MICROFILMED
BY
CENTRAL MICROFILM UNIT
PUBLIC ARCHIVES
OF
CANADA
OTTAWA, ONTARIO

MICROFILMÉ
PAR LE
SERVICE CENTRAL DU MICROFILM
ARCHIVES PUBLIQUES
DU
CANADA
OTTAWA, ONTARIO

24/1/74 DATE

REDUCTION 15X

L. Grosman OPERATOR
OPÉRATEUR

EXPOSURE
EXPOSITION 3.2

**NATIONAL LIBRARY
OF CANADA**

**CANADIAN THESES
ON MICROFILM**

**BIBLIOTHÈQUE
NATIONALE
DU CANADA**

**THÈSES CANADIENNES
SUR MICROFILM**

1 6 5 4 9

16549

NATIONAL LIBRARY
OTTAWA

BIBLIOTHÈQUE NATIONALE
OTTAWA

NAME OF AUTHOR.....Allan R. Smalley.....
TITLE OF THESIS.....Time Varying Applications of Practical Adaptive
.....Systems Using Gradient Algorithms.....
.....
UNIVERSITY.....Carleton.....
DEGREE FOR WHICH THESIS WAS PRESENTED.....Ph.D.....
YEAR THIS DEGREE GRANTED.....1973.....

Permission is hereby granted to THE NATIONAL LIBRARY
OF CANADA to microfilm this thesis and to lend or sell copies
of the film.

The author reserves other publication rights, and
neither the thesis nor extensive extracts from it may be
printed or otherwise reproduced without the author's
written permission.

(Signed) *A. R. Smalley*

PERMANENT ADDRESS:

183 MOUNTBATTEN AVE
.....
OTTAWA,
.....
ONTARIO K1H 5V6
.....

DATED *SEPT. 21,*19*73*

NL-91 (10-68)

TIME VARYING APPLICATIONS
OF
PRACTICAL ADAPTIVE SYSTEMS
USING
GRADIENT ALGORITHMS

by

ALLAN RALPH SMALLEY, B.SC. (ENG.), M.ENG.

A thesis submitted to the Faculty
of Graduate Studies in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy.

Carleton University
Ottawa, Ontario
September, 1973

The undersigned recommend to the Faculty
of Graduate Studies acceptance of the thesis

"Time Varying Applications of Practical
Adaptive Systems using Gradient Algorithms",

submitted by Allan Ralph Smalley, B.Sc. (Eng.), M.Eng.,
in partial fulfilment of the requirements for the
degree of Doctor of Philosophy.

.....

Thesis Supervisor

.....

Chairman, Division of Systems Engineering

Carleton University

Ottawa, Ontario

September, 1973

ABSTRACT

The design of communication or control systems involves considerable effort to obtain a system having the "best" performance as evaluated by a suitable criterion such as highest signal to noise ratio or minimum probability of error. That system having the best performance measurement may be termed the optimum system.

The difficulty of deriving optimum unconstrained systems for time varying applications and the resulting complexity of implementation of these optimum systems has led to the desire for much simpler adaptive systems that are capable of providing reasonable performance improvements with a minimum of complexity. The use of constrained systems, in which the form of the system is fixed and several parameters are then adaptively optimized, allows good suboptimum systems to be designed when the unconstrained solution may be impossible to obtain. In addition the complex unconstrained optimum system can often be implemented only by approximating it with a suboptimum system whose performance may not be very different from that of the constrained adaptive system. In fact the work of

this thesis suggests that in certain cases general adaptive solutions to problems (even of the constrained type) may yield no performance improvement over a fixed or simple adaptive system. However, it is generally necessary to deal with a specific problem using a trial and error process on the various adaptive techniques available for performance improvement.

This thesis examines some simple adaptive techniques that are suitable for on-line optimization of systems without introducing additional test signals or other methods which decrease the information rate of the system. The desire to use simple algorithms that do not require complex procedures has led to consideration of adaptive control by means of some of the gradient techniques with fixed step size parameters. These algorithms are studied both analytically and experimentally through the use of digital and hybrid computer simulations. Of particular interest is the tracking of time varying changes in the system environment.

The adaptive systems are constrained by fixing their form and selecting certain parameters to optimize while keeping the basic form intact. For example a first or second order low pass receive filter may

be selected and its bandwidth and sampling time optimized by the gradient methods. The selection of the parameters for optimization and the choice of performance criteria are most important. It is shown that proper selection of performance measures can reduce the parameter interaction effects and considerably simplify the adaptive processes needed.

The major example in the thesis is the use of adaptive receivers for digital Pulse Amplitude Modulation (PAM) systems with high intersymbol interference. In particular a method is derived for optimizing the analog prefilter in conjunction with the adaptive transversal equalizer part of the receiver. Some approximations are introduced into the algorithms, resulting in very simple hardware realizations of the adaptive loops for adjusting the prefilter bandwidth and sampling time. The system is then simulated on a hybrid computer.

In general this work shows that gradient techniques using fixed step size parameters yield very practical adaptive systems which can produce a significant improvement in performance over fixed systems.

ACKNOWLEDGEMENT

The author would like to express his appreciation to Dr. D.A. George for suggesting adaptive systems as a possible area of study, and for his consistent guidance through many fruitful discussions held during the preparation of this thesis.

This work would not have been possible without the financial help of a National Research Council Science Scholarship, and other assistance from the Government of Ontario and Carleton University.

The author especially wishes to thank his wife, Mary, for typing the draft and final versions of the thesis, and for her encouragement throughout the program.

CONTENTS

	Page
CHAPTER 1 - INTRODUCTION.....	1
1.1 Origin of Unknown and Time Varying Effects.....	3
1.2 General Adaptive System.....	4
1.3 Adaptive Receivers for Digital Communications.....	8
1.4 Maximization of the Performance Criterion.....	15
1.4.1 Solving for the Optimum Parameters.....	16
1.4.2 Suitability of Gradient Techniques.....	18
1.5 Related Work on Adaptive Systems.....	20
1.6 Summary.....	23
1.7 Thesis Outline.....	24
CHAPTER 2 - GRADIENT ALGORITHMS AND STOCHASTIC APPROXIMATION PROCEDURES.....	28
2.1 Some Common Gradient Algorithms..	31
2.1.1 The Robbins-Monro Algorithm.....	31
2.1.2 The Kiefer-Wolfowitz Algorithm...	33
2.1.3 A Modified Robbins-Monro Algorithm.....	35
2.1.4 Higher Order Algorithms.....	36
2.1.5 Quantized Gradient Algorithm.....	39
2.1.6 Tracking Algorithms and Time Varying Systems.....	40
2.2 Convergence for Time Invariant Systems.....	42
2.3 Convergence for the Quantized Gradient Algorithm.....	52
2.4 Time Varying Systems.....	54
2.5 Summary.....	63
CHAPTER 3 - ADAPTIVE TECHNIQUES FOR PAM COMMUNICATION RECEIVERS.....	65
3.1 The Basic PAM Communication Receiver.....	65
3.1.1 Choice of a Performance Criterion	69

CONTENTS (Cont'd)

	Page
3.1.2 Estimation of the Gradients.....	71
3.2 Rectangular Received Pulses and a Low Pass Receive Filter.....	75
3.2.1 Adaptive Adjustment of the Filter Gain.....	79
3.2.2 Adaptive Adjustment of the Sampling Time.....	84
3.2.3 Adaptive Adjustment of Receive Filter Bandwidth.....	89
3.2.4 Discussion of Computer Simulation Results.....	96
3.2.4.1 General Results.....	98
3.2.4.2 Mean Square Error and Maximum Signal to Noise Criteria.....	99
3.2.4.3 The Effects of Higher Order Algorithms.....	100
3.2.4.4 Some Preliminary Tracking Results	102
3.3 Second Order Received Pulses and a Low Pass Receive Filter.....	107
3.3.1 Adaptive Adjustment of the Receiver Parameters.....	108
3.3.2 Discussion of Computer Simulation Results.....	109
3.4 Summary.....	112
 CHAPTER 4 - TRACKING PERFORMANCE FOR TIME VARYING SYSTEMS.....	 114
4.1 Tracking Stochastic Variations...	115
4.2 A First Order Adaptive Model.....	119
4.3 Optimum Step Size for Sinusoidal Variations in Optimum Parameters.	124
4.4 Comparison with Experimental Results.....	129
4.4.1 Rectangular Pulses and a Low Pass Filter.....	129
4.5 Adaptive Estimation of the Step Size Parameters.....	132
4.5.1 Higher Orders of Adaption.....	134
4.5.2 Separation of "Steady State" and "Tracking" Errors.....	137
4.5.3 Multi-Mode Adaption.....	138
4.5.4 Other Methods of Adjusting Step Size Parameters.....	139
4.6 Summary.....	140

CONTENTS (Cont'd)

	Page
CHAPTER 5 - APPLICATION OF ADAPTIVE TECHNIQUES TO PAM COMMUNICATIONS WITH INTERSYMBOL INTERFERENCE.....	142
5.1 Intersymbol Interference in PAM Communication Systems.....	143
5.2 Analog Filter Adaption with Intersymbol Interference.....	147
5.3 Simplification of the Gradient Estimates.....	153
5.4 Adaption of the Sampling Time....	156
5.5 Simulation of a PAM System with Intersymbol Interference.....	158
5.5.1 Summary of the Results of the PAM Simulation.....	160
5.5.1.1 Additional Comments on Non- Matched Filter Effects.....	169
5.6 Summary.....	170
CHAPTER 6 - ADDITIONAL APPLICATIONS FOR ADAPTIVE SYSTEMS.....	172
6.1 Receivers for Nonlinear Modulation Systems.....	173
6.1.1 General Phase or Frequency Demodulation.....	174
6.1.2 Phase Locked Loop Demodulators...	176
6.1.3 Feedback Communication Systems and Variable Data Rates.....	179
6.2 Process Control in Manufacturing.	182
6.3 Adaptive Adjustment of Complex Systems.....	184
6.3.1 Diversity Systems.....	187
6.3.2 Adaptive Antenna Arrays.....	187
6.4 Some Future Trends in Adaptive Systems.....	188
CHAPTER 7 - THESIS SUMMARY.....	191
7.1 Topics for Further Study.....	199
REFERENCES	201
APPENDIX A	207
APPENDIX B	209

CONTENTS (Cont'd)

	Page
APPENDIX C	211
APPENDIX D	214
APPENDIX E	216
APPENDIX F - ESTIMATION OF SECOND PARTIAL DERIVATIVES.....	218
APPENDIX G - DETAILS OF THE HYBRID COMPUTER SIMULATION OF A PAM COMMUNICATIONS SYSTEM.....	220
G.1 The Transmitter and Channel.....	222
G.2 Analog Receiver Filter.....	223
G.3 Digital Receiver and Control.....	224

LIST OF ILLUSTRATIONS

<u>Figure</u>		<u>Page</u>
1.1	A Simple Adaptive System.....	5
1.2	A Basic PAM Communication System.....	11
3.1	A Simple PAM Receiver.....	66
3.2	Implementation of a Difference Equation.....	72
3.3	One Sided Difference Equation Implementation.....	74
3.4	A Simple Adaptive Binary PAM Receiver.....	78
3.5	Variable Relationships for Eqns. 3.25 and 3.26.....	88
3.6	MSE vs Convergence Time as a Function of Algorithm Order.....	101
3.7	Simple Binary PAM Receiver Simulation.....	111
4.1	A First Order Adaptive Model.....	120
4.2	Parallel Adaption of Step Size Parameters.....	135
4.3	Sequential Adaption of Step Size Parameters.....	136
5.1	A Binary PAM Communication System with Intersymbol Interference.....	148
5.2	Estimation of the Error Gradients for Analog Filter Parameters.....	154
5.3	Block Diagram of Simulated Adaptive PAM Receiver.....	159

LIST OF ILLUSTRATIONS (Cont'd)

<u>Figure</u>		<u>Page</u>
5.4	Typical Responses to Step Perturbations from Optimum for (a) Bandwidth; and (b) Sample Time; of the Analog Prefilter.....	161
5.5	Analog Prefilter Response to Step Changes in Received Pulse Bandwidth..	163
5.6	Performance of Various Receivers as a Function of SNR.....	164
5.7	The Effect of Sample Time on the Adaptive Receiver Performance.....	167
6.1	Generalized Phase or Frequency Modulation Receiver.....	175
6.2	A Phase Locked Receiver.....	178
6.3	A Simple Model of a Complex Manufacturing Process.....	183
6.4	Simplified Implementations of Complex Adaptive Systems.....	185
F.1	Estimation of Second Partial Derivatives.....	219
G.1	Hybrid Computer Simulation of a PAM System.....	221

LIST OF TABLES

<u>Table</u>		<u>Page</u>
3.1	Tracking Results.....	104
3.2	Adaptive Receiver Improvement as a Function of Cycle Time.....	106
4.1	Comparsion of Calculated and Measured Tracking Errors.....	130
4.2	MSE as a Function of Step Size Parameter.....	133

CHAPTER 1

INTRODUCTION

The design of communication or control systems involves considerable effort to obtain systems having the "best" performance in some sense. The performance is evaluated by means of a suitable criterion such as highest signal to noise ratio or minimum mean squared error, and that system having the best performance measure and still satisfying the design constraints may be termed the optimum system. Optimum is always relative to some performance measure such as the above, and in engineering the design constraints usually include cost, size, reliability, and efficiency.

If the system operates in a completely known, time invariant environment, it is usually a relatively straightforward problem to specify the system so that the performance will always be optimum. In practice, however, all systems are time varying or unknown to some extent. In some problems these effects are small enough to be ignored, but many systems in use today do not fall into this category. The designer may then design for an average environment (if such

an estimate is available) or he may design for a worst case situation (if this is known) and state that the system performance will always be better than, or equal to the performance for the worst case. However, worst case design may constrain the system to have a performance quite inferior to that of a system that can take environmental changes into account.

The best solution to the problems of unknown or time variant environments is to use an adaptive system. This is a system that has the ability to change its parameters in response to a change in its environment in order to keep its performance near optimum. Because it may be difficult or impossible to derive the analytically optimum adaptive system, it is often necessary to start with the optimum fixed system, and then augment it to make it adaptive. This is an example of constrained optimization, a technique which will be used here extensively.

This thesis is an examination of adaptive systems, particularly those of the type just mentioned. In the next section, the origin of time varying effects is considered in greater detail.

1.1 Origin of Unknown and Time Varying Effects

In control systems, slow system changes can come about because of changes in the input parameters. An example of this would be variations in the raw materials used in a refining process. Another example would be changes in noise levels. Often changes take place in the control system itself due to wear of mechanical parts or aging of electronic circuitry.

In communication systems changes can occur in the transmitted signal, and in the noise level, and performance variations can arise due to aging or drift problems in the receiver. But most of the unknown or time varying effects are caused by changes in the channel itself. For example, high frequency radio channels exhibit fading and time varying multipath effects.⁽¹⁾ For data links using coaxial, or twisted pair cables temperature changes can cause large variations in the channel impulse response. Problems of unknown characteristics; even when time invariant, arise when the channel is unknown because it has not been measured. An example of this is a time invariant randomly selected channel such as the dialed telephone network. There is available today a wide variety of data transmission modems that use

the telephone network, and there is considerable interest in operating these systems at high data rates. The telephone link being used is probably time invariant for all practical purposes, especially if the link is only used for a short time. However, channel characteristics can vary considerably between different links and so again a modem designed for worst case operation will often be quite inferior to one that is capable of adjusting itself to make best use of the channel at hand.

It is the objective of this thesis to study some systems having the ability to maintain optimum performance even though their environments may be initially unknown, or change in an unknown way. The work here is concerned with digital communication systems, but the methods of adaption to be presented apply to a large variety of problems.

The next section presents the general problem to be considered, and indicates its areas of application.

1.2 General Adaptive System

A general outline of the adaptive problem is given in this section. Figure 1.1 is a simplified

Figure 1.1 A Simple Adaptive System

representation of an adaptive system. A signal of some type, R , which has been corrupted by additive noise, N , is to be processed by the system shown in Figure 1.1 as $H(\bar{p})$. The characteristics of the received signal R may be time varying or initially unknown by the processing system, or both. The noise, N , is a sample function of some random process whose statistics may be unknown.

The system response is controlled by an adjustable set of K parameters $p_1, p_2, \dots, p_K$ which will be denoted here as the vector $\bar{p}$. The system output, X , is then compared with a reference or desired response, D , and the difference is defined as the system error, E . Clearly X and E are also functions of the parameter set $\bar{p}$. The reference value D may not always be available. However, in many control applications the desired output is known. For example, a manufacturing process may be required to produce articles with certain specified physical characteristics such as size or weight. In a ballistics problem, D is the required trajectory. In communications work D may represent a special training sequence known at the receiver, or as will be discussed later, D may be easily derived from the processed output of certain digital communication systems.

The processor is required to adjust the parameter set $\bar{p}$ to obtain the optimum system performance as measured by some suitable criterion. This criterion could be minimum mean squared error (MSE), maximum signal to noise ratio (SNR), or any other suitable measure of performance. Let V be a performance measure for the system. Since its value will depend on the values of the set $\bar{p}$ we may write $V(\bar{p})$.¹ The problem then is simply to adjust the set $\bar{p}$ to maximize (or minimize) the function V . As indicated in Figure 1.1 the system input, output, desired response and error can be examined and used to adjust the set $\bar{p}$.

The system of Figure 1.1 is completely general. The signal R can be time variant or time invariant. The processing system can be linear or nonlinear, although in general it is time invariant except for the changes caused by adjusting the set $\bar{p}$. The random noise processes involved may be stationary or non-stationary. The model could represent a communication system in which X is a noisy estimate of the information that was transmitted. The reference D is not likely to be the actual transmitted information unless a known test signal is being sent. However, a reference

1

V of course is also a function of the received signal and noise characteristics. However, only $\bar{p}$ will be shown as an explicit argument since these are the only parameters the system can adjust while seeking optimum performance.

value can be derived easily from the received signal for many digital communication systems, and this is described in the next section.

Figure 1.1 could also represent a control system that is required to have an output of the form of D . The system could be an identification setup in which the parameters $\bar{p}$ are assumed to specify some unknown, possibly time varying, system whose output is D in response to the same input $R + N$, and it is desired to learn these parameter values.

These types of problems have received considerable attention in recent years. (2,3,4,5) The problem to be considered in detail in this thesis is much more specific and is outlined in the next section. However, the methods of adaption are applicable, with slight modifications, to most of the problems that can be put into the form of Figure 1.1.

1.3 Adaptive Receivers for Digital Communications

In the past several years considerable literature has been published concerning adaptive communications receivers. (6-12) Some of these are

concerned with adaptive digital receivers such as Transversal Equalizers that are preceded by a fixed analog receive filter. All are essentially concerned with time invariant systems and consider only the transient phenomena during acquisition, and the steady state MSE attainable for time invariant systems. In fact tracking ability is often sacrificed in order to attain a smaller MSE. The study of tracking algorithms for time varying problems forms a large part of this thesis, and considerable simulation work was carried out to evaluate some of the results.

In Section 1.1 it was noted that time varying effects in communication systems arise most often in the channel, which may be time invariant but unknown because it is selected at random as in the dialed telephone network. In this thesis some adaptive receivers will be proposed to compensate for these effects and improve the performance of the system. The type of time variations to be considered here are slowly varying with respect to the digital data rate. Rapid variations, causing significant change within a few sample times, cannot be handled by these techniques. More discussion will be given in

Section 3.2 on what constitutes fast fading for the adaptive systems used.

The actual mechanism for the time variations will not be of concern because no attempts will be made to derive optimum time varying receivers based on detailed analysis of those phenomena. Instead a receiver that would be optimum if the channel were fixed may be used (or sub-optimum receivers used because of their simplicity) and these made adaptive by the techniques to be developed.

Because there is considerable interest in digital pulse amplitude modulation (PAM) communications, this thesis will concentrate on developing adaptive receiving techniques for these systems; however, the techniques can be applied to any system whose receiver can be put in the form of Figure 1.1.

A basic PAM system is shown in Figure 1.2. The message symbols are a discrete set of amplitudes that will be used to determine the pulse height transmitted over the channel. The pulses are sent at the data rate of one each T seconds. It is assumed that the original information is coded into the set $\bar{\theta}$ by a noiseless reversible process. The channel

Figure 1.2 A Basic PAM Communication System

is assumed to include the modulator and demodulator and all the system variables except the noise. These variables are represented by the parameter set $\bar{a}$. The receiver will only be concerned with these parameters through their effect on the received signal $r(t, \bar{a})$. The received noise $n(t)$ is assumed to be a random process that is additive and independent of the channel variations. It is assumed that this noise has a zero mean value and co-variance function $N_0 K(t, t')$. Because of the complications that arise in dealing with nonstationary processes the noise will usually be considered stationary so that $K(t, t') = K(t-t')$, although this is not necessary in general. In addition, if the noise is white then $K(t-t') = \delta(t-t')$ where $\delta(\cdot)$ is the Kronecker delta function.

The processing system $h(t, \bar{p})$ may or may not be an optimum filter. Its structure is fixed but it has a set of parameters $\bar{p}$ that can be adjusted to change the system performance. The filter response is thus a function of time varying parameters, but is not explicitly time variant. The output of this system consists of a signal portion, $x(t)$, and a noise portion $y(t)$ which arises in response to the input noise $n(t)$. Both of these quantities are also functions of $\bar{p}$.

This output is sampled at the data rate and then quantized to produce an estimate, $\hat{\theta}$, of the transmitted message symbol. It is assumed that symbols are received at multiples of the baud time T and any delay between transmitter and receiver has been ignored.

It is to be observed that the reference value is derived from the processed receive signal in this system. This is possible with digital systems that are working reasonably well, because it is known that at the sample times the estimate $x(kT)$ should have some definite characteristics. In this case its amplitude should be one of a discrete set of values. Since the receiver will usually be working near its optimum, the quantized estimate is usually correct, or in other words, the probability of error is small. By subtracting the quantized estimate from the analogue estimate an analogue error $e(kT)$ results which can be used as a measure of system performance. This type of reference must be used because if the desired sequence of message symbols were available at the receiver, there would be no need for a system to transmit them.¹

A function $V(t, \bar{p})$ will be defined as a suitable

1

An exception to this occurs when a known training sequence is sent to allow adjustment of the system before using it in a data communication mode.

performance measure. This could be signal to noise ratio (SNR) or mean squared error (MSE) as specific examples. In a digital system this performance need only be evaluated at the sample times and may be written as $V(kT; \bar{p}^k)$ or just $V(\bar{p}^k)$ which indicates that the performance at the sample time kT is a function of the parameter values at time kT .

The basic objective is to adjust the parameters $\bar{p}$ (and sometimes the sampling time) to maximize $V(kT, \bar{p})$ and thus optimize the receiver. In addition it is desired to keep changing these parameter values in response to changes occurring in the received signal in order that the performance stay maximized at all times. This ability to track time varying received signals is very useful, and it also greatly complicates the analysis as will be demonstrated in some examples later on.

In those problems where V should be minimized, - V will be maximized. Maximizing $V(\bar{p})$ will yield a set of optimum parameters $\bar{p}^*$ which give the best performance at the sampling time that can be obtained within the constraints imposed by the form of $h(t, \bar{p})$. Note that V is not restricted to be a linear function of $\bar{p}$.

1.4 Maximization of the Performance Criterion

In order for $V(\bar{p})$ to be a maximum with respect to $\bar{p}$ we have the following necessary condition:⁽¹³⁾

A continuous function $V(\bar{p})$ of K independent variables $p_1, p_2, \dots, p_K$ attains a maximum only for those values of $\bar{p}$ for which the K partial derivatives vanish simultaneously:

$$\frac{\partial V(\bar{p})}{\partial p_i} = 0, \quad i = 1, 2, \dots, K \quad 1.1$$

or for those values of $\bar{p}$ at which one or more of these derivatives is discontinuous.

