} f_m} \frac{1}{R} (e^{R^2} - 1) \tag{6.3.3}$$

with

$$R = \frac{A}{\sqrt{r_0}} \tag{6.3.4}$$

199

where c_k is the fading channel gain, $a = E(c_k)$, $r_0 = Var(c_k) = E(c_k^2) - |a|^2$. A is a predetermined level. Notice that

$$\beta = \frac{2\pi}{\lambda} \qquad (6.3.5)$$

where λ is wavelength and is known. Therefore measuring the level crossing rate would yield an estimate of vehicle speed v and fading channel bandwidth f_m.

By choosing R to be -5 dB, or $A = 0.5623\sqrt{r_0}$ as is suggested in [62], the vehicle speed is estimated by

$$v \approx \frac{1.6579}{\beta \bar{t}} = 6.1154\frac{\bar{n}}{\beta} \qquad (6.3.6)$$

6.3.2 Measuring Level Crossing Rate by Using Change Detectors

In this subsection the change detection algorithms studied in this thesis are applied to measure the level crossing rate $\bar{n}$. It is noted that for a given level, for example $R = -5$ dB, a level crossing detector is essentially a deep fade detector. For a given level R we count the number of level crossings n during a given period of time $(t_1, t_2]$, then the level crossing rate can be estimated using

$$\bar{n} \approx \frac{n}{t_2 - t_1}. \qquad (6.3.7)$$

Then the vehicle speed can be estimated using Eq.(6.3.6).

Notice that the observed signal $y_k = x_k * c_k + n_k$ contains a noise term n_k which may cause false alarms, i.e., that a level crossing is falsely reported. In addition, if the detection delay is too large, some level crossings may be missed. Therefore it is seen that the level crossing detector is actually a change detector. We now develop a change detection model and then apply the MAR and CUSUM procedures.

Consider the power of received signal y_i

$$z_i = |y_i|^2. \qquad (6.3.8)$$

200

It is easily seen that

$$z_i = |c_i|^2|x_i|^2 + c_i^* x_i^* n_i + c_i x_i n_i^* + |n_i|^2$$

$$\approx |c_i|^2 + c_i^* x_i^* n_i + c_i x_i n_i^* \tag{6.3.9}$$

where we used the fact that $|x_i|^2 = 1$ for constant-envelope modulation and assumed that the signal-to-noise ratio (SNR) is large such that the term $|n_i|^2$ is negligible compared to other terms. Since c_i is locally deterministic and is conditioned on data symbol x_i, z_i is a linear combination of zero-mean Gaussian random variables n_i. That is, z_i is Gaussian distributed with mean

$$E\{z_i\} = |c_i|^2 \tag{6.3.10}$$

and variance

$$Var\{z_i\} = 2E\{c_i^* x_i^* n_i c_i x_i n_i^*\}$$

$$= 2|c_i|^2|x_i|^2 \sigma_n^2$$

$$= 2|c_i|^2 \sigma_n^2 \tag{6.3.11}$$

where σ_n^2 is the variance of the noise n_i. The channel energy $|c_i|^2$ equals

$$A_0^2 = r_0 + |a|^2 \tag{6.3.12}$$

at normal level and drops to below

$$A_1^2 = 0.5623^2 r_0 + |a|^2 \tag{6.3.13}$$

during a fade, where $0.5623^2 r_0$ implies a 5 dB power drop and is proposed in [62] for measuring the level crossing rate. Therefore the problem is formally expressed as

$$H_0: \quad \text{Channel energy equals } A_0^2$$

$$H_1: \quad \text{Channel energy drops to } A_1^2$$

or equivalently

$$H_0: \quad z_i \sim N(A_0^2, 2A_0^2 \sigma_n^2)$$

$$H_1: \quad z_i \sim N(A_1^2, 2A_1^2 \sigma_n^2)$$

201

The densities of z_i under the above two hypotheses are

$$f_0(z_i) = \frac{1}{\sqrt{4\pi A_0^2 \sigma_n^2}} e^{-\frac{(z_i - A_0^2)^2}{4A_0^2 \sigma_n^2}} \qquad (6.3.14)$$

and

$$f_1(z_i) = \frac{1}{\sqrt{4\pi A_1^2 \sigma_n^2}} e^{-\frac{(z_i - A_1^2)^2}{4A_1^2 \sigma_n^2}} \qquad (6.3.15)$$

respectively. The one-sample likelihood ratio is then

$$
\begin{aligned}
l_i &\triangleq \frac{f_1(z_i)}{f_0(z_i)} \\
&= \frac{A_0}{A_1} \exp\left(\frac{A_0^2 - A_1^2}{4\sigma_n^2}\right) \exp\left(\frac{1}{4\sigma_n^2}\left(\frac{1}{A_0^2} - \frac{1}{A_1^2}\right) z_i^2\right) \\
&= A_e e^{A_s |y_i|^4}
\end{aligned}
\qquad (6.3.16)
$$

where

$$A_e \triangleq \frac{A_0}{A_1} \exp\left(\frac{A_0^2 - A_1^2}{4\sigma_n^2}\right) \qquad (6.3.17)$$

and

$$A_s \triangleq \frac{1}{4\sigma_n^2}\left(\frac{1}{A_0^2} - \frac{1}{A_1^2}\right). \qquad (6.3.18)$$

The one-sample log-likelihood ratio is

$$\ln l_i = \ln A_e + A_s |y_i|^4. \qquad (6.3.19)$$

The CUSUM statistic is

$$T_n = \max\{T_{n-1}, 1\} l_i \qquad (6.3.20)$$

with $T_0 = 1$. The CUSUM decision rule is that one decides a change has occurred as soon as

$$T_n \geq B_0 \qquad (6.3.21)$$

is satisfied, where B_0 is a decision threshold.

The MAR statistic is

$$\hat{r}(n) = \frac{T_n \pi_0^2 - (1 - \pi_0)}{T_n \pi_0^2 + 1 - \pi_0} n \qquad (6.3.22)$$

202

where $0 < \pi_0 < 0.5$ is the design parameter. The MAR decision rule is that one decides a change has occurred as soon as

$$\hat{r}(n) > \hat{r}(n-1).$$ (6.3.23)

From the above equations, it is seen that the CUSUM procedure is a sequential energy detector with automatic lower-value reset. The MAR statistic is a transform of the CUSUM statistic, and is a function of the sample energy $|y_i|^2$. The MAR procedures is sequential. It is noted that the detector is independent of the data symbols x_k if constant envelope modulation is used.

It is also desirable to detect the recovery of the channel power from a deep fade A_1^2 to the normal level A_0^2 in order to measure the number of level crossings. The algorithm is the same as the one described above except that the CUSUM statistic becomes

$$T_n' = \max\{T_{n-1}', 1\}\frac{1}{l_i}$$ (6.3.24)

with the decision threshold B_1 and the MAR statistic is

$$\hat{r}(n) = \frac{T_n'\pi_1^2 - (1-\pi_1)}{T_n'\pi_1^2 + 1 - \pi_1}n.$$ (6.3.25)

The combination of (6.3.20) and (6.3.24) represent a switched CUSUM detector which detects both the power drop from A_0^2 to A_1^2 and the power recovery from A_1^2 to A_0^2. Eqs.(6.3.22) and (6.3.25) represent a switched MAR detector accomplishing the same task.

We now try to design the thresholds B_0, B_1 for CUSUM and π_0, π_1 for MAR. According to the results in Section 3.6, the false alarm probability α for both CUSUM and MAR is a function of the actual change-time m. It has been observed that the false alarm probability is approximately a linear function of the actual change-time m. Therefore in order to maintain a fixed false alarm probability, we propose to vary the decision thresholds of both CUSUM and MAR linearly with time n which represents the time from the end of the most recent detected level crossing to the present moment

203

Real Vehicle Speed v
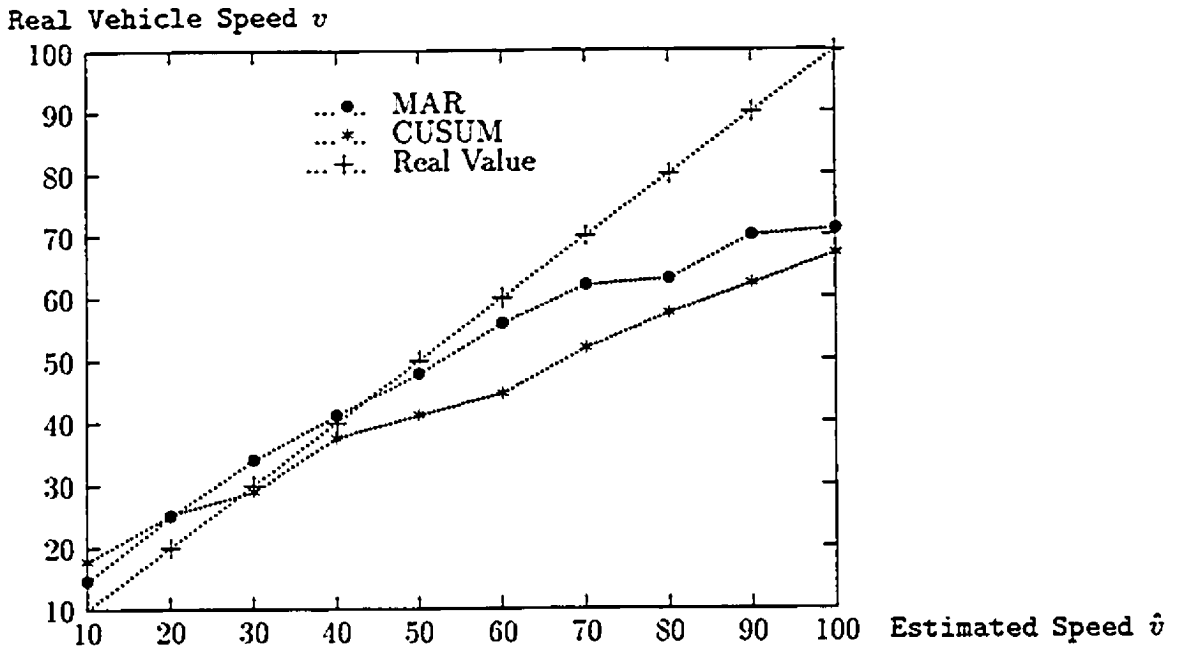
Figure 6.13: *The estimated vehicle speed using the MAR and CUSUM.* $f_s = 6000$ *symbols/second.* $f = 5$ *ghz. The fading channel is Rician (K = 4 dB).*

when a level crossing has not been detected ($n \leq m$). Formally, the parameters are designed as

$$B_{0n} = nB_0 \qquad B_{1n} = nB_1 \qquad (6.3.26)$$

and

$$\pi_{0n} = \frac{\pi_0}{n} \qquad \pi_{1n} = \frac{\pi_1}{n}. \qquad (6.3.27)$$

Using the above varying-threshold change-detection procedures and Eq.(6.3.6), the vehicle speed of a fading channel is estimated and plotted in Figure 6.13. It is seen that there exists certain errors between the estimated vehicle speed $\hat{v}$ and the actual speed v. Both procedures underestimate the speed when the actual value is large, and they over-estimate the speed when the actual value is small. The underestimation occurs for fast fading since level crossings occur so frequently that change-detection procedures cannot detect all level crossings and therefore under-estimate the level crossing rate. On the other hand, overestimation occurs for slow fading since level crossings do not occur frequently and the change-detection procedures contribute false alarms. It is interesting

Bit-Error-Rate $\log_{10} P_e$
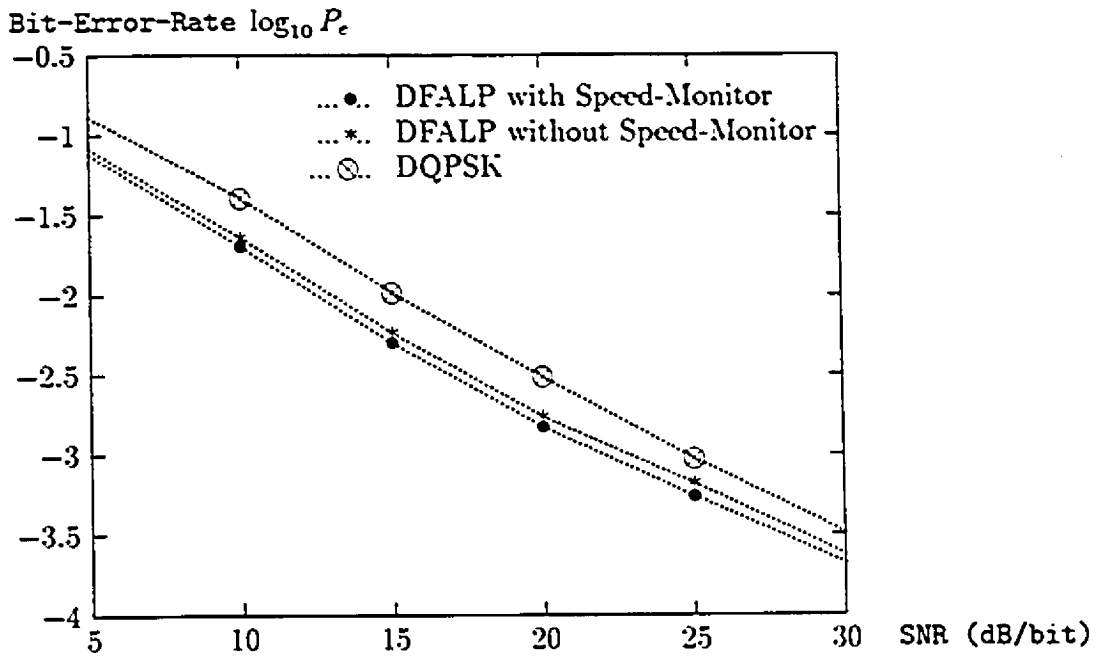
Figure 6.14: *The BER of the DFALP with or without a vehicle speed monitor and DQPSK. $K_t = 5$. The mobile speed changes from 60 km/hour to 6 km/hour. $f_s = 6000$ symbols/second. $f = 5$ ghz. The fading channel is Rician ($K = 4$ dB).*

that the MAR estimate is slightly closer to the actual value than the CUSUM estimate. Using a least-square-fit approach and the data shown in Figure 6.13, an empirical relationship between the estimated vehicle speed and the actual speed is obtained as

$$v \approx 0.0018\hat{v}^3 - 0.0352\hat{v}^2 + 1.0581\hat{v} - 2.1636 \qquad (6.3.28)$$

for the MAR estimate and

$$v \approx -0.0002\hat{v}^4 + 0.0174\hat{v}^3 - 0.6184\hat{v}^2 + 11.8242\hat{v} - 81.7958 \qquad (6.3.29)$$

for the CUSUM estimate.