In addition we have the following sufficiency conditions for V to attain a maximum:⁽¹³⁾

$$\begin{array}{lll} D_i > 0 & i = 2, 4, 6, \dots, K & \begin{array}{l} K \text{ even} \\ K-1 \text{ } K \text{ odd} \end{array} \\ D_i < 0 & i = 1, 3, 5, \dots, K-1 & \begin{array}{l} K \text{ even} \\ K \text{ } K \text{ odd} \end{array} \end{array} \quad 1.2$$

where D_i is the determinant given by

$$D_i = \begin{vmatrix} \frac{\partial^2 V}{\partial p_1 \partial p_1} & \frac{\partial^2 V}{\partial p_1 \partial p_2} & \dots & \frac{\partial^2 V}{\partial p_1 \partial p_i} \\ \frac{\partial^2 V}{\partial p_2 \partial p_1} & & & \vdots \\ \vdots & & & \vdots \\ \frac{\partial^2 V}{\partial p_i \partial p_1} & \dots & & \frac{\partial^2 V}{\partial p_i \partial p_i} \end{vmatrix} \quad 1.3$$

The location of the stationary points may be found by simultaneous solution of the Equations 1.1 whose roots satisfy Equations 1.2.

1.4.1 Solving for the Optimum Parameters

When the functional for $V(\bar{p})$ is known the optimum solution can be obtained by performing the matrix operations indicated in Eqns. 1.1, 1.2, and 1.3. This method can be quite complex and requires considerable computational capability. Furthermore, if the input is time varying the equations must be resolved fairly often to obtain a new set of parameters optimizing $V(\bar{p})$. In addition, many problems cannot be solved by this method.

An often better way is to solve the equations iteratively by choosing an initial solution estimate $\bar{p}^0$ and adjusting the $\bar{p}$ at each sample time (or multiples of the sampling time) in such a way that the sequence of parameter values

$$\bar{p}^0, \bar{p}^1, \bar{p}^2, \dots, \bar{p}^k, \dots$$

approaches $\bar{p}^*$ as k gets large. This iterative method of solution requires only simple computations at each step and can usually be implemented with very little hardware. In addition it has the built in ability to

track input signal variations by continuously iterating the $\bar{p}$ so that they follow the $\bar{p}^*$ as the input signal characteristics change. The disadvantage of the iterative technique is that it takes several steps to reach the optimum, and may have difficulty if a reasonable initial estimate of the parameter set is not available. In these cases, special start-up techniques may be required. The iterative technique is also limited in its tracking ability to time variations that are slow compared to the data rate. Details of tracking ability will be given in later sections and many actual systems have slow enough variations to be handled by this approach.

There are many techniques for the iterative solution of sets of equations, and for searching for the maximum of a function.⁽¹⁴⁻²²⁾ The simplest of these iterative techniques or stochastic approximation procedures are the gradient methods described in Chapter 2. Before examining these, some justifications for believing that gradient techniques are well suited to the problem at hand are given in the next section.

1.4.2 Suitability of Gradient Techniques

The use of gradient methods for adapting systems has become quite common, especially for PAM communication receivers. (9,10,11,12,23,24,33) The widespread use and apparent success of these methods arises in part for reasons considered in this section.

In general the system will be working reasonably well and it will be possible to pick a set of initial parameter values $\bar{p}^0$ such that the system is performing close to its optimum. For problems where this is not possible, special start-up techniques may be used, but we are more concerned here with the steady state and tracking results than with start-up or acquisition performance. The system will have to adapt itself on line while actual data is being processed and so the initial condition assumption is reasonable. This means that good estimates of the set of gradients, or partial derivatives $\frac{\partial V(\bar{p})}{\partial p_i}$, $i = 1, 2, \dots, K$ can be obtained without resorting to complex estimation procedures. Gradient methods need no apriori knowledge about process statistics, and no estimates of statistics or correlation functions need to be made. The algorithm simply increases or decreases each parameter value slightly according to the magnitude and sign of its present gradient. In addition, the performance

measure "hypersurface"; $V(\bar{p})$ will nearly always be a well behaved function of the parameters. It will have a unique maximum value; and "ridges" and "holes" in the surface will be uncommon. In many cases the variables have little interaction. This behaviour can be assured by selection of an appropriate form for the performance measure such as minimum mean square error. The performance measure will generally be well behaved for tracking problems in which the system is operating close to optimum and is only required to follow changes in the optimum values. For these cases the performance hypersurface should have no local extrema within the immediate vicinity of the optimum point.

The above characteristics indicate that simple gradient techniques may be sufficient in many cases to solve the adaption problem.

The processing system, such as a communication receiver; will often be fairly simple. It would therefore be unreasonable to use a complex system to do the iterative adjustments if the improvement in performance did not justify an adaptive system of greater complexity than the receiver itself.

Thus complex estimation or computation requirements should be minimized. These requirements for

simplicity seem to dictate that simple gradient techniques may also be necessary; and so it is reasonable to examine these as an adaptive scheme.

1.5 Related Work on Adaptive Systems

The use of adaptive systems has become very popular in many areas of communication and control. A brief list of some work related to the contents of this thesis is given in this section.

One area of special interest in communications is adaptive receivers for Pulse Amplitude Modulated (PAM) signals where intersymbol interference problems occur. For binary PAM in particular, the receivers are fairly simple and the gradient adaptive circuitry can also be made very inexpensive using modern digital techniques. Coll and George⁽⁵³⁾ were the first to use gradient methods with a MSE criterion in adapting transversal equalizers. George et al⁽¹¹⁾ showed that this technique could be applied to nonlinear equalizers for PAM as well. Many others^(6,7,9,10,12,30-32) have used gradient techniques in similar problems. However, these papers are all concerned with the acquisition of a time invariant optimum and the minimization of steady state mean square error. The gradient

techniques are examined in this thesis for use in tracking time varying optima, and the selection of an optimum fixed step size to do this. In addition, the analogue portion of the receiver will be made adaptive and the performance of the combined adaptive systems will be investigated.

An application of gradient techniques to adaptive antenna systems is discussed by Widrow et al⁽²⁴⁾. In this work minimum mean square error is used as a performance criterion for gradient algorithm adaption of the patterns of phased array antennas. In this work the signal statistics are assumed known and some results are given for time varying problems. More discussion of this work is found in Section 4.3 of this thesis.

Two general references to adaptive work in control systems are Stromer⁽⁵⁾ and Aseltine et al⁽⁴⁾. In recent years Kalman type filters have been considered for estimation and adaption⁽³⁹⁻⁴²⁾, although these filters are usually quite complex. Very little has been done in this area with time varying systems although some work is given by Jazwinski^(37,38).

In other control systems work McMurtry and Fu⁽²⁷⁾ have considered using a stochastic automaton

to do a random global search on a multimodal performance surface. Linear reinforcement is used as an averaging method to determine the control input most probable to produce optimum results. Work has also been done by McLaren⁽²⁶⁾ using a stochastic automaton to obtain monotonic convergence to the expected system performance by relating the probability distribution of the system response to performance and finding the best response on a probabilistic basis. These papers are concerned with long term convergence optima, not transient response or the tracking of time varying phenomena.

Some work by Riordon^(28,29) on the dual problems of estimation and control considers the transient effects as well. This is done by imposing the additional criterion that the total estimation cost after any number of iterations be minimized; that is, the lowest cost for maximum convergence is obtained. Again the disturbances are not considered to be time variant.

Some work on time varying adaptive systems has been done by Kushner⁽⁴³⁾ on parameter identification for linear systems. Further comparisons with this work are made in Chapter 4.

Other references to work in the literature are found in sections of this thesis dealing with specific problems.

1.6 Summary

This chapter introduced the idea of adaptive systems giving indications of the type of disturbances they can combat, and introducing an adaptive configuration that can be used for many types of problems. The emphasis is on systems that are simple, practical and easily constructed with analog or digital circuitry. The idea of using gradient techniques to iteratively adapt the systems to optimum performance was put forward because these methods tend to be easy to implement. This thesis is not a study of algorithms as such. Basic gradient algorithms will be used but some modifications to these will be considered that simplify the system, and attempts will be made to analyse the errors involved in these approximations. The work is more concerned with the choice of performance criteria; and with the selection of parameters to use to optimize the performance on line. The emphasis is on adaptive receivers for communication systems; but the methods used can be applied to a variety of problems as indicated in Section 1.2. In particular,

some results are presented for the tracking of time varying optima.

1.7 Thesis Outline

Chapter 2 of this thesis describes some common gradient techniques several of which are used in the analytical and computer simulation work of subsequent chapters. Some modifications to the basic algorithms are introduced which result in simpler hardware realizations of the algorithms. A proof of convergence (or stability) is indicated for several of these algorithms. These proofs are not mathematically rigorous; but are sufficient to show the important points in the analysis. References are made to more rigorous work available in the literature.

In Section 2.4 an analysis is done for the adaption of time varying systems by means of a simple fixed step size parameter gradient algorithm. It is demonstrated that this type of algorithm will also "converge" for time varying systems although determination of an optimum fixed step size is quite difficult. The problem of time varying systems is taken up again in Chapter 4.

Chapter 3 introduces basic Pulse Amplitude

Modulation (PAM) systems as a current and typical application for adaptive systems. A discussion of the performance criteria to use and methods of hardware implementation of gradient algorithms using these criteria is given. Section 3.2 chooses a specific example of rectangular received pulses into a low pass filter receiver for detailed analysis. Inter-symbol interference is absent in this system. The adaptive adjustment of filter gain, filter bandwidth and sampling time are analyzed and convergence is shown for practical gradient estimating procedures. Computer simulations are carried out and results of acquisition, steady state, and tracking performance are reported.

Section 3.3 reports briefly on a simulation using second order receive pulses and a low pass filter receiver. Of particular interest here is good tracking performance obtained when the transmitted pulse rate was increased providing considerable intersymbol interference. This section lays the groundwork for Chapter 5 which considers the adaptive performance of PAM systems in the presence of severe intersymbol interference.

Chapter 4 presents original work on adaptive

systems using gradient algorithms for tracking time varying phenomena. This work continues from the initial results of Section 2.3, and produces some new results for time varying systems. Section 4.1 derives the optimum fixed step size parameter to use when the time varying process is white. Section 4.2 considers a simple low pass filter model for the adaptive algorithm. This model is then used in Section 4.3 to derive the optimum fixed step size parameter to use when the adaptive system's adjustable parameters have optimum settings that are sinusoidal functions of time. Section 4.4 compares the analytical results of Section 4.3 with some computer simulation results. Section 4.5 considers some possible methods of adaptively estimating the step size parameters used in the algorithms of the main adaptive system.

Chapter 5 describes PAM systems in detail with emphasis on the use of adaptive receivers to combat intersymbol interference effects. This chapter introduces adaptive algorithms for optimizing the parameters of the analog prefilter in conjunction with the adaptive equalizer part of the receiver. Some approximations are incorporated into the algorithms, resulting in very simple digital hardware realizations

of adaptive loops for adjusting the prefilter bandwidth and sampling time.

Section 5.5 reports on the results of a hybrid computer simulation of a complete PAM system including a receiver with adaptive analog and digital portions. Results and conclusions of this study are reported and some additional comments are made about the use of adaptive equalizers in conjunction with non-matched prefilters.

Chapter 6 gives an outline of several other applications for simple adaptive systems of the type examined in this thesis.

Chapter 7 summarizes the study presented in this thesis and gives a discussion of some of the major conclusions.

CHAPTER 2

GRADIENT ALGORITHMS AND STOCHASTIC APPROXIMATION PROCEDURES

This chapter describes some of the common gradient techniques that will be used in the analytical and computer simulation work of subsequent chapters. Some modifications to the basic algorithms are introduced; and in particular fixed step size parameters are proposed. Since these new algorithms do not always satisfy the old asymptotic convergence criteria, proofs are outlined showing that the algorithms have sufficient stability and convergence capability for the problems considered in this thesis. These proofs are not mathematically rigorous, but are sufficient to show the important points in the analysis. References are made to more rigorous work available in the literature.

The basic idea of using gradient techniques to maximize a function $V(\bar{p})$ takes many forms, but all use the same manner of iteration. An initial estimate of the parameter set $\bar{p}^0$ is used to evaluate the derivatives or gradients of the performance measure at the point $\bar{p}^0$ with respect to the K parameters $\bar{p}$.

Throughout this thesis $\bar{p}^k$ represents the set of K parameter values after k steps, or at time kT.

Then the parameters are changed to obtain a new set of parameters $\bar{p}^1 = \bar{p}^0 + \bar{\Delta}^0$ such that $V(\bar{p}^1) \geq V(\bar{p}^0)$; that is the performance is improved. The increments $\bar{\Delta}^0$ are selected according to the sign (and usually the magnitude) of the individual gradient components as

$$\Delta_j^0 = \alpha_j \left. \frac{\partial V(\bar{p})}{\partial p_j} \right|_{\bar{p}=\bar{p}^0} \quad j=1,2,\dots,K \quad 2.1$$

where α_j is the step size parameter associated with the adjustment of p_j . The above process is repeated each sample time to yield a recursive relationship for the parameters

$$p_j^{k+1} = p_j^k + \alpha_j \frac{\partial V(\bar{p}^k)}{\partial p_j} \quad j=1,2,\dots,K \quad 2.2$$

where the notation $\frac{\partial V(\bar{p}^k)}{\partial p_j}$ has been substituted for $\left. \frac{\partial V(\bar{p})}{\partial p_j} \right|_{\bar{p}=\bar{p}^k}$.

This process continues until $V(\bar{p}^k)$ is sufficiently close $V(\bar{p}^*)$; or in the time varying case it continues indefinitely with V^k tracking V^* within some acceptable error range. The set $\bar{p}^*$ which maximizes

$V(\bar{p})$ is called the optimum parameter set. This maximum will often be denoted simply as V^* .

In physical systems however, it is not possible to get an exact estimate of the gradient components because of noise. When noisy gradient estimates are used, care must be taken to ensure convergence of the algorithm. These iterative techniques are usually called stochastic approximation procedures because of the random nature of the gradient estimates.

In this case the recursive relation is written

$$p_j^{k+1} = p_j^k + \alpha_j \left\{ \frac{\partial V(\bar{p}^k)}{\partial p_j} + z_j^k \right\} \quad 2.3$$

where z_j^k is the noise disturbing the estimate of the j^{th} gradient component at time k .

If this noise has zero mean, then it appears reasonable to expect that the mean value $E\{p_j^{k+1}\}$ will follow closely the sequence of values that p_j^{k+1} would follow if there were no noise. Care must be taken to guarantee stability, or ensure convergence when noise is present. In these problems rate of convergence, or tracking error, is of secondary importance with respect to stability. In the next sections the convergence conditions on the step size

parameters α_j are stated for some common types of gradient algorithms. In Section 2.2 a method of deriving these conditions, and an outline of a proof of convergence is given for a time invariant system. Several variations of the basic procedures will also be given along with convergence and performance analyses.

2.1 Some Common Gradient Algorithms

This section describes some of the algorithms available for iteratively finding the maximum of a function. They are shown as discrete time algorithms although many also have continuous counterparts. For additional information on these and other algorithms; the reader is referred to the literature (14,15,16,54).

2.1.1 The Robbins-Monro Algorithm

The Robbins-Monro (R-M) method was designed to find the root of a regression function⁽¹⁸⁾; here the derivative of the performance criterion is used as a regression function. Noisy observations of the derivatives or gradients are used in an iterative scheme to find the roots of $\frac{\partial V(\bar{p})}{\partial p_i} = 0$, $i = 1, 2, \dots, K$. These roots locate the maximum value of the performance

criterion (provided they also satisfy Eqns. 1.2).

The recursive relationship is

$$p_i^{k+1} = p_i^k + \alpha_i^k \frac{\partial V(\bar{p}^k)}{\partial p_i} + \alpha_i z_i^k, \quad i=1,2,\dots,K \quad 2.4$$

Here p_i^k is the value of the i^{th} parameter at time k ;

V is a performance criterion that is a function of the K variable parameter set $\bar{p}$;

z_i^k is the error, caused by noise, in estimating the gradient component at time k , and

α_i^k is the step size parameter for the i^{th} adjustment. (Its value can also be a function of time as indicated by the superscript time index. If it is time independent we write α_i ; if it is the same for all K parameters we write α^k , or just α .)

In the work of Robbins and Monro⁽¹⁸⁾ the step size parameters α^k are a monotonic decreasing sequence of constants which satisfies the following conditions:

$$\lim_{k \rightarrow \infty} \alpha^k = 0$$

$$0 < \sum_{k=1}^{\infty} (\alpha^k)^2 < \infty$$

$$\sum_{k=1}^{\infty} \alpha^k = \infty$$

These are sufficient to guarantee convergence.

Time variable step sizes such as $\alpha^k = \frac{a}{(1+k)}$ are common. (7,14,17,18) However, as will be shown later, this type of step size is not acceptable for tracking problems. Convergence in mean square and in probability can be shown for a decreasing step size. (7,18)

In this thesis an algorithm will be referred to as a R-M type algorithm if the iterations are made as in equation 2.4. However, the step size parameter α will usually be a constant.

2.1.2 The Kiefer-Wolfowitz Algorithm

The Kiefer-Wolfowitz (K-W) technique⁽¹⁹⁾ was originally proposed for finding the minimum of a regression function such as the performance function being considered here. Instead of measuring the gradients directly they can be estimated by a difference equation

$$\frac{\partial V(\bar{p})}{\partial p_i} \approx \frac{V(\bar{p} ; p_i + \delta) - V(\bar{p} ; p_i - \delta)}{2\delta} \quad 2.5$$

where δ is a small positive quantity.

The adaptive algorithm then becomes

$$p_i^{k+1} = p_i^k + \alpha_i^k \frac{V(p_i^k + \delta) - V(p_i^k - \delta)}{2\delta} + \alpha_i^k z_i^k \quad 2.6$$

In Eqn. 2.6 all parameters except p_i are assumed constant for both values of the i^{th} parameter and hence are not shown explicit arguments. Again z_i^k is the noise component of the estimated gradient.

In the work of Kiefer and Wolfowitz⁽¹⁹⁾ the step size parameter α^k and the quantity δ^k are both infinite sequences which satisfy the following conditions:

$$\lim_{k \rightarrow \infty} \alpha^k = 0$$

$$\lim_{k \rightarrow \infty} \delta^k = 0$$

$$\sum_{k=1}^{\infty} \alpha^k = \infty$$

$$\sum_{k=1}^{\infty} \alpha^k \delta^k < \infty$$

$$\sum_{k=1}^{\infty} (\alpha^k)^2 (\delta^k)^{-2} < \infty$$

These conditions are sufficient to guarantee convergence.

In this thesis an algorithm will be called a K-W type algorithm if the iterations are made as in Eqn. 2.6; but α and δ are kept constant, and therefore the above convergence criteria are not met. However, for the tracking problems considered in this work the above conditions are not necessary for the type of

convergence required. It will be shown (see Section 2.2) that the use of finite fixed step size parameters is sufficient for a tracking type convergence.

2.1.3 A Modified Robbins-Monro Algorithm

Another method that is very similar to the Kiefer-Wolfowitz method may be derived by approximating the gradient in a Robbins-Monro process from the Taylor series

$$V(\bar{p} ; p_i + \delta) = V(\bar{p}) + \frac{\partial V(\bar{p})}{\partial p_i} \delta + \frac{\partial^2 V(\bar{p})}{\partial p_i^2} \frac{\delta^2}{2} + \dots \quad 2.7$$

For small δ , terms with the powers of δ greater than first order may be ignored with small error, provided some assumption is made about the magnitude of the higher order derivatives. A sufficient condition is that all derivatives of order three or greater be less than or equal to the second derivative in magnitude. This assures that the higher order terms in the Taylor series are small with respect to the first derivative term.

Rearranging Eqn. 2.7 gives

$$\frac{\partial V(\bar{p})}{\partial p_i} \approx \frac{V(\bar{p} ; p_i + \delta) - V(\bar{p})}{\delta} \quad 2.8$$

The recursive relationship can now be written as

$$p_i^{k+1} = p_i^k + \alpha_i^k \frac{V(\bar{p}^k; p_i^k + \delta) - V(\bar{p}^k)}{\delta} + \alpha_i^k z_i^k \quad 2.9$$

This looks much like the K-W procedure, except that it is a one sided difference approximation to the derivative and thus puts a small bias on the mean parameter value. (See Section 3.2.3 for an outline of a convergence proof.) However, as will be shown later, it requires only half of the additional hardware to implement it that the K-W procedure requires; and the small bias term is usually negligible with respect to the noise and the parameters' time variations.

2.1.4 Higher Order Algorithms

Higher order in hill climbing algorithms usually refers to the use of higher order derivatives. However these are hard to estimate from physical measurements. In this work, "higher order" will refer to iterative methods making use of gradient estimates obtained in previous stages of iteration. All of the previous algorithms may be termed "first order" in that they make iterative adjustments based only on one data sample, and they iterate at the

data rate. Some basic types of higher order algorithms are illustrated in this section.

Higher order gradient methods smooth the effects of noisy measurements and may permit a more rapid convergence to occur. They have some of the ridge seeking properties of conjugate gradient methods⁽³⁰⁾ (and therefore fast convergence) but are simpler to implement and analyse.

Type I

Consider an M^{th} order iteration of the form:

$$p_i^{k+1} = p_i^k + \sum_{m=1}^{M-1} \alpha_{i,m}^k \frac{\widehat{\partial V(\bar{p}^{k-m})}}{\partial p_i} \quad 2.10$$

This iteration is a weighted sum of the previous M gradient estimates where $\frac{\widehat{\partial V(\bar{p}^m)}}{\partial p_i}$ is any noisy estimate of the i^{th} gradient component at time m , and $\alpha_{i,m}^k$ is the step size weighting for the i^{th} parameter, the m^{th} previous gradient and the iteration being calculated at time k . Usually $|\alpha_{i,m}^k| > |\alpha_{i,n}^k|$ for $m < n$; that is more recent gradient estimates are weighted more heavily. Proof of convergence is given for an algorithm of this type in (30).

Type II

Another type of higher order algorithm computes an iteration based on the size of the parameter changes as well as the present gradient. A second order algorithm of the type

$$p_i^{k+1} = p_i^k + \alpha_i \frac{\widehat{\partial V(\bar{p}^k)}}{\partial p_i} + \sum_{m=0}^{M-1} \beta_{i,m}^k (p_i^{k-m} - p_i^{k-m-1}) \quad 2.11$$

is considered in (31). In this thesis fixed step size parameters, $\alpha_i^k = \alpha_i$ will be used and any higher order methods will arise from the averaging of several noisy estimates of the gradient before iteration of the parameters. An algorithm of the form

$$p_i^{k+L} = p_i^k + \frac{\alpha_i}{M} \sum_{m=1}^M \frac{\widehat{\partial V(\bar{p}^{k-m+1})}}{\partial p_i} \quad 2.12$$

will be used in some of the work in this thesis. This is a type of M^{th} order algorithm because it uses a weighted sum of the M most recent derivative estimates in forming the iteration. It may also change the parameters every L sample intervals instead of each bit time. This algorithm is easy to implement since only one step size and one accumulating sum need be stored.

2.1.5 Quantized Gradient Algorithm

Another type of algorithm that is particularly useful in simple hardware implementations is the quantized gradient algorithm which uses only the sign of the estimated gradient to iterate the parameters. The iterative adjustments are of the form

$$p_i^{k+1} = p_i^k + \alpha_i \operatorname{sgn} \left[\frac{\partial V(\bar{p}^k)}{\partial p_i} + z_i^k \right] \quad 2.13$$

and a simple convergence proof for this algorithm will be given in Section 2.3.

An ingenious implementation of an algorithm like this was proposed by Lucky⁽⁹⁾. Here the sign of the gradient estimate drives an up/down counter. When the counter overflows (underflows) the parameter value p_i is reduced (increased) by the amount α_i and the counter is reset to its mid-value. This averages the gradient estimates and effectively provides a variable step size since the parameter is updated more often when the gradient is large (that is, consistently of one polarity) than when it is small.