Figures 6.14 and 6.15 show the BER performance of the DFALP with or without a vehicle speed monitor, where an MAR speed monitor is used. As a reference, the BER performance of the DQPSK detection is also plotted. From Figure 6.14 it is seen that when the vehicle speed decreases from 60 km/hour to 6 km/hour, the performance degenerates slightly if the vehicle speed monitor is not used to adjust the filter order and

205
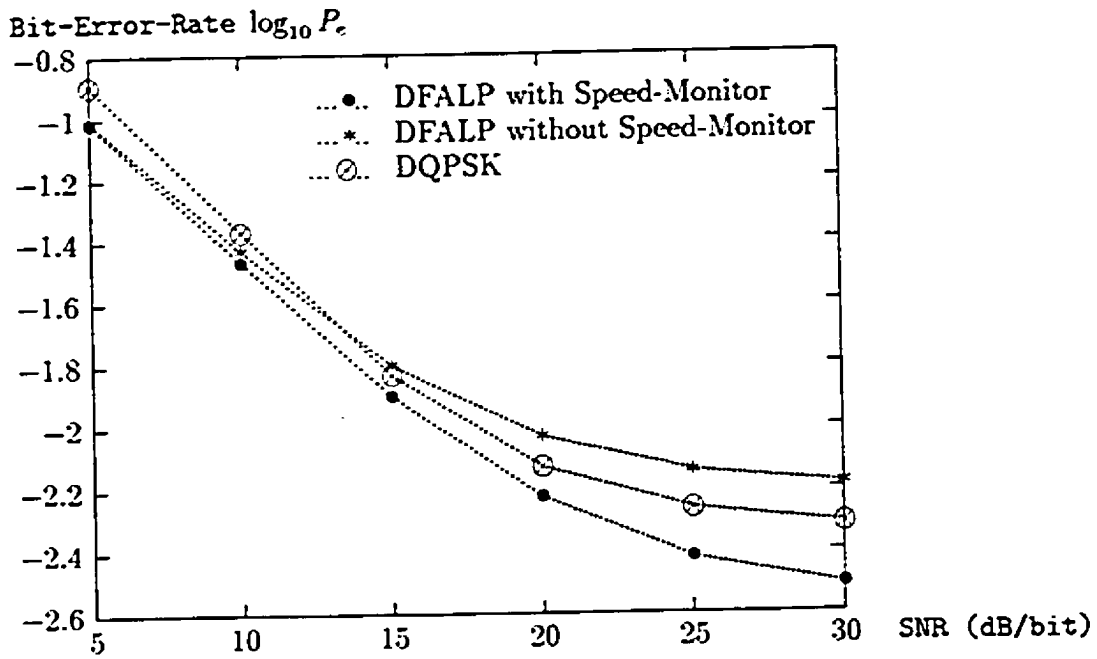
Figure 6.15: *The BER of the DFALP with or without a vehicle speed monitor and DQPSK.* $K_t = 5$. *The mobile speed changes from* 60 *km/hour to* 100 *km/hour.* $f_s = 6000$ *symbols/second.* $f = 5$ *ghz. The fading channel is Rician (K = 4 dB).*

tap weights. Figure 6.15 shows that when the vehicle speed increases from 60 km/hour to 100 km/hour, the performance degenerates quite significantly, particularly if the SNR is greater than 12 dB (even though its performance is not worse than the SADD according to Figures 6.5 and 6.6).

It is therefore recommended that a vehicle-speed monitor may be used if a larger complexity can be afforded by certain facilities such as base stations and large communications vehicles in order to achieve high quality communication. In other applications, where low complexity is desirable, e.g., in handsets of cellular telephone communications, it is recommended that the filter orders and tap weights are designed for a vehicle speed which is slightly larger than the most frequently-used speed. In the latter case, a vehicle-speed monitor is not required and the filter orders and LPF tap-weights are not adjusted even though the vehicle speed has changed.

206

6.4 DFALP Algorithm for Two-Beam Fading Channels

In a number of applications the frequency-selective fading channel may be more realistic. In this section we will briefly address the problem of estimating the frequency nonselective fading channels. We will provide two possible solutions to the problem without much elaboration. A detailed study is required in the future to further justify the proposals presented in this section.

For the frequency selective channel, a serious problem is the intersymbol interference (ISI) if it is not well equalized [98] or compensated for using some estimation technique. For a very slow fading channel, it is possible to approach this problem using a conventional adaptive equalizer with a fast tap-weight adaptation algorithm [25, 26, 55].

However, the conventional equalizer does not work well [25, 26, 55] in a fading channel whose normalized fading bandwidth is not less than 0.003, a situation for a typical cellular telephone channel involving fast moving vehicles or a satellite channel (see Figures 5.1-5.3). For this case a simple solution to the problem is to combine a conventional adaptive equalizer with the DFALP algorithm. The equalizer serves to decrease the ISI to an extent that its effect is not serious. Then the DFALP algorithm for nonselective fading channels is used to exactly estimate the channel gain to enhance the quality of the demodulation/decoding.

Another approach is to generalize the DFALP algorithm to track two-beam fading channels. We will now present this proposal. A two-beam fading channel is a special case of the frequency selective fading channels presented in section 5.1.2 when $l = 1$. The received signal may be written as

$$y_k = c_{0k}x_k + c_{1k}x_{k-1} + n_k \qquad (6.4.1)$$

where c_{0k} is the channel gain of the first equivalent beam and c_{1k} is the channel gain of the second equivalent beam. By an equivalent beam we mean that it is the output of

207

the actual beams passed through a linear matched filter and a symbol rate sampler. At the time $k - 1$ the received signal is

$$y_{k-1} = c_{0,k-1} x_{k-1} + c_{1,k-1} x_{k-2} + n_{k-1}. \qquad (6.4.2)$$

For a number of applications it is reasonable to assume that

$$c_{0k} \approx c_{0,k-1} \qquad c_{1k} \approx c_{1,k-1} \qquad (6.4.3)$$

which implies that the channel change from the time $k - 1$ to k is minimal. This assumption is not unrealistic for both slow and fast fading channels (see Figures 5.1-5.3). By using the above approximation, Eqs.(6.4.1) and (6.4.2) may be written as

$$\begin{pmatrix} y_k \\ y_{k-1} \end{pmatrix} = \begin{pmatrix} x_k & x_{k-1} \\ x_{k-1} & x_{k-2} \end{pmatrix} \begin{pmatrix} c_{0k} \\ c_{1k} \end{pmatrix} + \begin{pmatrix} n_k \\ n_{k-1} \end{pmatrix}. \qquad (6.4.4)$$

Following the DFALP algorithm for flat fading channels, the predicted channel gain at the time k is

$$\hat{c}_{0k} = \sum_{i=1}^{N} b_{0i}^* \tilde{c}_{0,k-i} = \vec{b}_0(k)^H \vec{\tilde{c}}_0(k) \qquad (6.4.5)$$

$$\hat{c}_{1k} = \sum_{i=1}^{N} b_{1i}^* \tilde{c}_{1,k-i} = \vec{b}_1(k)^H \vec{\tilde{c}}_1(k) \qquad (6.4.6)$$

where b_{0i} and b_{1i} are coefficients of the linear predictors for the first and second equivalent beams respectively. $\tilde{c}_{0,k-i}$ and $\tilde{c}_{1,k-i}$ are the past channel gains known from training symbols or decision feedback. The above equations generalize Step 1.1 of the DFALP algorithm presented in Section 6.1.