2.1.6 Tracking Algorithms and Time Varying Systems

In the literature,^(7,14,17,18) the step size parameters are often decreasing functions of time. This is suitable for acquisition type problems or time invariant systems. Under these restrictions and some conditions on the behaviour of $V(\bar{p})$ convergence in mean square and with probability one can be shown.^(7,16) One exception is Daniell⁽³⁴⁾ who alludes to the problem of time varying applications but really only considers the time invariant case and conjectures that the system performance is insensitive to many types of mild nonstationarity. This is just the quasi-static case for very slow time variations and gives no indication of the effects of time variations that are rapid enough to invalidate the quasi-static assumption.

R. de Figueiredo⁽³⁵⁾ considers the problem of estimating conditional probabilities in pattern recognition problems in nonlinear nonstationary environments. However, the non-linear form must be known for the algorithm to converge and this information is rarely available. V. Dupac⁽³⁶⁾ considers the convergence of R-M and K-W algorithms when the optimum root (or minimum) is a linear function of time. He is able to show convergence of the mean error to zero but able only to

bound the mean square parameter error within an order of magnitude. The step size parameters are decreasing functions of time; and an expression cannot be obtained that would permit selection of a fixed step size parameter to give a minimum total MSE for time variations.

There has been a great deal of work in the control system literature on Kalman Type filters for estimation and adaption.⁽³⁹⁻⁴²⁾ Very little work is available here for systems with time varying optima, although some work in this area is given by Jazwinski.^(37,38) However these filters are very complex to implement and will not be considered in this thesis. The problem with time varying systems is that even simple approximate expressions for a suitable fixed step size parameter are dependent on the characteristics of the time variations. These are nearly always unknown, and any attempt to estimate them leads to implementational complexities.

The only approach other than assuming quasi-stationarity is to use some simplified performance criteria such as response to step changes in the system parameters. It may be possible to get some idea of step size parameters from such criteria, or

possibly from some semi-empirical formulae based on estimates of the instantaneous error. Such formulae may be examined in conjunction with simulation work.

The problems of time varying systems are examined in Section 2.4. Some analysis similar to this was carried out by Kushner⁽⁴³⁾ who also considered neglected parameters and measurement errors.

In this thesis a fixed step size parameter will usually be assumed and convergence can be shown quite easily. However, because of the finite step size, the parameters do not converge in mean square; they always have a finite variance. It is this finite step size that allows the time varying parameters to be tracked. Proof of convergence for time varying systems is very difficult. A proof is given in the next section for time invariant systems with fixed step size parameters and the algorithm performance is discussed for application to time varying cases. The problem for time varying systems is taken up again in Section 2.4 and in Chapter 4 where some approximate results are obtained.

2.2 Convergence for Time Invariant Systems

In this section convergence will be shown for

a time invariant system when certain simplifying assumptions can be made. Some references to other work will be given where the restrictions have been relaxed.

Consider the first order Robbins-Monro type algorithm

$$p_i^{k+1} = p_i^k + \alpha_i \frac{\partial V(\bar{p}^k)}{\partial p_i} + \alpha_i z_i^k \quad i=1,2,\dots,K \quad 2.14$$

Here the iteration has been based on noisy estimates of the gradient

$$\widehat{\frac{\partial V(\bar{p}^k)}{\partial p_i}} = \frac{\partial V(\bar{p}^k)}{\partial p_i} + z_i^k \quad 2.15$$

where $\frac{\partial V(\bar{p}^k)}{\partial p_i}$ is the true gradient component and z_i^k is a zero mean additive noise term of variance

$$E\{z_i^m z_i^n\} = \sigma_{z_i}^2 \delta_{mn} \text{ indicating that the noise is assumed}$$

independent from sample time to sample time. This independence condition can be relaxed somewhat.⁽¹⁶⁾

It is further assumed that the noise characteristics are not functions of $\bar{p}$. This is often easy to arrange in practice. (See Section 3.2.)

Because of the noise effects the $\bar{p}$ are a set of random variables. It is assumed for convenience that

these parameters are independent, that the system is time invariant, and that the noise is stationary.

The optimum parameter set is $\bar{p}^*$ which satisfies Eqns. 1.1 and 1.2. The i^{th} parameter error at time k may be defined as

$$e_{p_i}^k \triangleq p_i^k - p_i^* \quad 2.16$$

This is also a random variable.

Using this definition in Eqn. 2.14 we can write a recursive relationship for the parameter errors.

$$\begin{aligned} e_{p_i}^{k+1} &= p_i^{k+1} - p_i^* \\ &= p_i^k - p_i^* + \alpha_i \frac{\partial V(\bar{p}^k)}{\partial p_i} + \alpha_i z_i^k \\ &= e_{p_i}^k + \alpha_i \frac{\partial V(\bar{p}^k)}{\partial p_i} + \alpha_i z_i^k \end{aligned} \quad 2.17$$

For a system that is working reasonably well p_i^k will be close to p_i^* ; and it may be assumed that $V(\bar{p}^k)$ is analytic in p_i^k so that we may write the Taylor series

$$\begin{aligned} V(\bar{p}^k ; p_i^k + e_{p_i}^k) &\approx V(\bar{p}^k ; p_i^*) + \frac{\partial V(\bar{p}^k)}{\partial p_i} \bigg|_{p_i^k = p_i^*} (p_i^k - p_i^* + e_{p_i}^k) \\ &\quad + \frac{1}{2} \frac{\partial^2 V(\bar{p}^k)}{\partial p_i^2} \bigg|_{p_i^k = p_i^*} (p_i^k + e_{p_i}^k - p_i^*)^2 + \text{higher order terms.} \end{aligned} \quad 2.18$$

In order for the higher order terms to be considered small with respect to the second derivative term it is sufficient to assume that the magnitude of the third and higher derivatives is less than or equal to the magnitude of the second derivative.

The expansion of Eqn. 2.18 could be made exact by using the mean value theorem, but the mean value is not known here. The expansion is exact for performance criteria of second order since all higher order derivatives are zero. The performance criterion is usually of low order, or may be closely approximated as a quadratic function around the optimum. Eqn. 2.18 is thus a reasonable approximation to make.

Now $\left. \frac{\partial V(\bar{p}^k)}{\partial p_i} \right|_{p_i^k = p_i^*} = 0$ by Eqn. 1.1 since p_i^* is a coordinate of the maximum of $V(\bar{p})$.

Rearranging and noting from Eqn. 2.16 that $p_i^* = p_i^k - e_{p_i}^k$

we have

$$\frac{V(\bar{p}^k; p_i^k + e_{p_i}^k) - V(\bar{p}^k; p_i^k - e_{p_i}^k)}{2 e_{p_i}^k} \approx \frac{1}{2} \left. \frac{\partial^2 V(\bar{p}^k)}{\partial p_i^2} \right|_{p_i^k = p_i^*} \frac{(2 e_{p_i}^k)^2}{2 e_{p_i}^k}$$

$$= \left. \frac{\partial^2 V(\bar{p}^k)}{\partial p_i^2} \right|_{p_i^k = p_i^*} (e_{p_i}^k) \quad 2.19$$

For operation near the optimum $e_{p_i}^k$ will be small.
Therefore

$$\begin{aligned} \frac{\partial V(\bar{p}^k)}{\partial p_i} &\approx \frac{V(\bar{p}^k ; p_i^k + e_{p_i}^k) - V(\bar{p}^k ; p_i^k - e_{p_i}^k)}{2 e_{p_i}^k} \\ &\approx \left. \frac{\partial^2 V(\bar{p}^k)}{\partial p_i^2} \right|_{\substack{(e_{p_i}^k) \\ p_i^k = p_i^*}} \end{aligned} \quad 2.20$$

We now have an expression for the first derivative in terms of the parameter error $e_{p_i}^k$. Substituting this back into Eqn. 2.17 gives

$$e_{p_i}^{k+1} = e_{p_i}^k \left(1 + \alpha_i \left. \frac{\partial^2 V(\bar{p}^k)}{\partial p_i^2} \right|_{p_i^k = p_i^*} \right) + \alpha_i z_i^k \quad 2.21$$

This error can not be made identically zero because of the random noise. However, the adaptive process is stable if the mean error tends to zero while the mean squared error remains bounded. It will also be shown that by choosing the proper sequence of step sizes the mean squared error can be made to vanish.

Taking expected values of both sides of Eqn. 2.21 gives

$$E\{e_{p_i}^{k+1}\} = E\left\{e_{p_i}^k \left(1 + \alpha_i \left. \frac{\partial^2 V(\bar{p}^k)}{\partial p_i^2} \right|_{p_i^k = p_i^*} \right) + \alpha_i z_i^k \right\} + \alpha_i E\{z_i^k\} \quad 2.22$$

Recall that z_i is a zero mean process. Also

$\frac{\partial^2 V(\bar{p}^k)}{\partial p_i^2} \bigg|_{p_i^k = p_i^*}$ is not a function of the random

variable p_i^k . Since the parameters are independent

$e_{p_i}^k$ is independent of $\frac{\partial^2 V(\bar{p}^k)}{\partial p_i} \bigg|_{p_i^k = p_i^*}$. Thus

$$E\{e_{p_i}^{k+1}\} = E\{e_{p_i}^k\} \left(1 + \alpha_i \frac{\overline{\partial^2 V(\bar{p}^k)}}{\partial p_i^2} \bigg|_{p_i^k = p_i^*} \right) \quad 2.23$$

where the overbar is used to indicate an expectation over the ensemble of $\bar{p}$ and not a time average.

Solving Eqn. 2.23 gives

$$E\{e_{p_i}^m\} = \frac{m-1}{\pi} \left(1 + \alpha_i \frac{\overline{\partial^2 V(\bar{p}^k)}}{\partial p_i^2} \bigg|_{p_i^k = p_i^*} \right) E\{e_{p_i}^0\} \quad 2.24$$

Now the initial parameter error $e_{p_i}^0$ will always be finite so the $E\{e_{p_i}^0\}$ is finite. The mean parameter error will thus approach zero geometrically with time if, at each step,

$$\left| 1 + \alpha_i \frac{\overline{\partial^2 V(\bar{p}^k)}}{\partial p_i^2} \bigg|_{p_i^k = p_i^*} \right| < 1 \quad 2.25$$

This sufficiency condition defines limits on the step size parameter α_i which will guarantee convergence.

The convergence will be very rapid if

$$\alpha_i \approx \frac{-1}{\frac{\partial^2 V(\bar{p}^k)}{\partial p_i^2} \bigg|_{p_i^k = p_i^*}} \quad 2.26$$

Thus for rapid convergence each parameter iteration algorithm has an optimum step size parameter. Eqn. 2.26 implies that the optimum step size parameter changes each iteration because it is a function of the random variables p_j^k for $j \neq i$ and thus is time varying during the initial transient. There are algorithms which compute the optimum step size parameter to use at each step such as the Fletcher-Powell method.^(22,14) However, algorithms of this type require either complex gradient estimates or considerable computation, and this amount of complexity is too great for the problems considered here. A fixed step size parameter will therefore be considered for the time being. Its value must be such that it always satisfies the stability or convergence bound of Eqn. 2.25 for all values of $\bar{p}$. These fixed step size parameters still guarantee that p_i^k will converge in mean value to p_i^* , $i = 1, 2, \dots, K$, but the algorithms are much easier to implement. They will of course converge more slowly to a steady state value.

Consider now the variance or mean square parameter error $E\{(e_{p_i}^k)^2\}$. From Eqn. 2.21

$$\begin{aligned}
 E\{(e_{p_i}^{k+1})^2\} &= E\{(e_{p_i}^k)^2\} \left(1 + \alpha_i \frac{\partial^2 V(\bar{p}^k)}{\partial p_i^2} \bigg|_{p_i^k = p_i^*} \right)^2 \\
 &+ \alpha_i^2 E\{(z_i^k)^2\} + 2 \alpha_i E\{z_i^k e_{p_i}^k\} \left(1 + \alpha_i \frac{\partial^2 V(\bar{p}^k)}{\partial p_i^2} \bigg|_{p_i^k = p_i^*} \right). \quad 2.27
 \end{aligned}$$

If the parameter error $e_{p_i}^k$ is independent of the noise then the last term is zero because the noise is zero mean. From Eqn. 2.21 this can be seen to hold if $E\{z_i^j z_i^k\} = \sigma_{z_i}^2 \delta_{jk}$, that is the noise samples are white as assumed. Eqn. 2.27 can be more concisely written as

$$E\{(e_{p_i}^{k+1})^2\} = E\{(e_{p_i}^k)^2\} \overline{H_{k,i}^2} + \alpha_i^2 \sigma_{z_i}^2 \quad 2.28$$

$$\text{where } H_{k,i} = 1 + \alpha_i \frac{\partial^2 V(\bar{p}^k)}{\partial p_i^2} \bigg|_{p_i^k = p_i^*} \quad 2.29$$

The subscripts indicate that for each parameter p_i , $i=1,2,\dots,K$, the value $H_{k,i}$ is a function of time k . The solution to Eqn. 2.28 is

$$E\{(e_{p_i}^m)^2\} = \pi_{j=0}^{m-1} \overline{H_{j,i}^2} E\{(e_{p_i}^0)^2\} + \alpha_i^2 \sigma_{z_i}^2 \left(1 + \sum_{j=0}^{m-2} \pi_{k=j+1}^{m-1} \overline{H_{k,i}^2} \right) \quad 2.30$$

It is desired to show that this mean square parameter error converges to a steady state value. In order to calculate the steady state value, it will be assumed that the $H_{j,i}$ are independent for different values of j . This restriction is not necessary but it simplifies the resulting expressions. In practice this condition is satisfied if the parameters are updated every M sample times instead of each sample time. M is selected large enough for the independence condition to be satisfied. Since α_i has been selected so that Eqn. 2.25 holds for all iterations we have

$$\left| \overline{H_{k,i}} \right| < 1 \quad 2.31$$

and

$$\overline{H_{k,i}^2} = \overline{H_i^2} < 1 \quad 2.32$$

Note that in the steady state $\overline{H_i^2}$ is not a function of time. The RHS of Eqn. 2.30 may be simplified to

$$\begin{aligned} & (\overline{H_i^2})^m E \left\{ (e_{p_i}^o)^2 \right\} + \alpha_i^2 \sigma_{z_i}^2 \sum_{j=0}^{m-1} (\overline{H_i^2})^j \\ &= (\overline{H_i^2})^m E \left\{ (e_{p_i}^o)^2 \right\} + \alpha_i^2 \sigma_{z_i}^2 \left(\frac{1 - (\overline{H_i^2})^m}{1 - \overline{H_i^2}} \right) \end{aligned} \quad 2.33$$

Defining $E \left\{ (e_{p_i})^2 \right\} = \sigma_{p_i}^2$ we have for large m (steady state)

$$\sigma_{p_i}^2 = \alpha_i^2 \sigma_{z_i}^2 \left(\frac{1}{1 - \overline{H_i^2}} \right) \quad 2.34$$

Thus the mean square parameter error converges to the asymptotic value given by Eqn. 2.34 where

$$\overline{H}_i^2 = \left(1 + \alpha_i \left. \frac{\partial^2 V(\bar{p})}{\partial p_i^2} \right|_{p_i = p_i^*} \right)^2 \quad 2.35$$

$$\text{or} \quad \overline{H}_i^2 = 1 + 2\alpha_i \left. \frac{\partial^2 V(\bar{p})}{\partial p_i^2} \right|_{p_i = p_i^*} + \alpha_i^2 \left(\left. \frac{\partial^2 V(\bar{p})}{\partial p_i^2} \right|_{p_i = p_i^*} \right)^2 \quad 2.36$$

Each parameter has a steady state MSE expression of the form of Eqn. 2.34 for $i=1,2,\dots,K$. Eqn. 2.36 can be evaluated when the performance criterion is known explicitly. However, the most important result here is the demonstration that the algorithm is still stable and converges with fixed step size parameters.

If α_i were a decreasing function of time then Eqn. 2.34 shows that the mean square parameter error could also be made to approach zero after many iterations. A treatment of algorithms of this type may be found in the literature. (7,16,18,19) Here the fixed finite step size parameter gives rise to a finite steady state error. This is the price paid for the ability to track changes in the parameters of time varying systems. To keep the steady state error small, we want a small step size, but Eqn. 2.26 shows that there

is an optimum step size for rapid convergence. There is thus a trade off in performance as a function of step size during the transient period even for systems that do not have time varying parameters.

For systems with time varying parameters, the basic iteration equation (Eqn. 2.14) shows that the step sizes cannot be allowed to go to zero or no further adaption can take place. The problem of optimum step sizes for time varying systems will be considered in Section 2.4. This is a difficult problem and is rarely mentioned in the literature. Daniell⁽³⁴⁾ alludes to this problem but really considers only the time invariant case. Some theoretical work with time varying systems has been considered by Kushner.⁽⁴³⁾ We expect, however, that if the parameters do not change too rapidly the steady state error will be approximately that given above. The dynamic or tracking error is more difficult to estimate, but some approximations are attempted in Section 2.4.

2.3 Convergence for the Quantized Gradient Algorithm

Consider the following iteration procedure using only the sign of the noisy gradient and not its magnitude.

$$p_i^{k+1} = p_i^k + \alpha_i \operatorname{sgn} \left(\frac{\partial v(\bar{p}^k)}{\partial p_i} + z_i^k \right) \quad 2.37$$

This kind of algorithm is very important because it can be realized using simple digital hardware and it will be considered in some of the examples in later sections. To show convergence, consider the error

$$e_{p_i}^{k+1} = p_i^{k+1} - p_i^* = e_{p_i}^k + \alpha_i \left[\operatorname{sgn} \frac{\partial^2 V(\bar{p}^k)}{\partial p_i^2} \bigg|_{p_i^k = p_i^*} e_{p_i}^k + z_i^k \right] \quad 2.38$$

Taking expected values gives

$$E\{e_{p_i}^{k+1}\} = E\{e_{p_i}^k\} + \alpha_i \Pr(z_i^k > - \frac{\partial^2 V}{\partial p_i^2} \bigg|_{p_i^k = p_i^*} e_{p_i}^k) - \alpha_i \Pr(z_i^k < - \frac{\partial^2 V}{\partial p_i^2} \bigg|_{p_i^k = p_i^*} e_{p_i}^k) \quad 2.39$$

Suppose for example z_i is a sample function of a zero mean Gaussian noise process of unit variance. Then

$$E\{e_{p_i}^{k+1}\} = E\{e_{p_i}^k\} + \alpha_i \left[1 - 2Q \left(\frac{\partial^2 V(\bar{p}^k)}{\partial p_i^2} \bigg|_{p_i^k = p_i^*} e_{p_i}^k \right) \right] \quad 2.40$$

for positive Q function arguments.¹

Since $Q(x) \leq \frac{1}{2}$ for all x , the term in brackets is always greater than, or equal to zero. Therefore

¹

This is the standard Q function

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy$$

$$E\left\{e_{p_i}^{k+1}\right\} \leq E\left\{e_{p_i}^k\right\}$$

at each step if α_i is a negative constant. Hence the mean error converges to zero. Gersho⁽⁴⁴⁾ shows that the mean square error will converge to a small value for this algorithm. If α_i were allowed to approach zero as k approaches ∞ , then the MSE would approach zero too.

Algorithms of the form

$$p_i^{k+1} = p_i^k + \alpha_i \operatorname{sgn} \left[\sum_{m=k-L}^k \left\{ \frac{\partial V(\bar{p}^m)}{\partial p_i} + z_i^m \right\} \right] \quad 2.41$$

which are of L^{th} order also satisfy convergence criteria and the averaging may aid convergence when the gradient estimates are very noisy.

The convergence times of quantized algorithms are usually longer. However they are much simpler to implement and can avoid problems that may arise using the R-M or K-W techniques if the performance function has inflection points.⁽¹⁴⁾

2.4 Time Varying Systems

This section presents some original work on the use of gradient algorithms for time varying

adaptive systems. New results are derived showing that with fixed step size parameters the algorithms are still stable, and that they are capable of tracking time varying parameters. These equations are used in Chapter 4 where expressions for the optimum fixed step size parameters for tracking algorithms are derived.

Suppose that the received signal R in Figure 1.1 is governed by a set of time varying parameters $\bar{a}(t)$, so that it may be written $R(\bar{a}(t))$. These parameters simply represent the net effect of the time varying distortion of R . Their exact form is not important since the system is only concerned with their effects on the performance criterion V . The optimum parameter set now becomes a function of time $\bar{p}^*(t)$. For discrete time problems the optimum parameter set will be defined as $\bar{g}^k$. The adaptive system must now track changes in the optimum as well as initially "acquire" a set of parameter values close to the optimum when it is started with a suboptimum estimate $\bar{p}^0$.

The same iterative procedure with fixed step size parameters α_i is used. The i^{th} parameter is

then updated according to the relation

$$p_i^{k+1} = p_i^k + \alpha_i \frac{\partial V(\bar{p}^k)}{\partial p_i} + \alpha_i z_i^k \quad 2.42$$

From Eqn. 2.20 we have

$$\frac{\partial V(\bar{p}^k)}{\partial p_i} \approx e_{p_i}^k \left. \frac{\partial^2 V(\bar{p}^k)}{\partial p_i^2} \right|_{p_i^k = g_i^k} \quad 2.43$$

The i^{th} parameter error at time $k+1$ is

$$e_{p_i}^{k+1} = p_i^{k+1} - g_i^{k+1} \quad 2.44$$

Substituting Eqns. 2.42 and 2.43 into 2.44 and adding and subtracting g_i^k on the R.H.S. yields

$$\begin{aligned} e_{p_i}^{k+1} &= p_i^k - g_i^k + \alpha_i e_{p_i}^k \left. \frac{\partial^2 V(\bar{p}^k)}{\partial p_i^2} \right|_{p_i^k = g_i^k} + \alpha_i z_i^k - (g_i^{k+1} - g_i^k) \\ &= e_{p_i}^k H_{k,i} + \alpha_i z_i^k - d_i^{k+1} \end{aligned} \quad 2.45$$

where

$$H_{k,i} = 1 + \alpha_i \left. \frac{\partial^2 V(\bar{p}^k)}{\partial p_i^2} \right|_{p_i^k = g_i^k} \quad 2.46$$

and $d_i^{k+1} = g_i^{k+1} - g_i^k$ is the change in the i^{th} optimum parameter value from time k to time $(k+1)$.

Eqn. 2.45 is very similar to Eqn. 2.21 except for the last term which represents a time varying parameter error due to changes in the optimum parameter set. Solving Eqn. 2.45 gives

$$e_{p_i}^m = e_{p_i}^0 \prod_{j=0}^{m-1} H_{j,i} + \alpha_i \sum_{n=0}^{m-1} z_i^n \prod_{j=n+1}^{m-1} H_{j,i} - \sum_{n=1}^m d_{i,j=n}^n \prod_{j=n}^{m-1} H_{j,i} \quad 2.47$$

where $\prod_{j=m}^{m-1} H_{j,i} \triangleq 1$

The mean error is obtained by taking expectations of both sides of Eqn. 2.47. The first term converges exponentially to zero for large m provided α_i is selected so that $|H_{j,i}| < 1$ for all values of j . This is a sufficient condition for convergence. Since the noise process is assumed to be zero mean, the second term has an expected value of zero. The optimum parameters $\bar{g}^k$ will be assumed to be stationary processes. Thus

$$E \{ d_i^m \} = E \{ g_i^m - g_i^{m-1} \} = \bar{g}_i - \bar{g}_i = 0 \quad 2.48$$

If the optimum parameters have deterministic time

variations, the average increment d_i^m must be zero or the parameters will become unbounded as m gets large. In this case as well the mean parameter errors thus converge to zero even for time varying parameters.

$$\therefore E\{e_{p_i}^{k+1}\} = 0 ; i = 1, 2, \dots, K. \quad 2.49$$

However, the mean square parameter errors are more important. From Eqn. 2.47 we can obtain

$$\begin{aligned} (e_{p_i}^m)^2 &= (e_{p_i}^o)^2 \left(\sum_{j=0}^{m-1} \pi_{j,i} H_{j,i} \right)^2 + \alpha_i^2 \left(\sum_{n=0}^{m-1} z_i^n \sum_{j=n+1}^{m-1} \pi_{j,i} H_{j,i} \right)^2 \\ &+ \left(\sum_{n=1}^m d_i^n \sum_{j=n}^{m-1} \pi_{j,i} H_{j,i} \right)^2 + 2(e_{p_i}^o \sum_{j=0}^{m-1} \pi_{j,i} H_{j,i}) \left(\alpha_i \sum_{n=0}^{m-1} z_i^n \sum_{j=n+1}^{m-1} \pi_{j,i} H_{j,i} \right) \\ &- 2(e_{p_i}^o \sum_{j=0}^{m-1} \pi_{j,i} H_{j,i}) \left(\sum_{n=1}^m d_i^n \sum_{j=n}^{m-1} \pi_{j,i} H_{j,i} \right) \\ &- 2\left(\alpha_i \sum_{n=0}^{m-1} z_i^n \sum_{j=n+1}^{m-1} \pi_{j,i} H_{j,i} \right) \left(\sum_{n=1}^m d_i^n \sum_{j=n}^{m-1} \pi_{j,i} H_{j,i} \right) \end{aligned} \quad 2.50$$

The MSE can be obtained by taking the expected value of Eqn. 2.50. The first term becomes negligible for large m since $|H_{j,i}| < 1$ is assumed to hold. The fourth and sixth terms are zero since the noise, z_i^n is zero mean and independent of the parameter sets $\bar{p}(t)$ and $\bar{a}(t)$. The fifth term has a zero expected value since

the d_i^n are independent of the $H_{j,i}$ and from Eqn. 2.48
 $E \{ d_i^n \} = 0$. Therefore

$$E \{ (e_{p_i}^m)^2 \} = \alpha_i^2 E \left\{ \left(\sum_{n=0}^{m-1} z_i^n \sum_{j=n+1}^{m-1} \pi H_{j,i} \right)^2 \right\} + E \left\{ \left(\sum_{n=1}^m d_i^n \sum_{j=n}^{m-1} \pi H_{j,i} \right)^2 \right\} \quad 2.51$$

The noise z_i^n will be assumed white as in Section 2.2.
 It will also be assumed that $H_{j,i}$ are independent for different values of j . This can be accomplished in practice by updating the estimates of the parameters every M sample times instead of each sample time. M is chosen large enough to guarantee the independence and usually $M > 1$ anyway because it may not be possible to implement the adaptive algorithms at the data rate.

Since the sequence z_i^n is white

$$E \left\{ \left(\sum_{n=0}^{m-1} z_i^n \sum_{j=n+1}^{m-1} \pi H_{j,i} \right)^2 \right\} = \sigma_{z_i}^2 \sum_{n=0}^{m-1} \sum_{j=n+1}^{m-1} \pi^2 E \{ H_{j,i}^2 \} \quad 2.52$$

where

$$\sigma_{z_i}^2 = E \{ (z_i)^2 \}$$

and the independence of the $H_{j,i}$ has been used.
 Because the optimum parameters are stationary,

$$E \{ H_{j,i}^2 \} = \overline{H_i^2} \quad \text{independent of } j. \quad \text{Thus the first}$$

term on the right of Eqn. 2.51 simplifies to

$$\alpha_i^2 \sigma_{z_i}^2 \sum_{n=0}^{m-1} (\overline{H_i^2})^n = \alpha_i^2 \sigma_{z_i}^2 \left(\frac{1 - (\overline{H_i^2})^m}{1 - \overline{H_i^2}} \right) \quad 2.53$$

As m becomes large, the residual MSE due to system noise reaches a steady state value of

$$\alpha_i^2 \sigma_{z_i}^2 \left(\frac{1}{1 - \overline{H_i^2}} \right) \quad 2.54$$

The second term of Eqn. 2.51 may be expanded as

$$\begin{aligned} & E \left\{ \left(\sum_{n=1}^m d_i^n \frac{1}{\pi} \sum_{j=n}^{m-1} H_{j,i} \right) \left(\sum_{k=1}^m d_i^k \frac{1}{\pi} \sum_{j=k}^{m-1} H_{j,i} \right) \right\} \\ &= \sum_{n=1}^m \sum_{k=1}^m E \left\{ d_i^n d_i^k \right\} E \left\{ \frac{1}{\pi} \sum_{j=n}^{m-1} H_{j,i} \frac{1}{\pi} \sum_{j=k}^{m-1} H_{j,i} \right\} \end{aligned}$$

using the independence of the d_i^j and $H_{j,i}$. If the variables H_j are independent, this can be factored into

$$\begin{aligned} & E \left\{ d_i^2 \right\} \sum_{j=0}^{m-1} (\overline{H_i^2})^j + E \left\{ d_i^2 \right\} 2 \sum_{j=1}^{m-2} \rho_{j,i} (\overline{H_i})^j \sum_{n=0}^{m-j-1} (\overline{H_i^2})^n \\ &= \frac{\sigma_{d_i}^2}{1 - \overline{H_i^2}} + 2 \sigma_{d_i}^2 \sum_{j=1}^{m-2} \rho_{j,i} (\overline{H_i})^j \sum_{n=0}^{m-j-1} (\overline{H_i^2})^n \text{ for large } m, \end{aligned}$$

where

$$\sigma_{d_i}^2 = E \left\{ d_i^2 \right\}$$

$$\text{and } \rho_{j,i} = \frac{1}{E\{d_i^2\}} E\{d_i^{j+m} d_i^m\} = \frac{1}{E\{d_i^2\}} E\{d_i^{-j+m} d_i^m\}$$

Thus Eqn. 2.51 may be written as

$$\begin{aligned} \sigma_{p_i}^2 = E\{(e_{p_i}^m)^2\} &= \frac{\alpha_i^2 \sigma_{z_i}^2}{1 - \overline{H_i^2}} + \frac{\sigma_{d_i}^2}{1 - \overline{H_i^2}} \\ &+ 2 \sigma_{d_i}^2 \sum_{j=1}^{m-2} \rho_{j,i} (\overline{H_i})^j \sum_{n=0}^{m-j-1} (\overline{H_i^2})^n \end{aligned} \quad 2.55$$

If the time variations were not present, the MSE would reduce to that given by Eqn. 2.35 of Section 2.2.

The mean square parameter error is bounded and the adaptive system will thus be stable provided α_i is chosen so that $|\overline{H_i}| < 1$ and $|\overline{H_i^2}| < 1$. The optimum value for α_i can be found by minimizing Eqn. 2.55 with respect to α_i . Note that $\overline{H_i}$ and $\overline{H_i^2}$ are also functions of α_i (see Eqn. 2.46).

If the form of the time variations is known, the second and third terms of Eqn. 2.55 can be evaluated and an expression for α_i obtained. For example if g_i is correlated only over one step, the third term of Eqn. 2.55 may be easily evaluated. (See Section 4.1). In other cases, only the first few $\rho_{j,i}$ of the third term may be important and the sum may be approximated.

In practice the real problem is that the nature of the time variations is not known. This has been a basic assumption in all the preceding discussion. It would thus appear that some procedure for estimating α_i (possibly adaptively) may be necessary for those systems in which the time variations are rapid. If the parameters are slowly varying, the system will be able to track even for a small α_i . Some other criterion is therefore needed to evaluate the adaptive system response to time varying parameters for given step size values. Such criteria may include response to step changes in the parameters, or tracking performance for sinusoidal parameter variations. There remains however, the problem of determining these α_i requirements from some given unknown environment.

The important result here is that the algorithms are stable and capable of tracking time varying environments. In Chapter 4 these results are verified by some simple computer simulations. The tracking requirements are not as severe here as the convergence requirements of K-W or R-M processes because there is no requirement for the steady state MSE to become zero. Consequently stability for tracking systems is more easily demonstrated, and is more easily

achieved in practice.

The finite MSE allowed opens up other possibilities for the selection of optimum criteria. In certain cases it may be possible to get gradient estimates that are less noisy by forcing the system to converge to values that are not optimum. The improved tracking ability may more than offset this, resulting in improved performance. An example of this can be found in (11) where a forward equalizer's taps were found to adapt more rapidly when the algorithm used the receiver's decision estimates rather than the actual signal values because the decisions were less "noisy" at the low error rates involved. It is easy to demonstrate stability for such a system, but convergence in the normal sense would obviously not be possible.

2.5 Summary

This chapter provided a short review of some of the gradient techniques that appear most promising for the type of problems considered in this thesis. These methods are easy to implement and appear to be sufficiently flexible for a variety of problems.

Some modifications to the basic algorithms were introduced; in particular one for the K-W procedure requiring only half the additional circuitry. Simple proofs of convergence were given for the R-M algorithm with fixed step sizes. A proof for the modified K-W technique will be given in Section 3.2. References were made to more general and rigorous proofs in the literature.

Some original results on adaption for time-varying systems were given and it was demonstrated that a fixed step size algorithm also converges for these systems. It was shown that the tracking requirements are much less severe than those for asymptotic convergence. Determination of an optimum set of step sizes is quite difficult even if the form of the time variations is known.

CHAPTER 3

ADAPTIVE TECHNIQUES FOR PAM COMMUNICATION RECEIVERS

In this chapter detailed calculations will be made for simple PAM communication receivers for systems of the type outlined in Section 1.3. The equations of Chapter 2 apply in most cases although some modifications are necessary. In Sections 3.2 and 3.3 computer simulation results are given for two specific examples. As many general conclusions as possible will be made about the operation of adaptive receiving filters on the basis of these examples. The received pulses are assumed to be independent and non overlapping for these examples. The effects of intersymbol interference and the adaption of receivers for this type of system are considered in Chapter 5.

3.1 The Basic PAM Communication Receiver

A simple digital PAM Communication System was shown in Figure 1.2 and described in Section 1.3. The receiver for this system is shown in Figure 3.1.

Figure 3.1 A Simple PAM Receiver

This receiver is made adaptive by adding the parameter adjusting equipment indicated in Figure 3.1. The analog receive filter contains a set of variable parameters which are to be optimized by the adaptive circuitry. In addition, the sample time will be selected to give optimum performance as measured by some suitable criterion.

The desired received signal is $r(t)$ where, for a PAM system,

$$r(t) = \theta_j p(t) \quad 3.1$$

and the pulse amplitude θ_j is one of a discrete set of values θ_i , $i = 1, 2, \dots, N$ each of which is assumed to occur with equal probability of $1/N$, and independently of all others. (These assumptions are made only to simplify the analysis.) The pulse shape $p(t)$ may be unknown or time varying or both. It is assumed to be of finite duration and the pulses occur with sufficient time separation to prevent overlap at the filter output at the sampling time T_s .

This received signal is masked by additive stationary Gaussian noise $n(t)$. In addition it is assumed that this noise is white. This assumption, and the Gaussian one, are not necessary but simplify the analysis. In addition, only zero mean noise is

allowed. Therefore

$$E\{n(t)\} = 0 \quad 3.2$$

$$\begin{aligned} R_n(t, t') &= R_n(t-t') \\ &= E\{n(t)n(t')\} \\ &= N_0 \delta(t-t') \end{aligned} \quad 3.3$$

where N_0 is the noise variance.

The filter output is sampled at time T_s . If the pulses are transmitted each T seconds, then the receiver will sample each T seconds. It is necessary however, to adjust the sample time with respect to the peak, or the beginning of the received pulse in order to obtain optimum performance.

The filter output is $x(t) + y(t)$ where

$$x(t) = \int_{-\infty}^t h(\tau)r(t-\tau) d\tau$$

$$\text{and } y(t) = \int_{-\infty}^t h(\tau)n(t-\tau) d\tau .$$

The sampled values are then quantized to one of the N θ_i values and this is used as an estimate, $\hat{\theta}$, of the transmitted pulse amplitude parameter. By subtracting this quantized estimate from the analog estimate $x(T_s) + y(T_s)$, an analog error is formed. This error could be used to monitor system performance and to adapt the receiver parameters.

3.1.1 Choice of a Performance Criterion

For many digital systems, the probability of error (P_e) is an important measure of system performance. However, it is not possible to measure P_e unless the transmitted symbols are known at the receiver in order for a comparison to be made. Clearly if this were the case the communication system would not be required! This situation can be approximated by using a training sequence along with the desired information. However, this degrades the information rate of the system. In addition a long-term average is required to approximate the P_e and so such a performance measure is unsuitable for transient and tracking system evaluation.

The mean squared error (MSE) is a common performance criterion which is also easy to use in analytical derivations. For a digital system this means the MSE at the sampling time. For the system of Figure 3.1 this would be $E\{e^2(T_s)\}$.

If the additive noise is Gaussian (as assumed here) then minimizing the MSE is equivalent to minimizing $P_e^{(1)}$. To get a good estimate of the MSE several samples of the instantaneous squared error

must be averaged. The greater the number of samples the more accurate the estimate. This inherent averaging requirement of the MSE means that the adaptive systems which estimate MSE are constrained to act fairly slowly.

For simple systems such as we have here, there is another possible criterion. It is well known that the optimum receive filter for a PAM system is a matched filter⁽¹⁾. A matched filter also provides the maximum signal to noise ratio (SNR) at the sampling time⁽¹⁾. Since this optimum filter provides maximum SNR, it seems very reasonable to adjust the receive filter to give maximum SNR as well. If this filter is of a form that allows it to become matched for the optimum parameter set $\bar{p}^*$ then the adaptive system will in fact adapt the filter into a matched filter. If its constrained form will not allow it to become matched its performance will usually still be very close to the best that this constrained filter can do.

Another advantage of the maximum SNR criterion is that it does not require any inherent averaging. The estimates given by the instantaneous SNR at one sample time are accurate enough.¹ For simple receive

¹ Note that for $SNR = \frac{(\text{signal})^2}{\sigma_y^2}$ this assumes that the noise variance is constant and that the signal estimates $\hat{\theta}$ are usually correct.

filters it is possible to use a modified receive filter that has an output noise variance, σ_y^2 , that is independent of the filter parameters. (This is discussed further in Section 3.2). For this modified filter, maximum signal out at the sample time is the same as maximum SNR (since the noise variance is constant) and this criterion will be shown to be very easy to implement (see Section 3.2).

Minimum mean square error and maximum SNR are thus expected to be common performance objectives in many systems. More discussion of their relative merits is given in Section 3.2.

3.1.2 Estimation of the Gradients

As discussed in Chapter 1 the adaptive techniques to be studied in this thesis are the simple gradient techniques. Of most use for physical systems, in which the gradient must be estimated by implementing difference equations, are the Kiefer-Wolfowitz type algorithms with fixed step size parameters. This algorithm was discussed in Section 2.1.2.

A difference equation estimate can be implemented as shown in Figure 3.2.

Figure 3.2 Implementation of a Difference Equation

This method requires two receive filters in addition to the filter in Figure 3.1 for each parameter being adapted. It gives a noisy but unbiased estimate of the derivative of the receive filter output with respect to the parameter p_i at the sampling time. For filters where the output noise variance is not a function of $\bar{p}$ as mentioned in the previous section, this derivative estimate is directly related to the derivative of SNR with respect to p_i . More detail will be given in Section 3.2.

The method of Figure 3.2 will give derivative estimates for non linear functions of the parameters, and for functions which have discontinuous derivatives. Thus this difference technique is applicable to a wide range of problems. In order to reduce the complexity only one additional filter could be used along with the receive filter as illustrated in Figure 3.3. This gives a one sided difference approximation and results in biased estimates. However, this bias is often small and this technique will be used in the simulation results of Section 3.2.

Another possibility for reducing equipment complexity is to use one additional receive filter

Figure 3.3 One Sided Difference Equation Implementation

sequentially. Its parameter values would be set to $p_i + \delta_i$ for one sample and $p_i - \delta_i$ for the next. The difference of these two samples would be used to form the difference estimate. This scheme requires more control logic and would react much more slowly because of its sequential nature.

One other way of obtaining derivative estimates is to use a filter which gives the desired derivative as an output. For example, when derivatives with respect to sampling time are required a standard analog computer differentiator can be used and its output sampled at the proper time. These type of estimates would be used with a R-M type algorithm since they are not difference approximations. Unfortunately the differentiator output will be very noisy. In addition, for variables other than time, it is usually very difficult to find a filter that will give the derivative.

A specific example will now be examined in detail.

3.2 Rectangular Received Pulses and a Low Pass Receive Filter

This section considers a particular PAM system

to illustrate the derivation of the adaptive gradient algorithms; to show how they may be easily and practically implemented and to demonstrate the performance improvement that can be provided. The receiver will consist of an analog filter which has adjustable gain, sampling time and bandwidth. The algorithms are derived in Sections 3.2.1 to 3.2.3 respectively and then used in a computer simulation.

As a specific example let the receive filter be a single pole low pass filter of the form

$$h(t) = A \sqrt{\alpha} e^{-\alpha t} \quad 3.4$$

A is a gain parameter which will be adapted to scale the filter output to the proper magnitude with respect to the fixed quantizer values. It will be selected to minimize the mean squared error $E\{e^2\}$. The parameter α is a receive filter bandwidth parameter that will be adjusted to provide maximum SNR at the sample time. The best sample time will also be selected by an adaptive loop to give maximum SNR at the sample time. A binary PAM system will be examined so the quantizer has only two output values which are set to ± 1 for scaling convenience. Note that different performance criteria are used for different

parameters.

The overall receiver is shown in Figure 3.4. The received signal $r(t)$ is of the form

$$r(t) = \theta P \quad 0 \leq t \leq T_1 \quad 3.5$$

where T_1 is the pulse duration when the leading edge of the pulse is assumed to arrive at the receiver at time $t = 0$. For this example it is assumed that the pulses are transmitted at the rate of one each T seconds.

Here θ is a binary random variable taking the values ± 1 with equal probability of $\frac{1}{2}$. In addition the θ 's are assumed to be independent from pulse to pulse. P is an amplitude factor introduced by the channel. In general it may be unknown, or time varying, or both. In addition the pulse duration T_1 may be unknown or time varying, or both. However, for this example the parameters are unknown, but fixed. The channel noise $n(t)$ is assumed to be stationary additive white Gaussian with zero mean and variance N_0 . It is also independent of the transmitted θ values. For this example

$$x(T_s) = \theta P \int_0^{T_s} \sqrt{\alpha} e^{-\alpha(T_s - \tau)} d\tau$$

Figure 3.4 A Simple Adaptive Binary PAM Receiver

$$= \frac{\theta P}{\sqrt{\alpha}} (1 - e^{-\alpha T_S}) \text{ for } 0 \leq T_S \leq T_1 \quad 3.6(a)$$

$$= \frac{\theta P}{\sqrt{\alpha}} e^{-\alpha T_S} (e^{\alpha T_1} - 1) \text{ for } T_S \geq T_1 \quad 3.6(b)$$

$$y(T_S) = \int_{-\infty}^{T_S} n(\tau) \sqrt{\alpha} e^{-\alpha(T_S - \tau)} d\tau \quad 3.7$$

Here $x(T_S)$ is the signal portion of the filter output due to $r(t)$ and $y(T_S)$ is the output portion due to $n(t)$.

The individual algorithms will now be examined in detail. It will be assumed that the adaptive loops operate essentially independently.

3.2.1 Adaptive Adjustment of the Filter Gain

The performance measure used in adapting A will be minimum mean square error

$$E\{e^2(T_S)\} = E\{(Ax(T_S) + Ay(T_S) - \hat{\theta})^2\} \quad 3.8$$

In Equation 3.8 it is assumed that the system is working quite well and that its probability of error is quite small. This means that $\hat{\theta}$ is a correct estimate of the transmitted symbol θ essentially all of the time. The use of $\hat{\theta}$ instead of the true value of θ then gives an excellent approximation to the true MSE.

The use of mean squared error implies long term averaging of many samples of squared error. An unbiased estimate of MSE will be used instead, in particular the instantaneous squared error. The average of only a few samples would give a better estimate; but would also result in slower adaption. More will be said about averaging time in a later section.

The performance criterion is therefore

$$e^2(T_S) = (Ax(T_S) + Ay(T_S) - \hat{\theta})^2 \quad 3.9$$

From Equation 2.4 of Section 2.1.1 the R-M type algorithm gives a parameter iteration of

$$\begin{aligned} A^{k+1} &= A^k + \epsilon \left. \frac{\partial e^2(T_S)}{\partial A} \right|_{A = A^k ; T_S = T_S^k} \\ &= A^k + \epsilon \frac{\partial}{\partial A} \left[A^k x(T_S^k) + A^k y(T_S^k) - \hat{\theta}^k \right]^2 \end{aligned} \quad 3.10$$

Equation 3.10 says that the value of A to be used for the reception of the $k+1^{\text{st}}$ received pulse is that value used for the k^{th} pulse incremented by ϵ times the derivative of the performance measure (squared error) with respect to A. This derivative is estimated at time k. Here ϵ is a step size parameter which will be assumed fixed in value.

Expanding Eqn. 3.10 yields

$$A^{k+1} = A^k + 2\epsilon \left[A^k x(T_S^k) + A^k y(T_S^k) - \hat{\theta}^k \right] \left[x(T_S^k) + y(T_S^k) \right] \quad 3.11(a)$$

Since all the terms on the right side of Eqn. 3.11(a) are available, the adaptive loop for adjusting A is easily implemented. Notice however that it is assumed that the filter output can be measured both before and after multiplication by the filter gain parameter A. If A were an integral part of the filter so that only the term $Ax(T_S) + Ay(T_S)$ were available then Eqn. 3.11(a) would have to be written as

$$A^{k+1} = A^k + 2\epsilon \left[A^k x(T_S^k) + A^k y(T_S^k) - \hat{\theta}^k \right] \cdot \left[A^k x(T_S^k) + A^k y(T_S^k) \right] \frac{1}{A^k} \quad 3.11(b)$$

This algorithm is nonlinear in A and hence very difficult to analyse. The equations of Chapter 2 cannot be applied. It is therefore important to build the receive filter in a way that makes adaption easy.

Let the optimum gain parameter be A^* . This is the value of A that minimizes the mean squared error of Eqn. 3.8 and is given by

$$\left. \frac{\partial E\{e^2(T_S)\}}{\partial A} \right|_{A=A^*} = 0 = 2E\{(A^* x(T_S) + A^* y(T_S) - \hat{\theta})(x(T_S) + y(T_S))\} \quad 3.12(a)$$

$$\therefore A^*(E\{x^2\} + E\{y^2\}) = E\{\hat{\theta} x(T_S)\} \quad 3.12(b)$$

Assuming a correct decision, $E\{\hat{\theta} x(T_S)\} = |x(T_S)|$

$$\therefore A^* = \frac{|x(T_S)|}{x^2 + \sigma_y^2} \quad 3.13$$

where $\sigma_y^2 = E\{y^2(T_S)\}$ and $E\{x^2\} = x^2$ since $x^2(T_S)$ is deterministic. Now define the gain parameter error at time k as

$$e_A^k = A^k - A^* \quad 3.14$$

Using Eqn. 3.11(a) we can write

$$\begin{aligned} e_A^{k+1} &= A^{k+1} - A^* \\ &= A^k - A^* + 2\epsilon A^k [x(T_S^k) + y(T_S^k)]^2 - 2\epsilon \hat{\theta}^k [x(T_S^k) + y(T_S^k)] \end{aligned}$$

Adding and subtracting $2\epsilon A^* [x(T_S^k) + y(T_S^k)]^2$ and regrouping gives

$$\begin{aligned} e_A^{k+1} &= e_A^k + 2\epsilon(A^k - A^*) [x(T_S^k) + y(T_S^k)]^2 + 2\epsilon [A^* (x(T_S^k) + y(T_S^k))^2 \\ &\quad - \hat{\theta}^k (x(T_S^k) + y(T_S^k))] \end{aligned}$$

$$\begin{aligned}
&= e_A^k \left[1 + 2\epsilon (x(T_S^k) + y(T_S^k))^2 \right] \\
&\quad + 2\epsilon \left[A^*(x(T_S^k) + y(T_S^k))^2 - \hat{\theta}(x(T_S^k) + y(T_S^k)) \right]
\end{aligned} \tag{3.15(a)}$$

Using the notation

$$x^k = x(T_S^k)$$

and $y^k = y(T_S^k)$

Eqn. 3.15(a) can be rewritten more conveniently as

$$e_A^{k+1} = e_A^k (1 + 2\epsilon (x^k + y^k)^2) + 2\epsilon \left[A^*(x^k + y^k)^2 - \hat{\theta}^k(x^k + y^k) \right] \tag{3.15(b)}$$

Eqn. 3.15(b) is the same form as Eqn. 2.21 where the equivalent additive noise term is

$$z_A^k = 2 \left[A^*(x^k + y^k)^2 - \hat{\theta}^k(x^k + y^k) \right] \tag{3.16}$$

It is shown in Appendix A that z_A above satisfies all the requirements placed on z_i of Eqn. 2.21. The results of Section 2.2 can be applied directly. The steady state mean error will therefore tend exponentially to zero. The adaptive system will be stable provided the step size parameter satisfies

$$0 < \epsilon < \frac{-2}{2(x^2 + y^2)} = \frac{-1}{x^2 + \sigma_y^2} \tag{3.17}$$

and the convergence will be very rapid initially if

$$\epsilon \approx \frac{-1}{2(x^2 + y^2)} \tag{3.18}$$

By substituting the appropriate values into Equation 2.34 we find that the steady state mean square error is

$$E\{e_A^2\} = \frac{4\epsilon^2(A^*(x+y)^2 - \hat{\theta}(x+y))^2}{1 - (1+2\epsilon(x+y)^2)^2} \quad 3.19(a)$$

In Appendix B it is shown that this can be reduced to

$$E\{e_A^2\} = \frac{-\epsilon\sigma_y^2(x^2 + \sigma_v^4)}{(x^2 + \sigma_y^2)^2(x^2 + \sigma_y^2 + \epsilon(x^4 + 6x^2\sigma_y^2 + 3\sigma_y^4))} \quad 3.19(b)$$

These results; like those of Section 2.2 are for acquisition problems only. Consideration of time varying tracking problems is left for Chapter 4. Note that none of the results in this section depend on the explicit form of x^k .