Then the data symbol at the time k may be estimated according to Eq.(6.4.1) as

$$\hat{x}_k = (y_k - \hat{c}_{1k} \bar{x}_{k-1})/\hat{c}_{0k} \qquad (6.4.7)$$

where $\bar{x}_{k-1}$ is the past detected data symbol or training symbol at the time $k - 1$. The above equation is a generalization of Step 1.2 of the DFALP.

Naturally, a tentative decision results in a possibly better estimate $\bar{x}_k$ of x_k given by

$$|\hat{x}_k - \bar{x}_k| = \min_{x_k \in D} |\hat{x}_k - x_k| \qquad (6.4.8)$$

208

where D is the signal constellation. The above equation generalizes Step 1.3 of the DFALP.

From Eq.(6.4.4) it is easily seen that a more accurate channel estimate may be obtained using

$$\begin{pmatrix} \bar{c}_{0k} \\ \bar{c}_{1k} \end{pmatrix} \approx \begin{pmatrix} \bar{x}_k & \bar{x}_{k-1} \\ \bar{x}_{k-1} & \bar{x}_{k-2} \end{pmatrix}^{-1} \begin{pmatrix} y_k \\ y_{k-1} \end{pmatrix} \tag{6.4.9}$$

where $\bar{x}_j$ are either decision-feedback data symbols or training symbols. It is assumed that

$$\bar{x}_k \bar{x}_{k-2} \neq \bar{x}_{k-1}^2. \tag{6.4.10}$$

Similar to Step 1.4 of the DFALP, a thresholding is used to check if

$$|\hat{c}_{0k} - \bar{c}_{0k}| \geq \beta \quad or \quad |\hat{c}_{1k} - \bar{c}_{1k}| \geq \beta \tag{6.4.11}$$

If the threshold is exceeded, it should be switched to the original predicted value $\hat{c}_{0k}$ or $\hat{c}_{1k}$ respectively.

Step 2 of the DFALP is easily generalized to

$$\vec{b}_0(k+1) = \vec{b}_0(k) + \mu(\bar{c}_{0k} - \hat{c}_{0k})^* \vec{\bar{c}}_0(k) \tag{6.4.12}$$

$$\vec{b}_1(k+1) = \vec{b}_1(k) + \mu(\bar{c}_{1k} - \hat{c}_{1k})^* \vec{\bar{c}}_1(k) \tag{6.4.13}$$

which update the linear predictor coefficients.

Then the third step is to low-pass-filter the estimated channel-gains using

$$\bar{c}_{0,k-D_f} = \sum_{i=0}^{2D_f} h_i \bar{c}_{0,k-i} \tag{6.4.14}$$

$$\bar{c}_{1,k-D_f} = \sum_{i=0}^{2D_f} h_i \bar{c}_{1,k-i} \tag{6.4.15}$$

which reduce the noise in the estimate.

Finally we output $\bar{c}_{i,k-D_f}$ and feedback $\bar{c}_{ik}$.

209

The above steps generalize the DFALP algorithm for flat fading channels. It is noted that following the same procedures the DFALP algorithm can be generalized to track other frequency selective fading channels.

It is noted that the frequency selective fading channel is essentially a time-varying finite-impulse-response (FIR) filter. Adaptive equalization is the conventional way to approach time-invariant or slowly time-varying FIR channels. An equalizer is in essence an approximation to the inverse FIR channel filter. Since a fading channel does not necessarily vary very slowly (see Figures 5.1-5.3), a conventional equalizer then has difficulty to adapt itself fast enough to follow the rapidly varying fading (FIR) channel. Even using training symbols and existing fast algorithms, for example the recursive least square algorithm, an adaptive equalizer takes at least 50 samples to converge [98] and therefore is too slow to track a fading channel with reasonable specifications (vehicle speed, symbol rate and wave frequency). On the other hand, the DFALP algorithm tries to predict each current tap weight of the time-varying FIR channel using the past known or decision-feedbacked tap weights. According to the well-known linear prediction theory [97], the only condition required for a linear predictor to work satisfactorily is that the covariance matrix of the random sequence be stationary or slowly time-varying. The value of the tap weights can be varying arbitrarily with time subject to the above condition. Based upon the above argument, it is seen that the DFALP algorithm may be a more viable approach to tracking the time-varying fading channel.

Chapter 7

CONCLUSION

In the final chapter we summarize the results in this thesis and propose future research directions for both change-detection and fading channel identification.

7.1 Change Detection

7.1.1 Major Results

In this thesis, a major result is that a new procedure, referred to as the minimum asymptotic risk (MAR) procedure, was proposed to quickly detect an abrupt change in random signals (Section 3.2). The new procedure is based upon an optimum procedure derived from Bayesian decision theoretic analysis. Theoretically, it was shown that under certain asymptotic conditions the new procedure's test statistic converges almost surely to the optimum decision statistic, and that the MAR decision rule is asymptotically equivalent to the optimum rule (Section 3.3). Extensive simulations were conducted to compare the performance of the MAR procedure to three well-known existing procedures described in Section 2.2, i.e., CUSUM, GRS and FSS, where a modified Shiryayev criterion introduced in Section 3.1 was used. It is shown that the MAR procedure has both a smaller

211

average delay and a larger false alarm average run length than CUSUM. FSS and a special case of the GRS procedure for a given false alarm probability. The MAR procedure also compares favorably with GRS in a number of situations. In the simulations both Gaussian and non-Gaussian distributions were used (Section 3.5). Bounds for the false alarm probability of the MAR, CUSUM and GRS procedures were derived in Section 3.6. It was shown that the bounds are tight for large parameter changes. The almost sure termination of each procedure was also proved.

In Chapter 4, the change detection procedures were extended to the case of unknown parameters, and non-i.i.d. observations. For change detection with unknown parameters, a multiple-hypothesis (MH) method was combined with the change-detection procedure discussed in Chapter 3. A number of properties of the MH approach were studied in Section 4.1, particularly convergence and the number of presumed parameters needed. For non-i.i.d. observations, a major approach to the change detection problem is to use an approximate CUSUM (A-CUSUM) statistic proposed by Blostein [17]. In Section 4.2 the A-CUSUM statistic was combined with MAR in order to address the large computational complexity problem. Simulations were conducted for some of the schemes proposed in this chapter.

Another important contribution in this thesis is a generalization of the well-known sequential probability ratio test (SPRT) to nonstationary situations (NSPRT). A generalized Wald's lower bound was provided. Using Bayesian decision theoretic analysis, it was proved that the NSPRT is optimal in the classical sense of Wald and Wolfowitz. Although optimal threshold determination remains an unsolved problem, a number of interesting properties of the optimum time-varying decision thresholds were provided in Section 4.2.