3.2.2 Adaptive Adjustment of the Sampling Time

The sampling time; T_s will be selected to give maximum SNR at the filter output at the sampling instant. Here we will use

$$SNR = \frac{E\{x^2(T_s)\}}{E\{y^2(T_s)\}} = \frac{\overline{x^2}}{\sigma_y^2} \quad 3.20$$

Since the filter output noise variance is not a function of T_s ; the SNR can be maximized by maximizing the

magnitude of the filter output at time T_S . Only a noisy estimate of the output is available so the adaptive loop must use the unbiased estimate

$$x(T_S) + y(T_S)$$

The derivative of the output at time k will be approximated by the difference equation

$$\frac{x(T_S^k + \delta) - x(T_S^k - \delta)}{2\delta} \hat{\theta}^k + \frac{y(T_S^k + \delta) - y(T_S^k - \delta)}{2\delta} \hat{\theta}^k \quad 3.21$$

Because it is desired to maximize the magnitude of the filter output; and the system is a binary one; it is necessary to take account of the sign of the filter output.

The magnitude of the derivative cannot be used since this is always positive by definition. However the effect of the received pulse sign can be removed by multiplying the derivative by the estimate $\hat{\theta}$, which is ± 1 . (If there are no decision errors this is the required derivative.) As long as the system is working reasonably well, the errors will seldom cause a step in the wrong direction.

The difference equation is implemented by sampling the output δ sec. before, and the δ sec. after the current estimate of the sampling time. Thus

for sample time adaption no additional receive filters are required. It is only necessary to store the sample at $T_S - \delta$ until the one at $T_S + \delta$ is taken. Thus a K-W type process as described in Section 2.1.2 will be used. The recursive relation for the sample time estimate is

$$T_S^{k+1} = T_S^k + \beta \left[\frac{x(T_S^k + \delta) - x(T_S^k - \delta)}{2\delta} \right] \hat{\theta}^k + \beta \left[\frac{y(T_S^k + \delta) - y(T_S^k - \delta)}{2\delta} \right] \hat{\theta}^k \quad 3.22$$

As before, β is a step size parameter. Eqn. 3.22 may be simplified to

$$T_S^{k+1} = T_S^k + \beta \left[\frac{x(T_S^k + \delta) - x(T_S^k - \delta)}{2\delta} \right] \hat{\theta}^k + \beta z_T^k \quad 3.23$$

where

$$z_T^k = \left[\frac{y(T_S^k + \delta) - y(T_S^k - \delta)}{2\delta} \right] \hat{\theta}^k \quad 3.24$$

Using Taylor Series expansions about the optimum value $x(T_S^*)$ for small δ gives

$$x(T_S^k + \delta) \approx x(T_S^*) + \left. \frac{\partial x(T_S)}{\partial T_S} \right|_{T_S = T_S^*} (T_S^k + \delta - T_S^*) + \frac{1}{2} \left. \frac{\partial^2 x(T_S)}{\partial T_S^2} \right|_{T_S = T_S^*} (T_S^k + \delta - T_S^*)^2 \quad 3.25$$

$$x(T_S^k - \delta) \approx x(T_S^*) + \left. \frac{\partial x(T_S)}{\partial T_S} \right|_{T_S = T_S^*} (T_S^k - \delta - T_S^*) + \frac{1}{2} \left. \frac{\partial^2 x(T_S)}{\partial T_S^2} \right|_{T_S = T_S^*} (T_S^k - \delta - T_S^*)^2 \quad 3.26$$

Figure 3.5 displays the relationship among the variables.

$$\text{Now } \left. \frac{\partial x(T_S)}{\partial T_S} \right|_{T_S=T_S^*} = 0 \quad \text{since } x(T_S^*) \text{ is a minimum with respect}$$

to T_S .

Define

$$e_T^k \triangleq T_S^k - T_S^* . \quad 3.27$$

Subtracting Eqn. 3.26 from Eqn. 3.25 gives

$$x(T_S^k + \delta) - x(T_S^k - \delta) \approx \frac{1}{2} \left. \frac{\partial^2 x(T_S)}{\partial T_S^2} \right|_{T_S=T_S^*} (4e_T^k \delta) \quad 3.28$$

Substituting Eqn. 3.28 into Eqn. 3.23 yields

$$T_S^{k+1} = T_S^k + \beta \left. \frac{\partial^2 x(T_S)}{\partial T_S^2} \right|_{T_S=T_S^*} (e_T^k) + \beta z_T^k \quad 3.29$$

Subtracting T_S^* from both sides yields an equation of the form of Eqn. 2.21:

$$e_T^{k+1} = e_T^k \left(1 + \beta \left. \frac{\partial^2 x(T_S^k)}{\partial T_S^2} \right|_{T_S=T_S^*} \right) + \beta z_T^k \quad 3.30$$

In Appendix C it is shown that z_T satisfies the requirements placed on z_i of Eqn. 2.21. The results of Section 2.2 can now be applied.

Figure 3.5 Variable Relationships for Eqns. 3.25, 3.26

The mean error $E \{e_T^m\}$ will converge to zero if

$$0 > \beta > \frac{-2}{\left. \frac{\partial^2 x(T_S)}{\partial T_S^2} \right|_{T_S=T_S^*}} \quad 3.31$$

The steady state mean square error is obtained by using Eqn. 2.34

$$E \{e_T^2\} = \frac{\beta^2 \sigma_{z_T}^2}{1 - \left(1 + \beta \left. \frac{\partial^2 x(T_S)}{\partial T_S^2} \right|_{T_S=T_S^*}\right)^2} \quad 3.32$$

where

$$\sigma_{z_T}^2 = E \{(z_T)^2\} = \frac{N_0}{4\delta^2} \left[1 - e^{-2\alpha\delta}\right] \quad 3.33$$

from Appendix C where δ is a small fixed positive quantity.

3.2.3 Adaptive Adjustment of Receive Filter Bandwidth

For this example the receive filter itself has only one adjustable parameter, the bandwidth (or filter time constant). The signal portion of the filter

output at the sampling time is

$$\begin{aligned} x(T_s; \alpha) &= \frac{\theta P}{\sqrt{\alpha}} (1 - e^{-\alpha T_s}) && \text{for } T_s \leq T_1 \\ &= \frac{\theta P}{\sqrt{\alpha}} e^{-\alpha T_s} (e^{\alpha T_1} - 1) && \text{for } T_s > T_1 \end{aligned}$$

The noise variance at the filter output is

$$\sigma_y^2 = \frac{N_0}{2}.$$

The parameter α will be adapted to achieve maximum SNR at the sampling time where

$$\text{SNR} = 2x^2(T_s)/N_0.$$

Note that the noise variance is independent of α . Thus a simpler criterion, equivalent to maximum SNR, is maximum absolute output at the sample time.

An estimate of the derivative of the output with respect to α will be obtained by using a one sided difference equation approximation as shown in Figure 3.3. This requires only one additional receiver filter but introduces a small bias into the estimate.

The recursive relationship for α is

$$\alpha^{k+1} = \alpha^k + \gamma \frac{x(T_s^k; \alpha^{k+\lambda}) - x(T_s^k; \alpha^k)}{\lambda} + \gamma z_\alpha^k \quad 3.34$$

where

$$z_{\alpha}^k = \frac{1}{\lambda} (y(T_S^k ; \alpha^{k+\lambda}) - y(T_S^k ; \alpha^k)) \quad 3.35$$

γ is the step size parameter and λ is a small positive perturbation on α in the additional receiver filter.
(cf. Figure 3.3)

This is the technique of Section 2.1.3. When the system is fairly close to its optimum settings the following Taylor Series approximation can be made:

$$x(T_S^k ; \alpha^{k+\lambda}) \approx x(T_S^k ; \alpha^*) + \frac{1}{2} \frac{\partial^2 x(T_S^k ; \alpha^k)}{\partial \alpha^2} \bigg|_{\alpha^k = \alpha^*} (\alpha^k + \lambda - \alpha^*)^2 \quad 3.36$$

where α^* is the optimum value of α and

$$\frac{\partial x(T_S^k ; \alpha^k)}{\partial \alpha} \bigg|_{\alpha^k = \alpha^*}$$

was taken as zero. Similarly

$$x(T_S^k ; \alpha^k) \approx x(T_S^k ; \alpha^*) + \frac{1}{2} \frac{\partial^2 x(T_S^k ; \alpha^k)}{\partial \alpha^2} \bigg|_{\alpha^k = \alpha^*} (\alpha^k - \alpha^*)^2 \quad 3.37$$

Defining $e_{\alpha}^k \triangleq \alpha^k - \alpha^*$ and subtracting Eqn. 3.37 from

Eqn. 3.36 yields

$$x(T_S^k ; \alpha^{k+\lambda}) - x(T_S^k ; \alpha^k) \approx \frac{1}{2} \left. \frac{\partial^2 x(T_S^k ; \alpha^k)}{\partial \alpha^2} \right|_{\alpha^k = \alpha^*} (\lambda^2 + 2\lambda e_\alpha^k) \quad 3.38$$

Substituting into Eqn. 3.34 and subtracting α^* from both sides gives

$$e_\alpha^{k+1} = e_\alpha^k \left(1 + \gamma \left. \frac{\partial^2 x(T_S^k ; \alpha^k)}{\partial \alpha^2} \right|_{\alpha^k = \alpha^*} \right) + \frac{\gamma \lambda}{2} \left. \frac{\partial^2 x(T_S^k ; \alpha^k)}{\partial \alpha^2} \right|_{\alpha^k = \alpha^*} + \gamma z_\alpha^k \quad 3.39$$

This approach introduces a bias term, but because γ and especially λ are usually small the bias term is usually quite small.

For ease of calculation define

$$H_k \triangleq 1 + \gamma \left. \frac{\partial^2 x(T_S^k ; \alpha^k)}{\partial \alpha^2} \right|_{\alpha^k = \alpha^*} \quad 3.40$$

and

$$X_k \triangleq \frac{\gamma \lambda}{2} \left. \frac{\partial^2 x(T_S^k ; \alpha^k)}{\partial \alpha^2} \right|_{\alpha^k = \alpha^*} \quad 3.41$$

Eqn. 3.39 can now be rewritten as

$$e_\alpha^{k+1} = e_\alpha^k H_k + X_k + \gamma z_\alpha^k \quad 3.42$$

Taking expected values

$$E\{e_{\alpha}^{k+1}\} = E\{e_{\alpha}^k\} \bar{H} + \bar{X}$$

3.43

since $E\{z_{\alpha}^k\} = 0$; $\bar{H}_k = \bar{H}$; and $\bar{X}_k = \bar{X}$.

$$\begin{aligned} \therefore E\{e_{\alpha}^m\} &= E\{e_{\alpha}^0\} \bar{H}^m + \bar{X} \sum_{j=0}^{m-1} \bar{H}^j \\ &= E\{e_{\alpha}^0\} \bar{H}^m + \frac{\bar{X}(1-\bar{H}^{m-1})}{1-\bar{H}} \end{aligned}$$

Thus if $|\bar{H}| < 1$ the steady state (m large) mean error is

$$E\{e_{\alpha}\}_{ss} = \frac{\bar{X}}{1-\bar{H}}$$

3.44(a)

$$\begin{aligned} &= \frac{\frac{\gamma\lambda}{2} \frac{\partial^2 x(T_S^k; \alpha^k)}{\partial \alpha^2} \Big|_{\alpha^k=\alpha^*}}{-\gamma \frac{\partial^2 x(T_S^k; \alpha^k)}{\partial \alpha^2} \Big|_{\alpha^k=\alpha^*}} \\ &= -\lambda/2 \end{aligned}$$

3.44(b)

Thus the steady state mean error is not zero because of the bias introduced by the one sided difference equation approximation. This could be eliminated by letting γ go to zero, but it is not possible to do this for tracking problems. The condition for

stability is just $|\bar{H}| < 1$ or

$$0 > \gamma > \frac{\frac{\partial^2 x(T_s; \alpha)}{\partial \alpha^2}}{\alpha = \alpha^*} \quad 3.45$$

as expected.

The mean square parameter error is

$$\begin{aligned} E\{(e_{\alpha}^{k+1})^2\} &= E\{(e_{\alpha}^k)^2\} \bar{H}^2 + \bar{X}^2 + \gamma^2 \sigma_{z_{\alpha}}^2 + 2\gamma E\{z_{\alpha}^k e_{\alpha}^k H_k\} \\ &\quad + 2E\{e_{\alpha}^k X_k H_k\} + 2\gamma \bar{X} E\{z_{\alpha}^k\} \end{aligned} \quad 3.46$$

Now z_{α}^k is independent of e_{α}^k and X_k , and has zero mean. The independence requires z_{α}^k to be white; this can be achieved by using samples several baud times apart.

$$\therefore E\{(e_{\alpha}^{k+1})^2\} = E\{(e_{\alpha}^k)^2\} \bar{H}^2 + \bar{X}^2 + \gamma^2 \sigma_{z_{\alpha}}^2 + 2\gamma \bar{X} E\{e_{\alpha}^k\} \quad 3.47$$

The solution to Eqn. 3.47 is

$$\begin{aligned} E\{(e_{\alpha}^m)^2\} &= E\{(e_{\alpha}^0)^2\} \bar{H}^{2m} + (\bar{X}^2 + \gamma^2 \sigma_{z_{\alpha}}^2) \frac{1 - \bar{H}^{2m}}{1 - \bar{H}^2} \\ &\quad + 2\gamma \bar{X} \sum_{j=0}^{m-1} \bar{H}^{2j} E\{e_{\alpha}^{m-1-j}\} \end{aligned} \quad 3.48$$

Now $|\bar{H}| < 1$ and $\bar{H}^2 < 1$ by Eqn. 3.45. Thus the first term tends to zero for large m . Considering the last term,

$\overline{H^2}^j$ is very small for large j ; and $E\{e_\alpha^{m-1-j}\}$ tends to $E\{e_\alpha\}_{ss}$ for small j (since m is large). Thus the last term is approximately

$$\begin{aligned} & 2\overline{XH} E\{e_\alpha\}_{ss} \sum_{j=0}^M \overline{H^2}^j \\ &= 2\overline{XH} (-\lambda/2) \frac{1 - \overline{H^2}^M}{1 - \overline{H^2}} \\ &= \frac{2\overline{XH}}{1 - \overline{H^2}} (-\lambda/2) \end{aligned}$$

where $0 < M < m$ and M is some moderately large number.

$$\therefore E\{e_\alpha^2\}_{ss} = \frac{1}{1 - \overline{H^2}} (\overline{X^2} + \gamma^2 \sigma_{z_\alpha}^2 - \overline{XH}\lambda)$$

Now

$$\overline{X^2} = \lambda^2/4 \overline{(H-1)^2}$$

$$\therefore \overline{X^2} - \overline{XH}\lambda = \frac{\lambda^2}{4} (1 - \overline{H^2})$$

$$\begin{aligned} \therefore E\{e_\alpha^2\}_{ss} &= \frac{\gamma^2 \sigma_{z_\alpha}^2}{1 - \left(1 + \gamma \frac{\partial^2 x(T_s^k; \alpha^k)}{\partial \alpha^2} \Big|_{\alpha^k = \alpha^*}\right)^2} + \frac{\lambda^2}{4} \quad 3.50 \\ &= \sigma_\alpha^2 + (\text{BIAS})^2 \end{aligned}$$

as expected where σ_{α}^2 is the steady state mean square parameter error that would occur for a double sided difference approximation. It can be shown that

$$\sigma_{z_{\alpha}}^2 = E\{z_{\alpha}^2\} = N_0 \left(1 - \frac{2\sqrt{\alpha(\alpha+\lambda)}}{2\alpha+\lambda}\right) \quad 3.51$$

Because the bias effect is small with respect to the noise variance (σ_{α}^2 term) for most choices γ and λ , the one sided difference approximation technique used in this derivation is quite justified for reducing system complexity. (The computer simulation results confirmed the small effect of the bias).

3.2.4 Discussion of Computer Simulation Results

This section gives a brief discussion of some computer simulation work carried out using the equations developed in the preceding sections. The simulation work was undertaken for the following reasons:

- 1) To show that the algorithms of Section 3.2.1 to 3.2.3 adapt satisfactorily,
- 2) To get some data on the MSE criterion vs the maximum SNR criterion,
- 3) To illustrate the effects of higher order gradients, and
- 4) To get some preliminary tracking performance data.

This example uses rectangular received pulses and a low pass receive filter. The rectangular pulses are not very realistic for a physical system but the results do illustrate, in a simple way, how adaptive systems behave. (A more realistic received pulse is considered in Section 3.3). It is assumed that the received pulses are sufficiently separated in time that they do not overlap at the receiver output and hence there is no intersymbol interference. (The more realistic case including intersymbol interference is considered in Chapter 5).

The receiver is configured as shown in Figure 3.4; and the filter gain A , the sampling time T_s , and the filter bandwidth α , are adapted according to Eqns. 3.11(a), 3.22; and 3.34 respectively. The program also included an option to average L gradient estimates as in Eqn. 2.12 and then iterate the parameters. Thus updates are made only every L sample times. This illustrates the effect of one simple type of higher order gradient method as defined in Section 2.1.4. When $L = 1$; all algorithms are first order. The effects of various order algorithms on transient, steady state, and tracking performance is evaluated.

For this work, the channel amplitude factor, P , of Eqn. 3.5 is fixed equal to 1.0 and therefore ignored.

Representative values for the step size parameters, the noise variance, and other parameters were selected and then fixed for all the simulation work. (Selection of optimum step size parameters for tracking is considered in Chapter 4).

3.2.4.1 General Results

A comparison was made between the analytical and experimental convergence ranges of the step size parameters. In general the experimental ranges were slightly narrower than the analytical results would indicate. This probably results from small parameter interactions, and the use of simplifying assumptions in the analysis. However, practical step size parameters are usually well below the stability limits so the differences between experiment and analysis are not significant. No stability problems were encountered in any of the simulations. Some work was done to check the effects of different step size parameters on the transient or initial acquisition time, and on the steady state mean square

error. The typical characteristic tradeoffs^(10,31) of rapid initial response for larger steady state error were found.

3.2.4.2 Mean Square Error and Maximum Signal to Noise Criteria

In addition to the performance criteria used in the analytical work of Sections 3.2.1 to 3.2.3, some simulation work was done using minimum MSE as a criterion for all three parameters. Transient performance simulations using this criterion showed that considerable parameter interaction occurred and longer acquisition times resulted than occurred when T_s and α were adapted using the maximum SNR (or maximum sample time output) criterion. The acquisition times of T_s and α were found to be quite sensitive to the initial value of A .

When all parameters are trying to force the MSE to zero it is possible to get non optimum sets of parameters that yield a fairly small value of MSE. This is known as a "ridge" characteristic in the performance surface,⁽¹⁴⁾ and is caused by parameter interaction. The simulation results indicated that using different performance criteria for different parameters can greatly reduce the interaction

problem. The results also indicate that the maximum SNR criterion produces much less parameter interaction for this example.

The MSE criterion requires inherent averaging by definition of mean. However, the maximum SNR (or maximum sample time output) criterion does not require inherent averaging and will therefore yield a faster transient response. This was also observed from the simulations. Section 3.2.4.3 illustrates that some averaging, by means of a type of higher order algorithm, can give some performance improvement for both criteria.

3.2.4.3 The Effects of Higher Order Algorithms

In this work, higher order effects were obtained by averaging the gradient estimates over L sample times and iterating the parameters after each L received pulses. The effect of this on transient response and on mean square error is illustrated in Figure 3.6 for a typical set of step sizes. The characteristic tradeoff of convergence time vs steady state error is evident.

More averaging gives better gradient estimates and hence lower MSE. However the rate of adaption

Figure 3.6 MSE vs Convergence Time as a Function of Algorithm Order

is slower so the transient response takes longer. Similar tradeoffs are found as a function of step size parameter values.

Tracking performance has requirements between acquisition and steady state, and so it might be expected that there would be an optimum algorithm order for tracking problems. This was found to be the case as illustrated by the simulation results reported in the next section.

3.2.4.4 Some Preliminary Tracking Results

This section describes some results obtained by using the adaptive system to follow a time varying received signal. The received pulse length was allowed to vary about its nominal value of $T_1 = 1.0$ according to

$$T_1 = 1.0 + 0.5 \sin \left(\frac{2\pi k}{N} \right) \quad 3.52$$

where $k = 0; 1, 2, \dots$ is the time index and N is the cycle time of the variation measured in numbers of received pulses.

The order of the algorithms was also varied from $L = 1$ to $L = 10$. No attempt was made to find the optimum fixed step sizes. A set of step sizes

was selected; by trial and error; to give good results for a cycle time $N = 500$ and a fifth order algorithm. These step size parameters were then used in all the tracking simulations. (Chapter 4 considers the analytic selection of optimum fixed step sizes).

In order to get some appreciation for the performance gain made by the tracking system the instantaneous signal to noise ratio was computed for each step in the iteration and compared with the optimum SNR that could have been attained if the filter were set optimally. This difference in SNR was averaged over the simulation run and is denoted $\overline{\text{SNRD}}$.

The results are presented in Table 3.1. For any cycle time there was found to be an optimum algorithm order giving minimum $\overline{\text{SNRD}}$. For slow variations (large N) the greater averaging of higher order algorithms gives more accurate iterations and a smaller "steady state" type of error. For rapid variations; the algorithms must be able to change the parameters quickly in order to follow changes in the received signal. Thus a lower order algorithm is more suitable. The order must be selected as a compromise between the speed of response and the

<u>Cycle Time (N)</u> <u>(No. of Rec'd Pulses)</u>	<u>Averaging Time (L)</u> <u>(No. of Rec'd Pulses)</u>	<u>SNRD</u> <u>(dB)</u>
1000	10	.286
	5	.307
	2	.414
	1	.707
500	10	.549
	5	.333
	2	.444
	1	.700
100	5	2.232
	3	1.070
	2	.694
	1	.789
50	2	2.246
	1	.976

Table 3.1 Tracking Results

increment accuracy similar to the results of Figure 3.6.

The value of $\overline{\text{SNRD}}$ for the adaptive system was compared with the same measurement for a system whose receiver was fixed at those values that are optimum only when the received pulse length is $T_1 = 1.0$. This allowed the adaptive receiver to be compared with the fixed receiver for various rates of change and the corresponding optimum algorithm orders. The results are given in Table 3.2. The improvement is over 2 dB for the slower variations. The adaptive system is better able to follow slow changes and therefore able to keep the system close to optimum. It would appear from these preliminary results that the simple adaptive system can track fairly large input signal changes and produce a considerable improvement in performance.

A second example that is more realistic is examined in the next section dealing with a second order received pulse and a low pass receive filter. The problem of selecting the optimum fixed step size parameters will be considered in Chapter 4.

<u>Cycle Time (N)</u> <u>(No. of Rec'd Pulses)</u>	<u>Algorithm Order (L)</u> <u>(No. of Rec'd Pulses)</u>	<u>Adapting SNRD</u> <u>(dB)</u>	<u>Fixed SNRD</u> <u>(dB)</u>	<u>Gain</u> <u>(dB)</u>
100	2	.69	2.47	1.78
500	5	.33	2.47	2.14

Table 3.2 Adaptive Receiver Improvement as a Function of Cycle Time

3.3 Second Order Received Pulses and a Low Pass Receive Filter

This section briefly describes another PAM system example that is more realistic because the received pulses are not rectangular but of the form

$$r(t) = \theta P t e^{-\beta t} \quad t \geq 0 \quad 3.53$$

Here again θ is a binary random variable taking the values ± 1 with equal probability of $\frac{1}{2}$; and the θ 's are independent from pulse to pulse. P is an amplitude factor; introduced by the channel; that is unknown but fixed in this example. It will be assumed that the pulses are transmitted at the rate of one each T seconds where T is large enough that intersymbol interference may be ignored. The parameter β determines the exact pulse shape that is received. β may be unknown; or time varying; or both. For now it is assumed to be fixed but unknown. The channel noise is assumed to be stationary additive white Gaussian with zero mean and variance N_0 . (Most of these assumptions are not necessary; but they simplify the analysis.)

The receive filter is the single pole low pass filter of Eqn. 3.4 used in the example of Section 3.2. The total receiver is the same as shown in

Figure 3.4. The noise sample at the filter output is the same as given in Eqn. 3.7. The sampled signal portion of the filter output is now given by

$$\begin{aligned}
 x(T_s) &= \theta P \int_0^{T_s} e^{-\beta\tau} \sqrt{\alpha} e^{-\alpha(T_s-\tau)} d\tau \\
 &= \frac{\theta P \sqrt{\alpha}}{(\beta-\alpha)^2} (e^{-\alpha T_s} - e^{-\beta T_s}) - \frac{\theta P \sqrt{\alpha}}{(\beta-\alpha)} T_s e^{-\beta T_s} \quad 3.54
 \end{aligned}$$

for $\beta \neq \alpha$

As in the example of Section 3.2, the receiver will adaptively adjust the filter gain A ; the sampling time T_s ; and the receiver bandwidth α to obtain optimum performance. The analytical work of Section 3.2 can be applied directly because the results are not dependent on the form of the received pulse.

3.3.1 Adaptive Adjustment of the Receiver Parameters

The filter gain, A , will be adjusted to minimize the mean square error as in Eqn. 3.8. Proof of convergence was given in Section 3.2.1 and the steady state MSE could be obtained by substituting into Eqn. 3.19(b).

The filter sampling time, T_s , is adjusted to

maximize the SNR at the sample time. All the equations of Section 3.2.2 apply here as well.

The receiver filter bandwidth α will again be adjusted to maximize the SNR at the sampling time as described in Section 3.2.3. All the previous results are valid for this example.

3.3.2 Discussion of Computer Simulation Results

The same simulations were carried out using the second order receive pulse as for the rectangular received pulse. The results were essentially identical indicating that the gradient algorithms were capable of satisfactorily adapting the receive filter parameters to their optimum values. The detailed results are not presented here because they do not add any further general knowledge about the performance of gradient algorithms. This example was also found to track satisfactorily with the same algorithm parameters that were picked for the time invariant tests and again it showed about a 2 dB improvement in $\overline{\text{SNRD}}$ over the fixed system.

The transmitted pulse rate was then increased until considerable pulse overlap or intersymbol interference occurred at the receiver filter output.

The same adaptive algorithm continued to perform satisfactorily and gave about the same tracking performance improvements if more samples were averaged to obtain the gradient estimates used in the algorithms. This is expected since; although the intersymbol interference effects tend to produce erroneous gradient estimates; the intersymbol interference has a zero mean value and therefore tends to average out when the higher order algorithms (order 10 to 20) are used.

For the case of intersymbol interference; a simple threshold detector following the receive filter is no longer an optimum detector. Hence for purposes of the simulation; errorless decision feedback to the gradient algorithms was simulated by using the sign of the actual transmitted pulses in the adaptive algorithms. Figure 3.7 illustrates the receiver configuration used in the simulation.

Because of this initial success in adapting the filter in the presence of both noise and intersymbol interference; it was decided to examine the performance of these high data rate systems; when a transversal equalizer is used in conjunction with the analog filter to remove the interference effects

Figure 3.7 Simple Binary PAM Receiver Simulation

from the decisions. In addition, it is desirable to attempt some optimization of the algorithms for this purpose. This important area of PAM communications will be examined in considerable detail in Chapter 5 which deals with intersymbol interference problems; and a simulation of an adaptive PAM receiver in which both the analog prefilter and the transversal equalizer are made adaptive using gradient algorithms.

3.4 Summary

This chapter introduced a basic PAM communication system and considered the adaptive reception of isolated pulses. The choice of performance criteria was discussed and practical methods of estimating the gradients were given. Section 3.2 considered the specific example of rectangular received pulses and a first order low pass filter. The adaptive adjustment of filter gain, filter bandwidth and sampling time were analyzed and convergence was shown for the practical gradient estimating procedures.

Computer simulations using the analytic techniques were carried out and the results reported

and discussed in Section 3.2.4. These results included tests of tracking performance and indicated that considerable performance improvements can be made using adaptive receivers. Section 3.3 discussed some computer work when second order pulses were received by the low pass filter. Of particular interest here were the good tracking performance results obtained under conditions of intersymbol interference.

This chapter lays the groundwork for Chapter 5 which considers the adaptive performance of PAM systems in the presence of severe intersymbol interference.

CHAPTER 4

TRACKING PERFORMANCE FOR TIME VARYING SYSTEMS

In Section 2.3 some expressions were derived for the performance of the simple gradient technique when the optimum parameter value is time varying. The optimum fixed step size could be obtained from Eqn. 2.55 by minimizing the mean square parameter error with respect to this step size. However this is difficult to do because of the complexity of the expression and the fact that in practice the form of the time variations is usually unknown. In Section 4.1 some simplifications are applied to Eqn. 2.55 to enable calculation of the optimum fixed step sizes to use for stochastic optimum parameter variations. Section 4.2 considers the adaptive process as a first order system and computes the tracking errors for sinusoidal variation of the optimum parameter. Section 4.3 combines this result with results from Chapter 2 to compute the best fixed tracking step size for this type of parameter variation. Subsequent sections compare these results with computer simulations to check the validity of the simple models.

4.1 Tracking Stochastic Variations

From Eqn. 2.55 we have

$$\sigma_{p_i}^2 = \frac{\alpha_i^2 \sigma_{z_i}^2}{1 - H_i^2} + \frac{\sigma_{d_i}^2}{1 - H_i^2} + 2\sigma_{d_i}^2 \sum_{j=1}^{m-2} \rho_{j,i} (\overline{H_i})^j \sum_{n=0}^{m-j-1} (\overline{H_i}^2)^n$$

where $\sigma_{p_i}^2$ is the i^{th} parameter MSE

α_i is the fixed step size parameter for the i^{th} parameter adaptive loop

$\sigma_{z_i}^2$ is the variance of the additive noise in the i^{th} gradient estimate

$\sigma_{d_i}^2$ is the variance of the optimum parameter process p_i^*

$\rho_{j,i}$ are correlation coefficients and H_i is defined in Eqn. 2.46.

As an extreme case suppose that p_i^* is a white process.

Then, as shown in Appendix D, $\rho_{1,i} = -\frac{1}{2}$, $\rho_{j,i} = 0$ for $j = 2, 3 \dots m-2$ and 2.55 simplifies to

$$\sigma_{p_i}^2 = \frac{\alpha_i^2 \sigma_{z_i}^2}{1 - H_i^2} + \frac{\sigma_{d_i}^2}{1 - H_i^2} (-\alpha_i Y) \quad 4.1(a)$$

$$= \frac{\sigma_{d_i}^2(Y) - \alpha_i \sigma_{z_i}^2}{2Y + \alpha_i Z} \quad 4.1(b)$$

$$\text{where } Y = \frac{\partial^2 V(\bar{p})}{\partial^2 p_i} \quad 4.2(a)$$

$$\text{and } Z = \left[\frac{\overline{\partial^2 v(\bar{p})}}{\partial p_i^2} \right]^2 \quad 4.2(b)$$

The mean square parameter error $\sigma_{p_i}^2$, has no minimum for a finite value of the step size parameter, α_i . It might appear that $\sigma_{p_i}^2$ can be made zero if

$$\alpha_i = \frac{\sigma_{d_i}^2}{\sigma_{z_i}^2} \quad \text{Y} = \frac{\sigma_{d_i}^2}{\sigma_{z_i}^2} \frac{\overline{\partial^2 v(\bar{p})}}{\partial p_i^2} \quad 4.3$$

but in Section 2.4 it was shown that for stability

$$|\bar{H}| < 1 \text{ or}$$

$$\left| 1 + \alpha_i \frac{\overline{\partial^2 v(\bar{p})}}{\partial p_i^2} \right| < 1$$

$$\therefore \alpha_i \frac{\overline{\partial^2 v(\bar{p})}}{\partial p_i^2} < 0 \quad 4.4$$

Thus α_i and $\frac{\overline{\partial^2 v(\bar{p})}}{\partial p_i^2}$ must be of opposite sign to ensure stability; and this condition is not satisfied by Eqn. 4.3.

If Eqn. 4.1 is compared with Eqn. 2.34 it can be seen that the effect of the optimum parameter variations is to add a second noise term to the mean square parameter error. Here as in Section 2.2 the error can be minimized only if α_i decreases with time in a special way

such as in the Robbins-Monro algorithm⁽¹⁸⁾. (See Section 2.1.1.) This makes the term $\frac{\alpha_i^2 \sigma_{z_i}^2}{1 - H_i^2}$ go to

zero but the effects of the stochastic parameter changes cannot be eliminated. In addition if α_i tends to zero, the algorithm will have no tracking capability.

Thus in the extreme case when the optimum parameter is a white process its effect on the adaptive procedure is the same as that of the noise process z_i which was an additive noise disturbance to the gradient estimate. This is expected since it is impossible to "track" a white process. A value of α_i would have to be picked for a tradeoff in performance between the initial transient response and the steady state error.

If the optimum process is correlated from step to step as we expect in practice then the optimum step size would be expected to be finite in order to follow changes and still keep the effects of additive noise small. To illustrate this trend suppose the optimum process is correlated only over one step, that is $\rho_{1,i} \neq 0$, $\rho_{2,i} \neq 0$, $\rho_{j,i} = 0$ for all $j \neq 1, 2$.

$$\begin{aligned} \therefore \sigma_{p_i}^2 = & \frac{\alpha_i^2 \sigma_{z_i}^2}{1 - \overline{H_i^2}} + \frac{\sigma_{d_i}^2}{1 - \overline{H_i^2}} + 2\sigma_{d_i}^2 \left[\rho_{1,i} \overline{H_i} \sum_{n=0}^{m-2} (\overline{H_i^2})^n \right. \\ & \left. + \rho_{2,i} (\overline{H_i})^2 \sum_{n=0}^{m-3} (\overline{H_i^2})^n \right] \end{aligned} \quad 4.5$$

$$\therefore \sigma_{p_i}^2 = \frac{\alpha_i^2 \sigma_{z_i}^2}{1 - \overline{H_i^2}} + \frac{\sigma_{d_i}^2}{1 - \overline{H_i^2}} \left[1 + 2\rho_{1,i} \overline{H_i} + 2\rho_{2,i} (\overline{H_i})^2 \right] \quad 4.6$$

since $m-2$ and $m-3$ are still large as $m \rightarrow \infty$.

It is shown in Appendix E that this can be written as

$$\sigma_{p_i}^2 \approx \frac{\alpha_i^2 \sigma_{z_i}^2 + \sigma_{d_i}^2 (A + \alpha_i B)}{-2\alpha_i \frac{\partial^2 V}{\partial p_i^2}} \quad 4.7$$

where A and B depend, in general, on the values of the $\rho_{j,i}$. The effect of greater correlation is to change the values of A and B and so change the effect of $\sigma_{d_i}^2$ only. This is certainly true if only the first few $\rho_{j,i}$ are non zero. Hence the optimum fixed step size is given in Appendix E as

$$\alpha_i^* \approx \frac{\sigma_{d_i}}{\sigma_{z_i}} \sqrt{A} \quad 4.8$$

As the optimum process becomes more correlated

or band limited the step size must be increased to better track the time variations. Usually in practice the optimum parameter process p_i^* will be highly correlated, or very narrow band (slowly time varying).

Eqn. 4.8 can be used to estimate the step size if something is known about the optimum process.

In the next section a simple model for the adaptive system is proposed and the tracking effects are investigated for sinusoidal variations in optima.

4.2 A First Order Adaptive Model

In this section a first order model for the adaptive loop is proposed as in Figure 4.1. The filter $G(s)$ will be considered a single pole low pass filter so that

$$G(s) = \frac{a}{s+a} \quad . \quad 4.9$$

The motivation for this comes from Eqn. 2.23 where the effect of the adaptive system on the mean square parameter error is seen to be exponential. The time constant for the model is obtained from

$$e^{-a} = \left. 1 + \alpha_i \frac{\overline{\delta^2 v(\bar{p})}}{\partial p_i^2} \right|_{p_i = p_i^*}$$

Figure 4.1 A First Order Adaptive Model

$$\therefore a = - \ln \left(1 + \alpha_i \frac{\overline{\partial^2 V}}{\partial p_i^2} \right) \quad 4.10$$

where the simpler notation has been used for the partial derivative;

$$\text{or } a \approx - \alpha_i \frac{\overline{\partial^2 V}}{\partial p_i^2} \quad 4.11$$

$$\text{when } \alpha_i \frac{\overline{\partial^2 V}}{\partial p_i^2} \ll 1$$

Here $1/a$ is the time constant of the model; or a is the 3 dB bandwidth.

This model has unity gain for a d.c. signal and therefore its estimate $\hat{p}_i$ of p_i^* converges in the mean to p_i^* . This is for a noiseless case. When noise is present it will be assumed additive zero mean as in Section 2.2.

Now suppose p_i^* is a sine function and the adaptive system must track its variations. For low frequency variations the estimate will be a sine function of the same frequency; but of lagging phase and possibly different amplitude. That is, if

$$p_i^* = P \sin \omega t \quad 4.12$$

$$\text{then } \hat{p}_i = A \sin (\omega t + \phi) \quad 4.13$$

$$\text{The mean square error } \overline{e^2} = \overline{(p_i^* - \hat{p}_i)^2} \quad 4.14$$

$$\begin{aligned} \therefore \overline{e^2} &= \overline{(P \sin \omega t - A \sin (\omega t + \phi))^2} \\ &= \frac{1}{T} \int_0^T (P \sin \omega t - A \sin (\omega t + \phi))^2 dt \\ &\quad \text{where } T = \frac{2\pi}{\omega} \end{aligned} \quad 4.15$$

$$\text{Now } A = P |G(\omega)| = \frac{Pa}{\sqrt{a^2 + \omega^2}} \quad \text{and} \quad \phi = \tan^{-1} \left(\frac{-\omega}{a} \right)$$

Performing the calculations of Eqn. 4.15 results in

$$\overline{e^2} = \frac{P^2 \omega^2}{2(a^2 + \omega^2)} \quad 4.16$$

where P is the peak amplitude of p_i^* ; ω is the frequency of p_i^* and a is the 3 dB filter bandwidth.

Eqn. 4.16 gives the mean square parameter error for a simple adaptive system tracking sinusoidal variations in the optimum parameter when no noise is present. This model should give reasonable results provided the frequency of variation is not too small with respect to the system "bandwidth". If the optimum changes very slowly then most of the parameter error will arise from variations of the estimate about the optimum point due to small amounts of noise

and due to a step size that is larger than required. Thus the step size should be reduced to reduce the effects of noise (Eqn. 2.35) and then the tracking formula would again give reasonable results. In Section 4.3 the optimum fixed step size is calculated for noise and time variations and it is shown by simulation work in subsequent sections that Eqn. 4.16 gives reasonable results.

Eqn. 4.16 indicates that if there were no noise then α should be large. This implies that the step size should be approximately

$$\alpha_i \approx \frac{-1}{\frac{\partial^2 V}{\partial p_i^2}}$$

which was also the result of Section 2.2, Eqn. 2.26 for rapid transient response.

If the time variations can be decomposed into a Fourier series of sinusoids then the mean square tracking error would be given by a sum of terms like Eqn. 4.16.

4.3 Optimum Step Size for Sinusoidal Variations in Optimum Parameters

If the model of Section 4.2 is used for tracking errors, and the expression for MSE due to additive noise of Section 2.2 is combined with it, we can estimate the total mean square parameter error as

$$\sigma_{p_i}^2 = \frac{p^2 \omega^2}{2(\omega^2 + a^2)} + \frac{\alpha_i^2 \sigma_{z_i}^2}{1 - \overline{H_i^2}} \quad 4.17$$

$$\text{or } \sigma_{p_i}^2 = \frac{p^2 \omega^2}{2(\omega^2 + a^2)} + \frac{\alpha_i^2 \sigma_{z_i}^2}{1 - \overline{(1 + \alpha_i \frac{\partial^2 V}{\partial p_i^2})^2}} \quad 4.18$$

$$\text{where } a = - \ln \left(1 + \alpha_i \frac{\partial^2 V}{\partial p_i^2} \right)$$

The optimum fixed step size α_i^* is found by setting the derivative of Eqn. 4.18, with respect to α_i , equal to zero.

Because of the difficulty posed in taking the derivative of a^2 with respect to α_i , an approximation will be introduced for analytical simplicity.

For good steady state performance

$$\alpha_i \frac{\partial^2 V}{\partial p_i^2} \ll 1 \quad 4.19$$

$$\text{Therefore } a \approx -\alpha_i \frac{\overline{\partial^2 V}}{\partial p_i^2} \quad 4.20$$

Substituting Eqn. 4.20 into 4.18 yields

$$\sigma_{p_i}^2 \doteq \frac{p^2 \omega^2}{2 \left[\omega^2 + \alpha_i^2 \left[\frac{\overline{\partial^2 V}}{\partial p_i^2} \right]^2 \right]} + \frac{\alpha_i \sigma_{z_i}^2}{-2 \frac{\overline{\partial^2 V}}{p_i^2} - \alpha_i \left[\frac{\overline{\partial^2 V}}{\partial p_i^2} \right]^2} \quad 4.21$$

Setting the derivative of Eqn. 4.21, with respect to α_i , equal to zero gives

$$0 \doteq \frac{-p^2 \omega^2 2\alpha_i \left[\frac{\overline{\partial^2 V}}{\partial p_i^2} \right]^2}{2 \left[\omega^2 + \alpha_i^2 \left[\frac{\overline{\partial^2 V}}{\partial p_i^2} \right]^2 \right]^2} + \frac{\sigma_{z_i}^2 \left[-2 \frac{\overline{\partial^2 V}}{\partial p_i^2} - \alpha_i \left[\frac{\overline{\partial^2 V}}{\partial p_i^2} \right]^2 \right] - \alpha_i \sigma_{z_i}^2 \left[-\left[\frac{\overline{\partial^2 V}}{\partial p_i^2} \right]^2 \right]}{\left[2 \frac{\overline{\partial^2 V}}{\partial p_i^2} + \alpha_i \left[\frac{\overline{\partial^2 V}}{\partial p_i^2} \right]^2 \right]^2} \quad 4.22$$

$$\therefore \frac{p^2 \omega^2 \alpha_i \frac{\overline{\partial^2 V}}{\partial p_i^2}}{2 \left[\omega^2 + \alpha_i^2 \left[\frac{\overline{\partial^2 V}}{\partial p_i^2} \right]^2 \right]^2} \doteq \frac{-\sigma_{z_i}^2}{\left[2 \frac{\overline{\partial^2 V}}{\partial p_i^2} + \alpha_i \left[\frac{\overline{\partial^2 V}}{\partial p_i^2} \right]^2 \right]^2} \quad 4.23$$

When all the terms of Eqn. 4.23 are known, α_i^* can be evaluated. However, in order to continue

analytically some further approximations will be made.

From Eqn. 4.19 we have

$$\alpha_i^2 \left[\frac{\overline{\partial^2 V}}{\partial p_i^2} \right]^2 \ll \alpha_i \frac{\overline{\partial^2 V}}{\partial p_i^2} \quad 4.24$$

In addition, the tracking model is reasonable only if $\omega \ll$ "bandwidth" of the adaptive loop; that is if

$$\omega \ll a^2 = \alpha_i^2 \left[\frac{\overline{\partial^2 V}}{\partial p_i^2} \right]^2 \quad 4.25$$

This approximation will usually hold in practice.

Using approximations 4.24 and 4.25 in Eqn. 4.23 gives

$$\frac{2P^2 \omega^2}{2\alpha_i^2 \left[\frac{\overline{\partial^2 V}}{\partial p_i^2} \right]^2} \approx - \frac{\alpha_i \sigma_{z_i}^2}{2 \frac{\overline{\partial^2 V}}{\partial p_i^2}} \quad 4.26$$

From this the optimum fixed α_i for sinusoidal tracking

$$\text{is} \quad \alpha_i^* \approx - \sqrt[3]{\frac{2P^2 \omega^2}{\sigma_{z_i}^2 \frac{\overline{\partial^2 V}}{\partial p_i^2}}} \quad 4.27$$

Note that the greater the noise power, the smaller the fixed step size. The greater the frequency or amplitude of the time variations, the larger the step size. Thus Eqn. 4.27 yields intuitively satisfying results.

The partial derivative $\frac{\partial^2 V}{\partial p_i^2}$ may be approximated as shown in Appendix F if necessary to aid in getting an approximate value for α_i^* . Thus the adaptive system could make an estimate of the proper step size by using Eqn. 4.27 as illustrated in Appendix F. However this would require estimates of P and ω and the result would really be optimum only for a sinusoidal variation in the optimum.

Some work on tracking problems has been done by Widrow⁽²⁵⁾ who states that the optimum step size results in the steady state error due to noise being approximately equal to twice the error due to tracking lag. These results are based on several approximations.

If the approximations of Eqns. 4.24 and 4.25 are applied to Eqn. 4.21 we have

$$\sigma_{p_i}^2 \approx \frac{P^2 \omega^2}{2\alpha_i^2 \left[\frac{\partial^2 V}{\partial p_i^2} \right]^2} + \frac{\alpha_i \sigma_{z_i}^2}{-2 \frac{\partial^2 V}{\partial p_i^2}} \quad 4.28$$

Comparing Eqn. 4.28 to Eqn. 4.26 shows that when α_i is selected according to Eqn. 4.27 then the steady state noise error component is equal to twice the tracking error component of the mean square parameter error. The results of this section are therefore in agreement with those of Widrow.

These results are for a single parameter variation. When there are several parameters adapting the problem is more difficult. If the parameters are independent then each will proceed as outlined above. Although they will have different step sizes, the net MSE will be minimized if the step sizes are chosen properly. When the parameters interact the problem is very complex. Considering the model of Figure 4.1 the components of this "filter" now become time varying functions of the other interacting parameters. Such a system is virtually impossible to analyse.

There remains some difficulty in estimating the parameters of Eqn. 4.27. Some discussion of possible adaptive estimation methods will be given in Section 4.5.

4.4 Comparison with Experimental Results

In this section the analytical results of Sections 4.1 to 4.3 will be compared with the results of some simple computer simulations. The specific example used in the simulations is introduced in the next section.