<div align="center">212</div>

7.1.2 Future Research in Change Detection

An interesting and practical problem in change detection is how to maintain a constant false alarm probability over the different possible change times in the random sequence. This problem has been encountered in the mobile communications application discussed in Section 6.3, where the vehicle speed is monitored by detecting the times of power drop and recovery of the fading channel. If the power drop and recovery occur relatively infrequently, then the actual change time is increased, and the thresholds for CUSUM and MAR must be changed in order to avoid an increase in false alarm rate. The false alarm probability bounds derived in Section 3.6 suggest that the threshold should be chosen to be time-varying, and this idea was used in our vehicle speed monitor. However, the adaptive thresholding algorithm lacks sound theoretical justification.

Much work remains to be done in change detection for dependent observations. Existing methods including those proposed in this thesis are mostly ad hoc. The trade-off between computational complexity and performance has not been carefully studied.

Finally, the fundamental problem of change-detection for i.i.d. observations has not been completely solved yet. Both theoretical and experimental studies show that none of the existing procedures including the new MAR procedure proposed in this thesis is nonasymptotically optimal under practical criteria. Therefore it may still be still possible to discover a more powerful algorithm.

7.2 Fading Channel Identification

7.2.1 Major Results

An important contribution of this thesis was to use decision feedback and adaptive linear prediction (DFALP) to track the phase and amplitude of fading channels (Section 6.1). This work was motivated by the observation that a globally optimum detection scheme proposed in [76] is approximately equivalent to a coherent detector and an MMSE

213

channel estimator as described in Section 5.2. The DFALP provides a good channel estimate for the coherent detector.

The performance of the DFALP algorithm was studied both analytically and experimentally in Section 6.2. An analytical estimate of the mean-square-error (MSE) and bit-error-rate (BER) of the DFALP algorithm was derived. An estimate of the MSE of the SADD algorithm was also derived analytically and compared to that of DFALP. This analytical result also explains an experimentally observed irreducible interpolation error problem of SADD found in Section 6.2. Using extensive simulations it has been shown that the new algorithm combined with CPSK outperforms its DPSK counterpart and the existing fading channel tracking algorithm. The parameter design problem for the DFALP algorithm was carefully studied in Section 6.1 and near-to-optimum filter orders and tap-weights are recommended.

In Section 6.3, the change detection procedures described in Chapter 3 were adapted to monitor the vehicle speed by detecting the power drop and recovery of a fading channel. The vehicle speed monitor was then used to adapt the fading channel tracking algorithms to maintain a high level of performance over different nonstationary situations. The performance improvement by using the vehicle speed monitor was demonstrated.

The extension of the new adaptive identification algorithm to frequency selective fading channels was also discussed briefly in Section 6.4.

7.2.2 Future Research in Fading Channel Identification

Chapters 5 and 6 have been devoted to developing the DFALP algorithm for tracking fading channels. Since the fading channel tracker provides both the channel phase and amplitude information, it is expected that combining DFALP with coding, interleaving and Viterbi soft decoding would yield a significant performance gain over the conventional DPSK scheme as reported in [57, 58]. However, this conjecture has to be confirmed

214

performance of trellis coded, interleaved CPSK with imperfect channel state information, while the performance of the same CPSK with perfect channel state information has been studied in [80, 54].

High-level quadratic amplitude modulation (QAM) has not been seriously considered in a fading channel environment since it has been perceived that a good estimate of the channel phase and amplitude is difficult to obtain, even though high-level QAM is bandwidth-efficient and has better performance than PSK if the channel state information is available [2]. Since the proposed DFALP channel tracker provides a good estimate of both channel phase and amplitude, it appears feasible to apply high-level QAM to a fading channel environment as is suggested in [2]. However, this problem should be studied carefully for the situation of imperfect channel state information.

Finally, Section 6.4 extended the DFALP channel estimation algorithm to frequency-selective fading channels. The performance of DFALP against existing schemes such as conventional equalizers and equalizer-type adaptive estimators [98] is a topic for future investigation.

Bibliography

[1] Abend, K. and B. D. Fritchman (1970), 'Statistical detection for communication channels with intersymbol interference,' *Proceedings of IEEE,* May, pp.779-785.

[2] Aghamohammadi, A. and H. Meyr (1990) 'On the error probability of linearly modulated signals on Rayleigh frequency-flat fading channels,' *IEEE Transactions on Communications,* Vol.38, No.11, November, pp.1966-1970.

[3] Alles, M. and S. Pasupathy (1992), 'Binary communication on the indoor wireless channel,' *Proceedings of the Biennial Symposium on Communications,* Kingston, Ontario, Canada, May 1992, pp.1-4.

[4] Andre-Obrecht, R. (1988) 'A new statistical approach for the automatic segmentation of continuous speech signals,' *IEEE Transactions on Acoustic, Speech, and Signal Processing,* Vol.ASSP-36, pp.29-40.

[5] Austin, M.E. (1967), 'Decision-feedback equalization for digital communication over dispersive channels,' MIT Lincoln Laboratory, Lexington, Mass., Tech. Report No.437, August.

[6] Basseville. M. etc. (1981) 'Edge detection using sequential methods for change in level.' *IEEE Transactions on ASSP,* 29:24-50.

[7] Basseville, M. (1988) 'Detecting changes in signals and systems: a survey,' *Automatica,* Vol.24, No.3, pp.309-326.

[8] Basseville, M. and A. Benveniste (1983), 'Sequential detection of abrupt changes in spectral characteristics of digital signals,' *IEEE Transactions on Information Theory*, Vol.IT-29, No.5, September, pp.709-724.

[9] Basseville, M. and A. Benveniste(ed.) (1986), *Detection of Abrupt Changes in Signals and Dynamical Systems*, Springer-verlag, New York.

[10] Basseville, M., A. Benveniste, G. Moustakides and A. Rougee (1987), 'Detection and diagnosis of changes in eigenstructure of nonstationary multivariable systems,' *Automatica*, Vol.23, No.4, pp.479-487.

[11] Bello, P.A. (1963), 'Characterization of randomly time-variant linear channels,' *IEEE Transactions on Communications Systems*, Vol. CS-11, pp.360-393.

[12] Benveniste, A, M. Basseville and G. M. Moustakides (1987), 'The asymptotic local approach to change detection and model validation,' *IEEE Transactions on Automatic Control*, Vol.AC-32, No.7, July, pp.583-592.

[13] Berger, J.O. (1985), *Statistical Decision Theory and Bayesian Analysis*, 2ed ed., Springer-verlag, New York.

[14] Biglieri, E. D. Divsalar, P.J. McLane and M. K. Simon (1991), *Introduction to Trellis-Coded Modulation with Applications*, Macmillan, New York.

[15] Billingsley, P. (1979), *Probability and Measure*, Wiley, New York.

[16] Blostein, S. D. (1988), *A Sequential Hypothesis Testing Approach to Detecting Small, Moving Objects in Image Sequences*. PhD thesis, University of Illinois at Urbana-Champaign.

[17] Blostein, S. D. (1991), 'Quickest detection of a time-varying change in distribution,' *IEEE Transactions on Information Theory*, Vol.IT-37, No.4, July, pp.1116-1122.

[18] Blostein, S. D. and T.S. Huang (1991), 'Detecting small, moving objects in image sequences using sequential hypothesis testing.' *IEEE Transactions on Signal Processing*, Vol.39, No.7, July, pp.1611-1629.

[19] Blostein, S.D. and Y. Liu (1991), 'On sequentially detecting an abrupt change in a stochastic process,' *Proceedings of IEEE International Conference on ASSP*, May, 1991, Toronto, pp.1341-4.

[20] Bojdecki, T. (1979), 'Probability maximizing approach to optimal stopping and its application to a disorder problem,' *Stochastics*, Vol. 3, pp.61-71.

[21] Bouvet, M. (1986), 'Bayesian quickest signal detection in a discrete-time observation,' *IEEE Transactions on Aerospace and Electronic Systems*, Vol.AES-22, No.2, March, pp.170-6.

[22] Chan, T. and P. Ho (1992), 'Bit-error probability of uncoded QPSK transmitted over a 2-ray frequency selective Rayleigh fading channel.' *Proceedings of IEEE Supercom/ICC*, Chicago, June.

[23] Chow, Y.S. and H. Robbins (1963), 'On optimal stopping rules,' *Wahrscheinlichkeitstheorie*, Vol.2, pp.33-49.

[24] Chow, Y.S., H. Robbins and D. Siegmund (1971), *Great Expectations: the Theory of Optimal Stopping*, Houghton Mifflin Company, Boston.

[25] Clark, A.P. and R. Harun (1986), 'Assessment of Kalman-filter channel estimator for an HF radio link,' *IEE Proceedings*, Vol.133, Part F, No.6, Oct., pp.513-521.

[26] Clark, A.P. and S. Hariharan (1989), 'Adaptive channel estimator for an HF radio link,' *IEEE Transactions on Communications*, Vol.37, No.9, Sept., pp.918-926.

[27] Cochlar, J. and I. Vrana (1978), 'On the optimum sequential test of two hypotheses for statistically dependent observations,' *Kybernetika*, Vol. 14, pp.57-69.

218

[28] D'Angelo. H. (1970), *Linear Time-Varying Systems*, Allyn and Bacon, Inc., Boston.

[29] Dodge, H.F. and H.G. Romig (1929), 'A method of sampling inspection,' *Bell System Technical Journal*, 8:613-631.

[30] Dudewicz, E.J. and S. N. Mishra(1988), *Modern Mathematical Statistics*, John Wiley and Sons, New York.

[31] Ewan, W.D. (1963), 'When and how to use CUSUM charts,' *Technometrics*, Vol.5, No.1, February, pp.1-22.

[32] Falconer, D. D. and L. Ljung (1978), 'Application of fast Kalmam estimation to adaptive equalization,' *IEEE Transactions on Communications*, Vol. Com-26, October, pp.1439-46.

[33] Ferguson, T. S. (1967), *Mathematical Statistics: a decision theoretic approach*. Academic Press. New York.

[34] Forney, G. D. Jr (1972), 'Maximum-likelihood sequence estimation of digital signaling in the presence of intersymbol interference,' *IEEE Transactions on Information Theory*, Vol. IT-18, May, pp.363-378.

[35] Francesco, R. J. D. (1990), 'Real-time speech segmentation using pitch and convexity jump models: Application to Variable Rate Speech Coding,' *IEEE Transactions on Acoustic, Speech, and Signal Processing*, Vol. ASSP-38, No.5 May, pp.741-748.

[36] Gagliardi, R.M. and I. S. Reed (1965), 'On the sequential detection of emerging targets,' *IEEE Transactions on Information Theory*, April, pp.260-263.

[37] Gardner, F.M. (1977), 'Handup in phase lock loops,' *IEEE Transactions on Communications*, Vol.25, No.10, Oct., pp.1210-1214.

[38] Ghosh, B. K. (1970), *Sequential Tests of Statistial Hypotheses*, Reading, Mass.: Addison-Wesley, pp.30-31.

[39] Girshick, M.A. and H. Rubin (1952), 'A Bayes approach to a quality control model.' *Annals of Mathematical Statistics*, vol.23, 114-125.

[40] Godard, D.N. (1974), 'Channel equalization using a Kalmam filter for fast data transmission,' *IBM Journal of Research and Development*, Vol. 18, May, pp.267-273.

[41] Goel, A. L. (1968), *A Comparative and Economic Investigation of Shewhart $\bar{X}$-Chart and Cumulative Sum Control Charts*, PhD Dissertation, University of Wisconsin-Madison Department of Mechanical Engineering.

[42] Goel, A. L. (1982), 'Cumulative sum control charts,' in L. Johnson and S. Kotz (ed. chief), *Encyclopaedia of Statistical Science, Vol. 2*. John Wiley and Sons, Inc.

[43] Gottfried, B. S. and J. Weisman (1973), *Introduction to Optimization Theory*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

[44] Govindarajulu, Z. (1975), *Sequential Statistical Procedure*, Academic Press, New York.

[45] Haddad, A.H. (1986), 'On detection-estimation schemes for uncertain systems,' in I. F. Blake and H. V. Poor (eds.) (1986), *Communications and Networks*. Springer-Verlag. New York.

[46] Haeb, R. and H. Meyr (1989), 'A systematic approach to carrier recovery and detection of digitally phase modulated signals on fading channels,' *IEEE Transactions on Communications*, Vol.37, No.7, July, pp.748-754.

[47] Hajek, J. and Z. Sidak (1967), *Theory of Rank Tests*, Academic, New York.

[48] Haykin, S.S.(ed.) (1976), *Detection and Estimation: with application to radar*. Dowden, Hutchinson and Ross, Inc.

[49] Haykin, S.S. (1991), *Adaptive Filter Theory*. Englewood Cliffs, NJ : Prentice Hall.

[50] Himmelblau, D.M. (1978), *Fault Detection and Diagnosis in Chemical and Petro-chemical Processes*, Elsevier Scientific Publishing Co., New York.

[51] Hines, W. G. S. (1976), 'A simple monitor of a system with sudden parame-ter changes,' *IEEE Transactions on Information Theory*, Vol.IT-22, No.2 March, pp.210-216.

[52] Hines, W. G. S. (1976), ' Improving a simple monitor of a system with sudden parameter changes,' *IEEE Transactions on Information Theory*, Vol.IT-22, No.3 July, pp.496-499.

[53] Ho, P. and D. Fung (1991), 'Error performance of multiple symbol differential de-tection of PSK signals transmitted over correlated Rayleigh fading channels,' *Pro-ceedings of IEEE ICC*, pp.568-574.

[54] Ho, P. and D. Fung (1992), 'Error performance of interleaved trellis-coded PSK modulation in correlated Rayleigh fading channels,' *IEEE Transactions on Com-munications*, vol.40, No12, December, pp.1800-1809.

[55] Hsu, F.M. (1982), 'Square root Kalman filtering for high-speed data received over fading dispersive HF channels,' *IEEE Transactions on Information Theory*, Vol.28, No.5, Sept., pp.753-763.

[56] Huber, P.J. (1981), *Robust Statistics*, Wiley, New York.

[57] Irvine, G.T. and P.J. McLane (1992), 'Symbol-aided plus decision-directed reception for PSK/TCM modulation on shadowed mobile satellite fading channels,' *IEEE Journal on Selected Areas in Communications*, Vol.10, No.8, October, pp.1289-1299.