4.4.1 Rectangular Pulses and a Low Pass Filter

This is the example described in Section 3.2. Here the formulae will be tested using the gain A as the adaptive parameter.

Eqn. 3.15(b) gives the recursive relationship for the parameter error as

$$e_A^{k+1} = e_A^k (1 + 2\epsilon(x^k + y^k)^2) + 2\epsilon[A^*(x^k + y^k)^2 - \hat{\theta}^k(x^k + y^k)]$$

where all quantities were defined in Section 3.2.

The adaptive system "bandwidth" here is

$$a_A = -\ln(1 + 2\epsilon(\overline{x^2} + \overline{y^2}))$$

A simulation was run in which the received pulse parameters were made to vary in such a way that the optimum value of the gain parameter, A , was a sinusoidal function of time. Table 4.1 is a comparison of the calculated and measured tracking

<u>Frequency (ω)</u>	<u>Step Size</u>	<u>Calculated Tracking Error</u>	<u>Measured</u>
$1.26 \times 10^{-2} \text{ r/sec.}$	-0.5	1.77×10^{-4}	4.08×10^{-4}
	-0.25	.000934	.00160
	-0.1	.00658	.00911
	-0.05	.0258	.0292
	-0.01	.216	.204
$2.09 \times 10^{-2} \text{ r/sec.}$	-0.5	4.92×10^{-4}	11.1×10^{-4}
	-0.25	.00258	.00429
	-0.1	.0176	.0222
	-0.05	.0622	.0638
	-0.01	.266	.306

Table 4.1 COMPARISON OF CALCULATED AND MEASURED TRACKING ERRORS

errors. The additive noise was zero during this simulation. The frequency, ω , is calculated assuming a pulse transmission rate of one pulse per second.

The results are fairly good for step sizes $|\epsilon| < 0.1$. For larger step sizes the error is not purely tracking error but partly due to the estimate wandering about the optimum as well. For smaller step sizes the estimate always lags the optimum value as discussed in Section 4.2. The maximum tracking error, $P^2/2$, occurs when $\alpha \ll \omega$. This value here is 0.307 and it is observed that as $|\epsilon|$ gets smaller (and since a_A gets smaller) the tracking error approaches this value.

Another source of error is that a_A is not really constant but changes slightly as the received signal varies, and for large ϵ the tracking error is quite sensitive to the value of a_A . However the operating step size for this example was found from simulation results to be around $\epsilon = -0.1$.

A second simulation, which included the effects of additive receiver noise, was run to compare calculated and experimental values of ϵ^* the optimum tracking step size parameter. Approximate calculated values were obtained by substituting into Eqn. 4.27. The

results are as follows:

	<u>Calculated ϵ^*</u>	<u>Experimental ϵ^*</u>
For $\omega = 2.09 \times 10^{-2}$	.19	.22
For $\omega = 1.26 \times 10^{-2}$	.14	.17

These results are quite reasonable considering the approximations involved. The actual MSE is fairly insensitive to the value of ϵ near ϵ^* as indicated in Table 4.2

4.5 Adaptive Estimation of the Step Size Parameters

The work so far in this thesis has shown that adaptive systems using gradient algorithms can yield significant performance improvements even with fixed step size parameters in the algorithms. It has been shown that optimum values for these parameters should be selected. It may be difficult to estimate the proper parameter values even with the aid of Eqn. 2.55 or the work of Sections 4.1 and 4.3. This section briefly considers some methods of estimating the optimum fixed step size parameter values by means of additional adaptive circuitry.

<u>Frequency (ω)</u>	<u>ϵ</u>	<u>MSE</u>
2.094×10^{-2} r/sec.	-.50	.0209642
	-.25	.0105355
	-.23	.0103297
	-.22	.0102918
	-.21	.0103061
	-.20	.0103811
	-.17	.0110910
	-.10	.0193424
1.256×10^{-2} r/sec.	-.50	.0220812
	-.25	.0113980
	-.20	.0105224
	-.18	.0103764
	-.17	.0103679
	-.16	.0104138
	-.15	.0105263
	-.14	.0107209
	-.10	.0129036

Table 4.2 MSE AS A FUNCTION OF STEP SIZE PARAMETER

4.5.1 Higher Orders of Adaption

For simple adaptive systems the derivatives of the performance criterion, V , with respect to the step size parameter, α , could be estimated as shown in Figure 4.2. In this system another adaptive loop, which adjusts the step size parameters used in the main adaptive system, has been added around the main system. As shown in Figure 4.2 this requires duplication of the main adaptive system for each step size parameter used. A more practical method of operation as shown in Figure 4.3 would probably be used. The parameter α would be set to a value α_1 for some period of time and the performance estimated by some suitable criterion. Then α would be changed to a value α_2 and the performance again measured. The gradient would be estimated from $V(\alpha_1) - V(\alpha_2)$ and a decision made on whether or not the performance was improved. Only one main adaptive system is now required.

When several step size parameters are to be adjusted using this technique, only one can be perturbed at a time. The estimation procedure would optimize each step size in turn and then start over continuously repeating the sequence. If the initial step size parameter estimates were within the stability

Figure 4.2 Parallel Adaption of Step Size Parameters

Figure 4.3 Sequential Adaptation of Step Size Parameters

bounds then the system should slowly converge to nearly optimum parameter values.

4.5.2 Separation of "Steady State" and "Tracking" Errors

Another method of selecting a step size parameter α , that does not require an estimate of $\frac{\partial V}{\partial \alpha}$, is suggested by the results of Section 4.3. If a mean square error criterion is used as a performance measure then the results of Eqn. 4.28 indicate that for the example considered the total MSE is minimized when α is selected so that the "steady state noise" error component is equal to twice the "tracking" error component. Similar results were obtained by Widrow⁽²⁵⁾ for other examples.

The steady state noise is a high frequency component caused by small quantized changes in the step size parameters and their "hunting" about the optimum parameter value. The tracking error is a low frequency slowly varying error component caused by the delay of the adaptive system in responding to input changes. If these errors are separately filtered and compared, the above criterion regarding their relative values could be used to adapt the step size parameter fairly close to its optimum value. However, it is often very difficult to physically measure these errors.

4.5.3 Multi-Mode Adaption

A multi-mode adaptive system is one which has two or more methods of selecting the step size parameter depending on whether the system is in a transient (or acquisition) mode, a steady state mode, or a time varying (tracking) mode.

During the initial acquisition phase the step size parameters should be selected to give a rapid transient response. The optimum step size parameters can be evaluated using Eqn. 2.26 if estimates for the second derivatives, $\frac{\partial^2 V(\bar{p})}{\partial p_i^2}$, are available.¹ These estimates may be made by using measurements of the system performance for various parameter values p_i in the equations of Appendix F, or by some similar method. During this phase of the adaption it is only necessary to guarantee stability of the algorithm. The mean squared error effects need not be considered.

When the transients of acquisition have disappeared (as indicated perhaps by a very slow change in the step size parameters) the adaptive process can be switched to the second mode. For time invariant problems this mode would involve a steady state approach to the asymptotic mean squared error.

¹

This is essentially the Newton-Raphson method. (57)

A small fixed step size parameter could be selected, or some type of decreasing step size parameter algorithm could be used. For time varying situations a tracking mode would follow the acquisition mode. Again a fixed step size parameter value could be used, having been computed in advance by some means such as those illustrated in Sections 4.1 to 4.3.

A multi-mode adaptive procedure similar to this is suggested by Wilde⁽¹⁴⁾ who considers a three stage process. A large step size parameter value is used initially. Then the step size parameter is monotonically decreased at a moderate rate until the optimum value is approached. The system then switches to the third mode in which the value is slowly reduced, eventually approaching zero. This third mode is obviously suitable only for time invariant systems that approach asymptotic steady state values.

4.5.4 Other Methods of Adjusting Step Size Parameters

There are many optimum seeking procedures which compute the optimum step for each iteration. One common type of procedure having this property is the conjugate gradient algorithm, such as the Fletcher-Powell method⁽²²⁾ or the Fletcher-Reeves method⁽²⁰⁾.

These algorithms compute the optimum magnitude and direction of each iteration. The problem with these methods is that they usually require a great deal of computation to get the required estimates of step size. This computational complexity cannot be considered for most of the simple systems considered here which require "on-line" techniques. Another problem occurs when the gradients used in the computations must be evaluated experimentally. The resultant inaccuracies may remove the advantage to be gained by this more complex method, or they may introduce problems when experimentally derived matrices need to be inverted. These methods do have considerable application in analytical work, but appear to be unsuitable for the type of simple on line adaptive systems considered in this thesis.

4.6 Summary

This chapter has examined the use of adaptive systems using gradient algorithms for tracking time varying phenomena based on the results of the time varying analysis of Section 2.3. Section 4.1 derived the optimum fixed step size parameter when the time varying parameter is a white noise process. Section

4.2 considered a simple low pass filter model for the adaptive algorithm. This model was used in Section 4.3 to derive the optimum fixed step size parameter to use when the adaptive system's adjustable parameters have optimum settings that are sinusoidal functions of time. Section 4.4 compared the analytical results of Section 4.3 with some computer simulation results to verify the model and assumptions used. Section 4.5 considered some possible methods of adaptively estimating the step size parameters in an adaptive system. These methods could be used to obtain performance improvement over that of the fixed step size parameter algorithms; or just to get a reasonable estimate of the required step size parameter value to be used in a fixed step size algorithm.

CHAPTER 5

APPLICATION OF ADAPTIVE TECHNIQUES TO PAM COMMUNICATIONS WITH INTERSYMBOL INTERFERENCE

The subject of PAM communications and the use of gradient algorithms to adaptively optimize receive filters was introduced in Chapter 3. In Section 3.3 preliminary results of some computer simulation work were discussed which indicated that the adaptive receive filters would also perform well when the received signal suffered from considerable intersymbol interference. That simulation work did not use a proper transversal equalizer type of receiver.

In this chapter the problem of intersymbol interference and the use of optimum receivers to combat it are introduced more fully. Methods of adapting the analog prefilter, as well as the digital equalizer portion of these receivers, are derived analytically. The results of a rather complex hybrid computer simulation of an adaptive PAM system are also reported.

5.1 Intersymbol Interference in PAM Communication Systems

The overlap of adjacent information symbols at the receiver of a PAM system is caused by the use of a bandlimited channel for high data rate transmission. The optimum receiver must therefore counteract the effects of both noise and intersymbol interference. In recent years considerable effort has been made to analyse and improve receivers for these types of digital PAM systems.^(9-12,30-33,45-49) Of particular interest is the use of adaptive equalizers in many of these receivers to adjust the digital portion of the receiver when the intersymbol interference is not accurately known.

The optimum linear receiver was first derived by George⁽⁵⁰⁾ and by Tufts⁽⁵¹⁾. The use of a receiver having a decision feedback or nonlinear equalizer was first proposed by Austin⁽⁵²⁾. Both receivers consist of a filter matched to the received pulse followed by a sampled data filter or transversal equalizer. Since a matched filter gives optimum SNR at the sample time (for white Gaussian noise at the input) the analog prefilters of these receivers can be made to adaptively track changes in the received

signal characteristics by using maximum SNR at the sample time as a performance criterion. The adaptive techniques described in Section 3.3 can be used with very little change. (The sign of the received pulses would now be available at the output of the equalizer; this was simulated in the work of Section 3.3 by using the sign of the transmitted pulses since an equalizer was not used in that simulation.)

This method of adaption assumes that the analog filter's parameters can always be set so that it stays matched as the received pulse shape varies. An adaptive system using this performance criterion is therefore easy to implement and may give reasonable performance improvement by optimizing the analog as well as the digital part of the receiver. As long as changes do not take place too rapidly, the tap gains of the equalizer will always re-adapt themselves to a new optimum set of values as dictated by the receiver pulse changes and the new analog filter parameters.

In the literature the analog prefilter is usually assumed to be a matched filter. However, since matched filters are usually difficult to build,

and especially because the channel characteristics (and therefore the received pulse characteristics) are usually not known in advance, the analog pre-filters used in practice are normally approximations to a matched filter, or simple nonmatched filters such as a first or second order low pass filter. Since a mismatched filter is not an optimum SNR filter, it is no longer an optimum scheme to set it for maximum SNR even though the idea seems reasonable. The filter must be set so that there is minimum distortion produced by the combined effects of noise and intersymbol interference when the filter operates in conjunction with the transversal equalizer. It is indeed fortuitous that the matched filter achieves this effect at the point of optimum SNR. It might turn out however that tracking nonmatched filters for maximum SNR at the filter output is a good suboptimum criterion to use.

Another reason for using an adaptive prefilter may occur when the equalizer is not ideal. This means that it is of finite length and that its tap and sample values have finite quantization. When the SNR is high, the noise on the analog filter output samples will be insignificant with respect to the equalizer quantization error and truncation error

(due to a finite number of taps). In this case the equalizer is intersymbol interference limited and the prefilter should be adjusted to reduce the interference (by widening its bandwidth for example) at the expense of more noise on the samples. This could be done by using a MSE criterion to minimize the combined effects of intersymbol interference and noise at the equalizer output. A filter which must remain matched is insensitive to SNR and is really optimum only if the equalizer is ideal.

This should be even more important if the equalizer is of the decision feedback type. At high SNR this equalizer can efficiently cancel the effects of previous pulses. The prefilter should be adjusted to minimize the effects of following pulses at the expense of more noise (which is small anyway at high SNR) and more previous pulse intersymbol interference (which can be removed by the feedback equalizer). This may require extra feedback taps but these are very simple to add. Clearly the required asymmetric prefilter response cannot be achieved by a matched filter. It would therefore appear that an adaptive prefilter could set itself to give a performance improvement over other fixed prefilters under certain conditions. Some work

along these lines for time invariant systems has very recently been published by Mark and Budihardjo^(55,56) but the tracking performance using the adaptive pre-filters was not investigated.

The next section develops an adaptive scheme for a simple PAM system.

5.2 Analog Filter Adaption with Intersymbol Interference

This section develops an adaptive scheme for adjusting the bandwidth of the receiver's analog prefilter using a MSE performance criterion.

A simple binary PAM system is shown in Figure 5.1. The receiver consists of an analog prefilter, $h(t; \alpha)$, which in general is not matched to the received pulse, and a transversal equalizer which consists of a tapped delay line operating at the symbol transmission rate $1/T$ and a set of tap values $a(Q), \dots, a(1), a(0), a(-1), \dots, a(-P)$. The tap values are summed to produce an analog estimate of the received symbol, $\tilde{\theta}$. This estimate is then sliced to produce a digital estimate, $\hat{\theta}$, having value of ± 1 .

A message symbol $\theta(i)$ is transmitted each T seconds so the transmitted signal is

$$s(t) = \sum_{i=-\infty}^{\infty} \theta(k-i)p(t + iT - kT) \quad kT \leq t \leq (k+1)T \quad 5.1$$

Figure 5.1 A Binary PAM Communication System with Intersymbol Interference

where $\theta(i)$ is a binary random variable taking values of ± 1 with equal probability,

$$E\{\theta(i)\} = 0 \quad 5.2(a)$$

$$E\{\theta(i) \theta(k)\} = \delta(i-k) \quad 5.2(b)$$

and $p(t)$ is the transmitted pulse shape.

After passing through the noisy channel the input to the receiver is

$$x(t) = \sum_{i=-\infty}^{\infty} \theta(k-i)q(t+iT-kT) + n_1(t) \quad kT \leq t \leq (k+1)T \quad 5.3$$

where $q(\tau) = p(\tau) * c(\tau)$,¹ and $n_1(t)$ is zero mean white Gaussian noise of power spectral density N_0 watts /Hz.

The output of the analog receiver filter may now be written as

$$w(t) = \sum_{i=-\infty}^{\infty} \theta(k-i)z(t+iT-kT) + n(t) \quad kT \leq t \leq (k+1)T \quad 5.4$$

where $z(\tau) = q(\tau) * h(\tau, \alpha)$
and $n(\tau) = n_1(\tau) * h(\tau, \alpha)$.

The output of the analog filter is sampled each T seconds. Taking the sampling time as $t = kT$

1

The "*" symbol is used to denote convolution.

(delay between transmitter and receiver has been ignored) we have

$$w(k) = \sum_{i=-\infty}^{\infty} \theta(k-i)z(i) + n(k) \quad 5.5$$

where $n(k)$ is the sampled noise value at the input to the delay line.

If $z(\tau) \approx 0$ for τ outside the range $-LT \leq \tau \leq MT$ we may write

$$w(k) = \sum_{i=-M+k}^{L+k} \theta(i)z(k-i) + n(k) \quad 5.6$$

The linear transversal equalizer of Figure 5.1 forms the estimate of the transmitted symbol at time zero as

$$\tilde{\theta}(0) = \sum_{k=-P}^Q a(k)w(k) \quad 5.7$$

The analog error is $\tilde{\theta}(0) - \theta(0)$ where $\theta(0)$ is the transmitted symbol. The mean square error is therefore

$$E\{e^2\} = E\{(\tilde{\theta}(0) - \theta(0))^2\} \quad 5.8$$

The values of the tap gains $a(k)$ can be selected to minimize the MSE, and this minimization is often performed by an adaptive system using gradient algorithms.⁽⁹⁻¹²⁾ In the remainder of the development

work in this section, it is assumed that the transversal equalizer taps are kept at their optimum values by an adaptive system.

A method will now be investigated to adapt the analog filter parameters in order to minimize the MSE as well. For simplicity this example has considered only one parameter, α . However if several parameters are to be adapted, a system as derived below would be used for each parameter.

The optimum α is given by the solution of the equation

$$\frac{\partial}{\partial \alpha} E\{(\tilde{\theta}(0) - \theta(0))^2\} = 0 = 2E\{(\tilde{\theta}(0) - \theta(0))\left(\frac{\partial \tilde{\theta}(0)}{\partial \alpha}\right)\} \quad 5.9$$

This equation can be solved iteratively by an on-line adaptive system using a gradient algorithm of the form

$$\alpha^{k+N} = \alpha^k - \epsilon \frac{\partial E\{e^2\}}{\partial \alpha} \quad 5.10$$

where ϵ is a fixed step size parameter and N is some integer greater than 1 implying that α need not be updated each sample time because of the averaging required to evaluate $E\{e^2\}$. The algorithm of Eqn. 5.10 is of the form described in Section 3.2.3 which dealt with the reception of isolated pulses.

If the system is working reasonably well the digital estimate $\hat{\theta}(0)$ may be used as an estimate of $\theta(0)$ and the receiver can estimate the error, e , as shown in Figure 5.1. It therefore remains only to measure

$$\begin{aligned} \frac{\partial \tilde{\theta}(0)}{\partial \alpha} &= \sum_{k=-P}^Q \frac{\partial}{\partial \alpha} (a(k)w(k)) \\ &= \sum_{k=-P}^Q \left(\frac{\partial a(k)}{\partial \alpha} w(k) + a(k) \frac{\partial w(k)}{\partial \alpha} \right) \end{aligned} \quad 5.11$$

The tap gains, $a(k)$, are functions of the value of α , but they are probably fairly insensitive to small changes in α . It is therefore a reasonable approximation to assume that

$$\sum_{k=-P}^Q \frac{\partial a(k)}{\partial \alpha} w(k)$$

is small with respect to

$$\sum_{k=-P}^Q a(k) \frac{\partial w(k)}{\partial \alpha} .$$

$\frac{\partial \tilde{\theta}(0)}{\partial \alpha}$ may therefore be approximated as

$$\sum_{k=-P}^Q a(k) \frac{\partial w(k)}{\partial \alpha} .$$

Eqn. 5.10 may now be rewritten as

$$\alpha^k + N = \alpha^k - 2\epsilon E\{(\tilde{\theta}(0) - \hat{\theta}(0)) \left(\sum_{k=-P}^Q a(k) \frac{\partial w(k)}{\partial \alpha} \right)\} \quad 5.12$$

If now $\frac{\partial w(k)}{\partial \alpha}$ is approximated as

$$\frac{w(k, \alpha + \delta) - w(k, \alpha)}{\delta}$$

then an algorithm may be used that is similar to the one sided Kiefer-Wolfowitz type algorithm of Section 2.1.3. Convergence for this algorithm was shown in the example of Section 3.2.3.

An estimate of the derivative of the MSE may be implemented as shown in Figure 5.2. This approach would appear to require considerable additional circuitry especially if there are several analog filter parameters to be adapted. However, because the additional tapped delay line holds differences, its input amplitude range may be quite restricted and the delay line may need only a few levels. The scaling factor $1/\delta$ can be incorporated into the tap gains or the summer gain. The next section introduces an approximation that results in a very simple replacement for the additional adaptive circuitry.

5.3 Simplification of the Gradient Estimates

From the observation that the additional tapped delay line needs only a few amplitude levels,

Figure 5.2 Estimation of the Error Gradients for Analog Filter Parameters

one is led to consider what would happen if the auxiliary delay lines are given only two levels, that is they hold only the sign of the sampled difference. In this case

$$\frac{\partial \tilde{\theta}(0)}{\partial \alpha} \approx \sum_{k=-P}^Q a(k) \operatorname{sgn} \left[\frac{\partial w(k)}{\partial \alpha} \right]$$

If the sign of the error is used as well, a signed correlation can be performed instead of the full range analog correlation required in Eqn. 5.12. A gradient algorithm of this type was introduced in Sections 2.1.5 and 2.3.

Eqn. 5.12 can be rewritten as

$$\alpha^{k+N} = \alpha^k - 2\epsilon E \left\{ \operatorname{sgn} [\tilde{\theta}(0) - \hat{\theta}(0)] \sum_{k=-P}^Q a(k) \operatorname{sgn} \left[\frac{\partial w(k)}{\partial \alpha} \right] \right\} \quad 5.13$$

or, hard limiting the summer output as well

$$\alpha^{k+N} = \alpha^k - 2\epsilon E \left\{ \operatorname{sgn} [e] \operatorname{sgn} \left[\sum_{k=-P}^Q a(k) \operatorname{sgn} \left[\frac{\partial w(k)}{\partial \alpha} \right] \right] \right\} \quad 5.14$$

These approximations can be implemented with very simple hardware. The tapped delay line is replaced by a binary shift register and the multiplier at the output of the summer can be replaced by an "exclusive or" logic gate. The averager can be replaced with a summer. All of the adaptive circuitry for the analog filter is therefore simple to construct

including the extra analog filter if it is only a simple low pass filter. Actual adjustment of the tap gains could be implemented using the overflowing accumulator method described by Lucky⁽⁹⁾ (see also Section 2.1.5).

This adaptive system is slightly more complex than the maximum SNR criterion used in Section 3.3. However, it provides a good practical adaptive technique to use when non-matched analog prefilters are used in these receivers. It will at least give a good comparison of performance with tracking as opposed to using fixed filter parameters even if the approximations introduced cause the parameter settings to be slightly off their optimal values. In Section 5.5 the results are presented for a computer simulation using the adaptive techniques described in this section.

5.4 Adaption of the Sampling Time

This section briefly indicates how the receiver sampling time can be adaptively set to minimize the mean squared error at the output of the transversal equalizer.

An adaptive algorithm similar to Eqn. 5.10

will be used. In this case we must evaluate

$$\frac{\partial E\{e^2\}}{\partial T_s} = 2E\left\{e \left(\frac{\partial \tilde{\theta}(0)}{\partial T_s} \right)\right\} \quad 5.15$$

where

$$\frac{\partial \tilde{\theta}(0)}{\partial T_s} = \sum_{k=-P}^Q \left(\frac{\partial a(k)}{\partial T_s} w(k) + a(k) \frac{\partial w(k)}{\partial T_s} \right) \quad 5.16$$

Again the first term on the right of Eqn. 5.16 will be assumed small with respect to the second term, and will be ignored. Also the derivative $\frac{\partial w(k)}{\partial T_s}$ will be approximated by the Kiefer-Wolfowitz type algorithm of Section 2.1.3 as

$$\frac{\partial w(k)}{\partial T_s} \approx \frac{w(k; T_s + \delta/2) - w(k; T_s - \delta/2)}{\delta} \quad 5.17$$

where δ is a small positive quantity. Convergence for this algorithm was shown in Section 3.2.2. This requires that the analog filter be sampled $\delta/2$ seconds before, and $\delta/2$ seconds after the current sample time. The difference between the two values is then used as the required estimate of Eqn. 5.17. Note that only the main analog filter is required here. The same simplifications, using the binary shift register and only the signs of the error and the summer output, can also be applied to this algorithm.