[58] Irvine, G.T. (1990), *Symbol-Aided Plus Decision-Directed Phase Estimation for Shadowed Mobile Satellite Fading Channels*, MSc. Thesis, Queen's University, Kingston, On, Canada.

[59] Jakes, W. (ed.) (1974), *Microwave Mobile Communications* , John Wiley and Sons, New York.

[60] Kam, P.Y. (1991), 'Optimal detection of digital data over nonselective Rayleigh fading channel with diversity reception,' *IEEE Transactions on Communications,* Vol.39, No.2, Feb., pp.214-219.

[61] Kolmogorov, A. N. (1956), *Foundations of the Theory of Probability,* Chelsea, New York.

[62] Lee, W.Y.C. (1982), *Mobile Communications Engineering,* McGraw-Hill, New York.

[63] Lehmann, E.L. (1986), *Testing Statistical Hypotheses,* Wiley: New York.

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

[65] Lin, S. and D.J. Costello Jr. (1983), *Error Control Coding,* Prentice Hall, Englewood Cliffs, New Jersey.

[66] Ling, F. and J. G. Proakis (1985), 'Adaptive lattice decision-feedback equalizers-their performance and application to time-variant multipath channels,' *IEEE Transactions on Communications,* Vol. Com-33, No.4, April, pp.348-356.

[67] Liporace, L.A. (1971), 'Variance of Bayes estimates,' *IEEE Transactions on Information Theory,* Vol. IT-17, No.6, November, pp.665-9.

[68] Liu, W. and P. Ho (1992), 'Differential Detection of PSK signals in frequency selective Rayleigh fading channel.' *Proceedings of the Biennial Symposium on Communications,* Kingston, Ontario, Canada, May 1992, pp.383-386.

[69] Liu, Y. and S. D. Blostein (1990), 'Optimality of the sequential probability ratio test for non-stationary observations.' *Proceedings of International Symposium on Information Theory and its Applications.* Hawaii, U.S.A., pp.199-202.

[70] Liu, Y. and S. D. Blostein (1992a), 'Optimality of the sequential probability ratio test for non-stationary observations,' *IEEE Transactions on Information Theory.* January, Vol.32, No.1, pp.177-182.

[71] Liu, Y. and S. D. Blostein (1992b), 'Quickest detection of an abrupt change in a random sequence with finite change-time,' *IEEE Transactions on Information Theory.* (submitted)

[72] Liu, Y. and S. D. Blostein (1992c), 'A modified Bayesian procedure for quickest signal change detection,' *Proceedings of the Biennial Symposium on Communications,* Kingston, Ontario, Canada, May 1992, pp.406-409.

[73] Liu, Y. and S. D. Blostein, 'Quickest detection of an abrupt change in a random sequence with finite change-time,' *Proceedings of IEEE International Symposium on Information Theory,* Jan. 17-22, 1993, San Antonio, TX, USA.

[74] Liu, Y. and S.D. Blostein (1992), 'Maximum likelihood sequence estimation for Rayleigh fading channels,' *IEEE Transactions on Communications.* (submitted)

[75] Ljung, L. and S. Gunnarsson (1990), 'Adaptation and tracking in system identification-a survey,' *Automatica.* Vol.26, No.1, pp.7-21.

[76] Lodge, J.H. and M.L. Moher (1990), *Maximum likelihood sequence estimation of CPM signals transmitted over Rayleigh flat-fading channels. IEEE Transactions on Communications,* Vol.38, No.6, June, pp787-794.

[77] Lorden, G. (1971), 'Procedures for reacting to a change in distribution.' *Annals of Mathematical Statistics,* 42(6): 1897-1908.

[78] Magill, D.T. (1965), 'Optimal adaptive estimation of sampled stochastic processes,' *IEEE Transactions on Automatic Control,* Vol. AC-10, pp.434-439.

[79] Mazo, J.E. (1991), 'Exact matched filter bound for two-beam Rayleigh fading'. *IEEE Transactions on Communications*, Vol.39, No.7, July, pp.1027-1030.

[80] McLane, P.J., P.H. Wittke, P.K.M. Ho and C. Loo (1988), 'PSK and DPSK trellis codes for fast fading, shadowed mobile satellite communication channel,' *IEEE Transactions on Communications*, Vol.36, No.11, Nov., pp.1242-5.

[81] Mehra, R.K. and J. Peschon (1971) 'An innovation approach to fault detection and diagnosis in dynamic systems,' *Automatica*, Vol.7, pp.637-640.

[82] Moher, M.L. and J.H. Lodge (1989), *TCMP-A modulation and coding strategy for Rician fading channels. IEEE Journal on Selected Areas in Communications*, Vol.7, No.9, December, pp.1347-1355.

[83] Monsen, P. (1971), 'Feedback equalization for fading dispersive channels,' *IEEE Transactions on Information Theory*, Vol. IT-17, Jan., pp.56-64.

[84] Morf, M., B. Dickinson, T. Kailath and A. Vieira (1977), 'Efficient solution of covariance equations for linear prediction,' *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-25, October, pp.429-433.

[85] Moustakides, G. V. (1986a), 'Optimal stopping times for detecting changes in distributions.' *Annals of Statistics*, Vol.14, No.4, pp.1379-1387.

[86] Moustakides, G. V. (1986b), 'The problem of diagnosis with respect to physical parameters for changes in structures.' *Research Report IRISA*, No.295.

[87] Moustakides, G. V. and A. Benveniste (1986), 'Detecting changes in the AR parameters of a nonstationary ARMA process.' *Stochastics*, Vol.16, pp.137-155.

[88] Newbold, P.M. and Y.-C. Ho (1968), 'Detection of changes in the characteristics of a Gauss-Markov process.' *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-4, No.5 September, pp.707-718.

[89] Page, E. S. (1954). 'Continuous inspection schemes.' *Biometrika*, 41:100-115.

[90] Papantoni-Kazakos, P. (1979). 'Algorithms for monitoring changes in quality of communication links.' *IEEE Transactions on Communications*, COM-27:682-692.

[91] Pelkowitz, L. (1987), 'The general discrete time disorder problem.' *Stochastics*, Vol. 20, pp.89-110.

[92] Pelkowitz, L. and S. C. Schwartz (1987), 'Asymptotically optimum sample size for quickest detection,' *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-23, No.2 March, pp.263-272.

[93] Perriot-Mathonna, D.M. (1984). 'Improvements in the application of stochastic estimation algorithms—parameter jump detection.' *IEEE Transactions on Automatic Control*, Vol. AC-29, No.11, Nov., pp.962-969.

[94] Pollak, M. and D. Siegmund (1985), 'A diffusion process and its applications to detecting a change in the drift of Brownian motion.' *Biometrika*, Vol.72, No.2, pp.267-80.

[95] Pollak, M. (1985), 'Optimal detection of a change in distribution.' *Annals of Statistics*, Vol.13, No.1, pp.206-227.

[96] Pollak, M. and D. Siegmund (1991), 'Sequential detection of a change in a normal mean when the initial value is unknown.' *Annals of Statistics*, Vol.19, No.1, pp.394-416.

[97] Poor, H.V. (1988), *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York.

[98] Proakis, J.G. (1989), *Digital Communications*, McGraw-Hill, second ed..

[99] Reudink, D.O. (1974), 'Properties of mobile radio propagation above 400 MHz Band,' *IEEE Transactions on Vehicular Technology*, Vol. VT-23, pp.143-159, Nov.

225

[100] Rice, M.D. (1991), *Adaptive Error Control over Slowly Varying Channels*, PhD thesis, School of Electrical Engineering, Georgia Institute of Technology.

[101] Ritov, Y. (1990), 'Decision theoretic optimality of the CUSUM procedure,' *Annals of Statistics*, Vol.18, No.3, 1464-9.

[102] Roberts, S.W. (1966), 'A comparison of some control chart procedure,' *Technometrics*, Vol.8, No.3, August.

[103] Samuel, E. (1970), 'Randomized sequential tests,' *Journal of the American Statistical Association*, Vol.65, 431-7.

[104] Sanderson, A.C. and J. Segen (1980), 'Hierarchical modelling of EEG signals,' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, No.2, pp.405-414.

[105] Sato, Y. (1975), 'A method of self-recovering equalization for multilevel amplitude-modulation systems,' *IEEE Transactions on Communications*, Vol. Com-23, June, pp.679-682.

[106] Satorius, E.H. and J.D. Pack (1981), 'Application of least squares lattice algorithms,' *IEEE Transactions on Communications*, Vol. Com-29, February, pp.899-905.

[107] Schwartz, M., W. R. Bennett and S. Stein 1966, *Communication Systems and Techniques*, McGraw-Hill, New York.

[108] Segen, J. and A.C. Sanderson (1980), 'Detecting change in a time-series,' *IEEE Transactions on Information Theory*, Vol. IT-26, No.2, March, pp.249-255.

[109] Shewhart, W.A. (1931), *The Economic Control of quality of a Manufacctured Product*, Van Nostrand.

[110] Shiryayev, A.N. (1963), 'On optimum methods in quickest detection problems.' *Theory of Probability and its Applications*, Vol. 8, No.1.

[111] Shiryayev, A.N. (1978), *Optimal Stopping Rules*, Springer-Verlag, New York.

[112] Shiryayev, A.N. (1984), *Probability*, Springer-Verlag, New York.

[113] Siegmund, David (1985), *Sequential Analysis*. Springer-Verlag, New York.

[114] Tantaratana, S. (1986), 'Sequential detection of a positive signal,' in I. F. Blake and H. V. Poor (eds.) (1986), *Communications and Networks*. Springer-Verlag. New York.

[115] Tugnait, J.K. (1982a) 'Detection and estimation for abruptly changing systems,' *Automatica*, Vol.18, pp.607-615.

[116] Tugnait, J.K. (1982b) 'Adaptive estimation and identification for discrete systems with Markov jump parameters,' *IEEE Trans. Automatic Control*. Vol.27, pp.1054-1064.

[117] Van Trees, H.L. (1968), *Detection, Estimation and Modulation Theory*, Wiley, New York.

[118] Viswanadham, N., V.V.S. Sarma and M.G. Singh, *Reliability of Computer and Control Systems*, North-Holland, New York, 1987, chapter 3.

[119] Wald, A. (1947), *Sequential Analysis*. John Wiley and Sons.

[120] Wald, A and J. Wolfowitz (1948), Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics*, 53:581-586.

[121] Widrow, B. and M.E. Hoff, Jr. (1960), 'Adptive switching circuits,' *IRE WESCON Conv. Rec.*, Part 4, pp.96-104.

[122] Widrow, B., J.M. McCool and M. Ball (1975), 'The complex LMS algorithm' *Proceeding of IEEE* , Vol.63, pp.719-720.

[123] Widrow, B. and J.M. McCool (1976), 'A comparison of adaptive algorithms based on the methods of steepest descent and random search' *IEEE Trans. on Antennas and Propagation*, Vol.AP-24, Sept.

[124] Widrow, B. and S.D. Stearns (1985), *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, N.J.

[125] Wijsman, R.A. (1963), 'Existence, uniqueness and monotonicity of sequential probability ratio tests,' *Annals of Mathematical Statistics*, Vol.34, 1541-8.

[126] Willsky, A.S (1986), 'Detection of abrupt changes in dynamic systems,' in M. Basseville and A. Benveniste(ed.), *Detection of Abrupt Changes in Signals and Dynamical Systems*. 1986, Springer-Verlag, New York.

[127] Willsky, A.S and H.L. Jones (1976), 'A generalized likelihood ratio approach to the detection and estimation of jumps in linear syatems,' *IEEE Transactions on Automatic Control*, 21:108-112.

[128] Willsky, A.S (1976), 'A survey of design methods for failure detection in dynamic systems,' *automatica*, Vol.12, pp.601-611.

[129] Wittke, P.H., W.S. Smith and L.L. Campbell (1988), 'Infinite series of interference variables with Cantor-type distributions,' *IEEE Transactions on Information Theory*, Vol.34. No.6, November.

VITA

Yong Liu

- **EDUCATION**

  1987   MEng in Electrical Engineering, Shanghai Chiao Tung University, Shanghai, P.R. China.

  1984   BEng in Electrical Engineering, Shanghai Chiao Tung University, Shanghai, P.R. China.

- **AWARDS**

  1992-3   Dean's Award. Queen's University.

  1992-3   Queen's Graduate Award, Queen's University.

  1991-2   Queen's Graduate Fellowship, Queen's University.

  1990-1   Dean's Award, Queen's University.

- **EXPERIENCE**

  1990-93   Research Assistant, Canadian Institute for Telecommunications Research, Kingston, Ontario, Canada.

  1987-89   Engineer, Shanghai Telecom Equipment Co., Shanghai, P.R. China.

  1987-89   Instructor, Shanghai Chiao Tung University, Shanghai, P.R. China.

• PUBLICATIONS

Liu, Y. and S. D. Blostein, 'Optimality of the sequential probability ratio test for non-stationary observations,' *IEEE Transactions on Information Theory*, Vol.32, No.1, Jan. 1992, pp.177-182.

Liu, Y. and S. D. Blostein, 'A modified Bayesian procedure for quickest signal change detection,' *Proceedings of the Biennial Symposium on Communications*, Kingston, Ontario, Canada, May 1992, pp.406-409.

Liu, Y. and S. D. Blostein, 'A new procedure for quickest signal change detection and its application to fading channel change detection,' *Proceedings of the CITR Annual Research Conference*, Montreal, Canada, May 1992.

Blostein, S.D. and Y. Liu , 'On sequentially detecting an abrupt change in a stochastic process,' *Proceedings of IEEE International Conference on ASSP* , May, 1991, Toronto, pp.1341-4.

Liu, Y. and S. D. Blostein, 'Optimality of the sequential probability ratio test for non-stationary observations,' *Proceedings of International Symposium on Information Theory and its Applications*. Hawaii, U.S.A., pp.199-202, Dec. 1990.

Liu, Y. and S. D. Blostein, 'Quickest detection of an abrupt change in a random sequence with finite change-time,' *Proceedings of IEEE International Symposium on Information Theory*, Jan. 17-22, 1993, San Antonio, TX, USA.

Liu, Y. and S.D. Blostein, 'Maximum likelihood sequence estimation for Rayleigh fading channels,' *IEEE Transactions on Communications*, 1992. (submitted)

Liu, Y. and S. D. Blostein, 'Quickest detection of an abrupt change in a random sequence with finite change-time,' *IEEE Transactions on Information Theory*, 1992. (submitted)