5.5 Simulation of a PAM System with Intersymbol Interference

This section gives a brief description of the system simulated and a summary of some results and conclusions based on the work.

The example used is the second order received pulse and a receiver consisting of a single pole low pass analog prefilter and a transversal equalizer. The equalizer used in the simulation was a decision feedback equalizer, not a linear one as discussed in Sections 5.2 to 5.4. The equalizer used had 3 forward taps and 8 feedback taps. This equalizer was made adaptive in a manner similar to that used by George et al⁽¹¹⁾, and Lucky⁽⁹⁾. The analog prefilter was also made adaptive as described in Sections 5.3 and 5.4.¹ Both its bandwidth, α , and the sample time, T_s were adjusted to their optimum values. Figure 5.3 shows the salient features of the system simulated. Further detail of the hybrid simulation setup is given in Appendix G. Some of the more important results and conclusions are summarized in the next section.

I

It is quite straightforward to show that the gradient estimates used in Sections 5.3 and 5.4 are unchanged in form if a feedback equalizer is used because the derivative of the feedback tap value with respect to the analog filter parameter will also be a small term. (cf. Eqn. 5.11)

5.5.1 Summary of the Results of the PAM Simulation

One of the most important results of the simulation work is that the dual adaptive system was found to be quite stable and capable of seeking out a joint set of prefilter settings and equalizer tap values to produce optimum performance. The stability is demonstrated by the transient results shown in Figure 5.4; where the analog filter parameters were individually perturbed from their optimum settings and allowed to re-adapt to the optimum value. This indicates that there are definite optimum values for the sample time and bandwidth corresponding to any particular received pulse shape and set of equalizer tap values. When both parameters were perturbed simultaneously the system re-adapted to the optimum performance; but the set of prefilter and tap values were different and depended on the particular perturbation. More will be said later about this performance insensitivity to the parameter values.

For non-matched prefilters, the optimum sampling time is not easily determined analytically. If matched filters are used the optimum sampling time is at the peak of the isolated pulse response of the filter. The system was found to be quite capable of

Figure 5.4 Typical Responses to Step Perturbations from Optimum for (a) Bandwidth, and (b) Sample Time, of the Analog Prefilter

adaptively solving the problem of joint selection of sampling time and filter bandwidth. This is shown in Figure 5.5 where step changes were made to the received pulse bandwidth. Both the sampling time and the receive filter bandwidth adapted to new optimum values in conjunction with new equalizer tap values.

It appears that no serious interaction problems occur between the two adaptive systems.

Figure 5.6 shows a series of performance curves where the probability of bit error has been plotted as a function of SNR. For time invariant received pulses, curve no. 1 shows the performance of the dual adaptive system with all parameters adapting, and curve no. 2 shows the performance of the system when all parameters are fixed at a set of optimum values. The small degradation in performance incurred by letting the parameters adapt is due to noisy estimates which cause the parameters to wander about their optimum values. These two curves can be compared to the performance of a receiver having a matched filter and fixed optimum tap values (curve no. 3). It can be seen that the fixed low pass filter performs only about 1.2 dB worse than the matched filter which is considerably more difficult

SNR = 13.7 dB
EQUALIZER ADAPTING

Figure 5.5 Analog Prefilter Response to Step Changes in Received Pulse Bandwidth

Figure 5.6 Performance of Various Receivers
as a Function of SNR

to construct. Curve no. 4 is the isolated pulse lower bound shown for reference.

The results for tracking performance are particularly interesting. For this test the bandwidth of the received pulse was made to vary sinusoidally over a range of $\pm 20\%$ from nominal with a period equal to about 75,000 received pulses. (This is not a rapid variation rate, but the equalizer tap values are much slower to adapt than the analog parameters and it was considered desirable to allow the equalizer to follow the variations fairly closely.) The average probability of error has been plotted against the nominal SNR as curve no. 5 of Figure 5.6. For comparison, the analog filter parameters were then fixed at the optimum values for the nominal received pulse, and the simulation was repeated. The results were essentially identical to the adaptive case and so only one curve has been shown. These results are somewhat surprising but occur for two main reasons.

First it appears that the adaptive equalizer has sufficient degrees of freedom to compensate for the improper bandwidth and sampling time used in the prefilter. To show this, the tracking simulation

was repeated with both the analog parameters and the equalizer taps fixed at values that were optimum for the nominal received pulse. The results are shown as curve no. 6 of Figure 5.6. In this case the performance is about 0.8 dB worse indicating that all of the tracking improvement can be made by the equalizer alone for this example.

The second reason for the apparent lack of improvement with the adaptive analog filter is that the SNR of the filter output samples is very insensitive to the sampling time. The filter output is a third order pulse which has a broad peak, and so the equalizer input changes very little with sampling time. This is illustrated in Figure 5.7 where the performance of a time invariant system is shown for various fixed sampling times over a $\pm 20\%$ range from nominal. Similar results were observed as a function of prefilter bandwidth.

In general it may be said that this particular example exhibited a broad optimum with respect to the bandwidth and sampling time parameters. As a result there were many combinations of sampling time, bandwidth and equalizer tap values that were stable

Figure 5.7 The Effect of Sample Time on the Adaptive Receiver Performance

and gave the same performance within the measurement accuracy of the criterion used. This may be partly due to interaction between the two adaptive systems since they both use the same MSE criterion. (See Section 3.2.4).

If the prefilter were matched, then a maximum filter output SNR criterion could be used to keep it matched while tracking. The prefilter would then adapt completely independently of the equalizer (since the filter parameters depend only on the received pulse shape) and interaction would be avoided. In addition the analog filter's parameters would be more sensitively defined. Therefore, at low SNR it is probable that a system with a tracking matched filter would show more improvement over a fixed system than the example treated in this section.

In general the use of a low pass filter is probably a good suboptimum choice because the loss in performance is only a little more than 1.0 dB. The tradeoffs are easier filter implementation and reduced sensitivity to sampling time near the optimum. This sensitivity with a matched filter may cause some of the potential gain to be eroded by the noisy gradient estimates. For some examples such as this

one, it may not be necessary to adapt the analog filter at all because of the flexibility of the adaptive equalizer.

5.5.1.1 Additional Comments on Non-Matched Filter Effects

The problem of having to select the proper sampling time was mentioned in the previous section. Related to this problem is that of selecting the position of the main tap in the forward portion of the equalizer. For a decision feedback equalizer the pulse peak sample (or time zero sample when estimating $\theta(0)$) is usually over the last tap of the forward portion of the equalizer at the time the decision on $\hat{\theta}(0)$ is made. However, this is true only for a matched prefilter. The simulation results of this example indicated that the main sample should be over the first tap of the 3 tap forward equalizer. There seems to be no way to determine this a priori. A trial and error method was required. This is identified as an area requiring further work and clarification; with the exception of some work by Mark and Haykin⁽⁴⁸⁾ it appears to be studiously avoided in the literature.

It was observed that more feedback taps are required on the equalizer when a mismatched filter is used than when a matched filter is used. These extra taps provide the additional degrees of freedom the receiver requires to partially compensate for the mismatched filter.

With this example, difficulties were encountered in adapting the tap values, of the forward part of the equalizer, according to the signed correlations between the error and decisions for that tap. (This technique was employed in (11) to give a performance improvement.) However adaption using signed correlations between the error and the forward delay line sample value for the tap worked well. This could be a result of the fact that the main tap is the leftmost (first) tap on the forward equalizer. It appears that the former technique may not always work when mismatched filters are used. More study of this point is also required.

5.6 Summary

This chapter introduced, in detail, PAM systems with intersymbol interference and the use of adaptive receivers to track time varying received signals. A new technique suggested for these receivers was the

adaption of the parameters of the analog prefilter in addition to the use of an adaptive equalizer. It was pointed out that for practical receivers using mismatched filters, the straightforward criterion of maximum SNR at the filter output is no longer an optimum criterion.

The original work presented in Section 5.2 derived an algorithm for adapting the prefilter bandwidth based on a minimum MSE criterion at the equalizer output. Section 5.3 introduced some approximations that resulted in a very simple hardware realization of the adaptive algorithm. Section 5.4 outlined a similar algorithm for optimizing the sampling time.

Section 5.5 reported on the results of a hybrid computer simulation of a complete PAM system including a receiver with both an adaptive equalizer and an adaptive mismatched prefilter. Several results and conclusions based on the simulation work were given in Section 5.5.1 and additional comments pertaining to the use of equalizers with non-matched prefilters were given in Section 5.5.1.1.

CHAPTER 6

ADDITIONAL APPLICATIONS FOR ADAPTIVE SYSTEMS

The work of this thesis has shown analytically and experimentally that the use of gradient algorithms leads to simple and practical adaptive systems. It has also been demonstrated that the algorithms are stable when used to track time varying parameters and the performance improvements have been verified by computer simulations.

Specific performance results are difficult to obtain because the problems are often impossible to analyse mathematically and simulation results are more easily obtained for simple systems. The PAM system of Chapter 5 represented a fairly complex system with more than one set of parameters being adapted simultaneously. The results of that work indicated that the system did have improved tracking performance but that all of the gain could be made by the adaptive equalizer alone. Thus in many adaptive systems only some of the parameters may need to be made adaptive and some care should be taken to

pick those that give the best results. Unfortunately this selection is very difficult to make without some simulation or actual system tests.

This chapter indicates some of the many other areas of application for the simple and practical adaptive systems described in the preceding chapters. The systems are not analysed in any detail. It is an advantage of these adaptive systems that as long as a few simple criteria are satisfied, convergence of the system or tracking capability is assured. The basic criterion is that the performance measure selected should be unimodal near the performance optimum. As long as reasonable performance gradient estimates can be obtained, an adaptive system can be proposed without doing detailed analysis. This is especially useful for non-linear systems such as those described in the following sections.

6.1 Receivers for Nonlinear Modulation Systems

Nonlinear modulation is commonly used in systems that are more power limited than bandwidth limited, or in systems having a severe amplitude nonlinearity such as a satellite transponder operating in the saturated region. Terrestrial radio relay

equipment usually operates with non-linear power amplifiers as well and therefore phase or frequency modulation techniques are required. Some of the newest radio systems are digital and, since digital systems can provide a reference at the receiver for measuring their performance, the discussion in the following sections will be limited mainly to digital modulation.

6.1.1 General Phase or Frequency Demodulation

Figure 6.1 is a block diagram of a general phase or frequency modulation receiver showing a channel filter preceding the demodulator and a baseband filter following the demodulator. For a digital system the output is quantized and an error signal can be formed from the difference between the quantizer input and output. The system performance is a function of the bandwidth of the channel filter, and of the gain and bandwidth of the baseband filter. If the channel response changes, or if the SNR fluctuates (due to fading) then these filters can be adjusted to keep the receiver optimum. If the system is working fairly close to optimum, a MSE criterion will have a unique minimum point and

Figure 6.1 Generalized Phase or Frequency Modulation Receiver

gradient algorithms can be used as described in previous chapters to track variations in the received signal.

The gradients could be estimated either by using two filters with slightly different bandwidths (see Section 3.1.2) or in a sequential manner using only one filter and perturbing its bandwidth. In the second method the MSE is evaluated at one setting and then the bandwidth is changed and the MSE evaluated again. The difference between the two MSE values can be used as an estimate of the performance gradient. This method requires much less hardware since, for the channel filter adaption, only one demodulator is necessary. However, the first value of the MSE must be stored.

It is a great advantage of these gradient algorithm systems that they can be applied without having to calculate the gradients analytically. For nonlinear systems this can be quite difficult.

6.1.2 Phase Locked Loop Demodulators

The techniques of filter optimization and tracking derived in previous chapters can also be

directly applied to the demodulator itself in phase or frequency modulation systems. In this section a phase locked loop receiver has been selected as a demodulator. A simple block diagram of the receiver is shown in Figure 6.2. It can be shown⁽⁵⁷⁾ that the loop filter; the voltage controlled oscillator (VCO) characteristics and the post-loop filter are all functions of three things: the bandwidth of the modulating spectrum, the modulation deviation ratio, and the system modulation-bandwidth SNR. If any of these are changed, either deliberately or as a result of time variations in the channel, then the demodulator will no longer be optimum.

The parameters of the various filters in this demodulator can be adaptively adjusted to track variations in the received signal as discussed in Section 6.1.1. The post-loop filter corresponds to the baseband filter of Figure 6.1. The demodulator may be preceded by a channel filter in some instances, although the inherent filtering capabilities of the loop often provide satisfactory operation with only an amplitude limiter at the input.

Figure 6.2 A Phase Locked Receiver

Phase locked loops are often used to obtain frequency references by locking onto a pilot frequency. The loop must first capture or acquire the pilot signal which, in general, will be offset from the idle frequency of the loop and masked by noise. In this mode the loop may have a search capability for sweeping across a band of frequencies. Once locked however, the loop bandwidth should be as narrow as possible to reduce the phase jitter on the reference frequency output. There is therefore a requirement for some type of adaptive loop filter to optimize the performance for these two modes. In addition, if the reference frequency drifts, the loop filter must compromise between output phase jitter and tracking ability. This is essentially the standard tracking compromise discussed in earlier sections of this thesis. It should be possible to implement a simple adaptive system to optimize the phase locked loop under these conditions by using the same gradient algorithms to adjust the loop gain and bandwidth.

6.1.3 Feedback Communication Systems and Variable Data Rates

It was noted in the previous section that the performance of a phase or frequency modulation system

is a function of the bandwidth of the modulating spectrum, the modulation deviation ratio, and the system modulation-bandwidth SNR. Any fixed receiver is optimum only for one combination of these three parameters. Changes in SNR may result from changing noise levels in the channel, or from signal fades. If the SNR is degraded sufficiently the demodulator will go below threshold cutting off communication. However, the total system communication may not have to be disrupted. For a given SNR the system can be adapted to operate above threshold if the bandwidth is changed, or for a phase shift keyed (PSK) system, if the data rate is reduced.

These changes must be made at the transmitter as well as at the receiver, and, since the sensing of the performance must be done at the receiver, a feedback channel from the receiver to the transmitter is required.¹

For PSK systems a reduction of the symbol rate gives a reduction of the bandwidth (for small modulation index)⁽⁵⁸⁾ allowing the channel filter bandwidth to be reduced by approximately the same proportion. This reduces the noise allowed into the

1

Since many communication systems operate in a full duplex mode, the required feedback channels may already exist.

demodulator and operation above threshold can be re-established. The adaptive systems described in Section 6.1.2 will continue to optimize the various filters under the new conditions provided the filters are changed to new nominal values each time a major change is made in the transmit characteristics.

The data source and sink in such a system must be capable of operating with varying data rates or large buffers would be required at each end. Some systems, such as computer data communication networks, can readily adapt to changing data rates. In practice the data rate would not be continuously variable but would probably use rates related by a factor of 2 (due to the ease of deriving the various clock frequencies). Similarly the channel filters would be changed in nominal 3dB steps. In addition some hysteresis must be built into the system that adjusts the data rates to prevent oscillation between two rates. This can be done by sensing two or more MSE (or other performance criterion) values as high and low "alarms". If a high MSE setting is exceeded, the rate is reduced. If the MSE drops below a low setting, the rate is increased. If the limits are far enough apart, oscillations will be prevented.

Simple adaptive systems can therefore be successfully used in the joint optimization of the trasmitters and receivers.

6.2 Process Control in Manufacturing

A manufacturing process can be modeled as a complex system operating on many inputs, a_i , which may include temperature or pressure settings and time of processing. Figure 6.3 shows a model in which the desired product is represented by the value of d , and the process error by e . In this model the system is trying to produce an output with specified characteristics by varying the amount, w_i , of each input. (In other instances the object may be to minimize the cost of production while maintaing the quality within certain limits.)

If a MSE criterion is used as a performance measure then gradient algorithms can be used to adjust the input gains w_i . If the system is near optimum, and it can be assumed that the performance surface is unimodal, the algorithms will all be stable and will be able to track changes in the input parameters. These changes could be differences in quality or concentration of the raw materials; changes could

Figure 6.3 A Simple Model of a Complex Manufacturing Process

also occur in the response of the processor itself due to aging or corrosion. This is another example of a system for which an analytical solution for the optimum parameter set is often impossible to obtain.

6.3 Adaptive Adjustment of Complex Systems

The process control example of the previous section adapted a complex system to changing conditions by, in effect, changing the gain of a set of input signals. Two more general systems of this type are shown in Figure 6.4. In Figure 6.4(a) only one signal is processed but it has been split and passed through several simple adaptive filters having variable gains and bandwidths. Each filter could be controlled by an adaptive algorithm which attempts to maximize some performance measure defined either with respect to a reference value, or obtained using a quantized system output signal. If, for example, the filters have nearly disjoint frequency passbands the system could be used as an adaptive analog equalizer to compensate for amplitude and phase anomalies in a channel, or to track time varying frequency selective fading if the fluctuations are not too rapid.

This configuration could also be used with the

(a)

(b)

Figure 6.4 Simplified Implementations of
Complex Adaptive Systems

output summer removed as an adaptive frequency division demultiplexer for signals that have been multiplexed by a system like that of Figure 6.4(b). In this communication system the input signals may be given various bandwidths and powers to make the best overall use of the communication channel. The receiver adaptively corrects not only for changes made at the transmitter but for any changes that occur in the channel as well. In addition information from the receiver may be sent back to the transmitter via a feedback channel to improve the overall channel use.

The division of such a complex adaptive system into several parallel smaller and simpler ones makes the implementation much more practical.

The system of Figure 6.4(b) is a generalized version of Figure 6.3 in which each input is filtered in addition to having an adaptive gain control. These filters can also be controlled by adaptive algorithms again yielding a complex adaptive system that is practical to implement. It could be used in process control applications; as a frequency multiplexer; or any other adaptive application of the form of Figure 6.4(b).

6.3.1 Diversity Systems

Another type of system that fits the form of Figure 6.4(b) is the receiver for a diversity communication system. In radio relay systems space diversity is often used so that each a_i would become the signal received at a separate antenna. These signals can be optimally combined to improve the performance of the link. In a digital system, each repeater regenerates the digital signal before re-transmission so that an error measurement can be made. This can be used as previously described to adjust the receive filters in response to changing conditions in the channel to optimize each hop of the relay system.

Frequency diversity is also used on radio relay systems and with HF radio systems to combat frequency selective fading. A frequency diversity receiver would take the form of Figure 6.4(a) and the filters would be adapted to optimally process the received signal.

6.3.2 Adaptive Antenna Arrays

The model of Figure 6.4(b) also applies to arrays of receiving antennas which are really just

more complex space diversity receive systems. However, by proper geometric arrangement of the antennas; and by proper signal combining in amplitude and phase; the array can be made to take on special net antenna patterns. In particular the pattern can be made to have a null in the direction of a strong interfering signal. Some work in this area has been done by B. Widrow et al⁽²⁴⁾. Another good reference to general adaptive antenna arrays is (59). Adaptive methods using gradient algorithms appear to be well suited to the practical implementation of these systems which often involve a large number of antenna elements.

6.4 Some Future Trends in Adaptive Systems

For adaptive digital systems, gradient algorithms can usually be implemented using digital techniques. The continuing reduction in cost of integrated digital circuitry has made it feasible to apply these algorithms in systems that would have been impractical only a few years ago.

However there are other algorithms and adaptive approaches that can outperform gradient algorithms in many systems. The drawback with many of these methods is that they require complex estimation procedures,

such as Kalman filtering⁽³⁹⁻⁴²⁾ and other recursive methods; or they require considerable computational capability for the inversion of matrices or similar operations.^(13,14) For example, an adaptive PAM receiver structure proposed by Quershi and Newhall⁽⁶⁰⁾ performs within 2 dB of the isolated pulse lower bound and is considerably simpler than the optimum nonlinear receiver. It has an adaptive equalizer to partially correct for intersymbol interference and to force the combined response of the channel and equalizer to be as close as possible to a desired or reference response. The equalized output is then fed to a Viterbi detector. This detector can be fairly complex to implement but yields considerably improved performance. The system is also capable of tracking time variations in the channel.

Some recent developments in the field of computer technology have led to the production of the so called micro-computer or micro-processor. These processors, using read-only memories and requiring only one printed circuit card are now becoming available for \$500-\$1000 depending on the memory size and speed. It seems most likely that these computers will provide the practical means for the

implementation of these more sophisticated adaptive techniques. The future therefore appears to offer even greater adaptive performance gains as the technology advances and the procedures are refined; but the simple gradient techniques appear to be the ideal solution for many current problems.

CHAPTER 7

THESIS SUMMARY

The continuing drive for improved performance at lower cost has led to a search for methods of automatically optimizing systems to account for factors that are either unknown to the designer or subject to change while the system is in operation. It is highly desirable to do these adjustments on-line using only the normal system signals. In addition, because the basic fixed systems are often simple and inexpensive, the adaptive techniques should be constrained to be simple and practical to implement. In a preliminary investigation, gradient algorithms demonstrated a capability for practical hardware implementation and sufficient flexibility for use in a wide range of adaptive systems.

These results initiated the work of this thesis which has presented a study of the performance of adaptive systems using gradient algorithms. It has been shown analytically and experimentally that the use of these algorithms leads to simple and practical

adaptive systems. Performance results have been obtained for the application of these adaptive techniques to relatively simple communication or control systems demonstrating the improvement that results from the ability to track time varying changes in the environment.

Another approach applied here to obtain practical adaptive systems was the use of constrained optimization, in which the form of the system was fixed and several parameters were then adaptively optimized, allowing good suboptimum systems to be implemented when the unconstrained solution may be impossible to obtain or too complex to construct.

The study did not present detailed performance comparisons of many different algorithms. A few common types were selected and some changes proposed to make the physical realizations more practical. The work was more concerned with the selection of a good constrained system, what parameters should be selected for adaption and what performance criteria should be used in the adaptive algorithms. For example, comparisons of minimum MSE and maximum SNR were made in Sections 3.2.4 and 5.5.1. In particular the desirability of avoiding interaction problems by using different performance criteria

for different parameters was stressed.

It was demonstrated that joint selection of the operating system and the adaptive algorithm can lead to considerable simplification in the hardware implementation. One example of this was discussed in Sections 3.1.1 and 3.2.2 where it was shown that the receive filter can be specified so that its output noise power is independent of its bandwidth. This allowed a maximum SNR criterion to be replaced by a maximum output sample criterion with a great simplification in the resultant adaptive system. Another example was noted in Section 3.2.1 where the system signals available for measurement were selected so that the iteration equation would be linear instead of non linear in the adaptive parameter. (Eqn. 3.11(a) vs. Eqn. 3.11(b)). It is important to make these joint design decisions whenever this freedom is available.

Another significant point about the work in this thesis is that fixed step size parameter values were used in all the algorithms. Reference was made to algorithms that are capable of continuously adjusting their step size parameters. Those algorithms

that make the step size approach zero asymptotically are unsuitable for tracking work. Most other algorithms which adjust their step sizes throughout the adaption are quite complex; have very large computational requirements; and are not suitable for on-line adaption of simple processing systems.

One notable exception to the above comments on variable step size parameter algorithms is the Lucky algorithm⁽⁹⁾. This is actually a type of quantized algorithm (see Sections 2.1.5 and 2.3) that has been implemented in a special way. The sign of the gradient is used to increment or decrement a counter. When the counter overflows or underflows the parameter estimate is decreased or increased by a fixed amount; and the counter is reset to its mid-value. This averages the gradient estimates and has the effect of a variable step size parameter. Simulation work^(11,54) indicates that this algorithm usually exhibits superior performance in acquisition type problems. The dynamics of the algorithm are very complex; and both the counter size and the fixed parameter increment must be properly selected for optimum performance.

It is because of the difficulty in adaptively adjusting the step size parameters that this thesis concentrated on the use of fixed step size parameters. It was shown that these algorithms are stable and converge for time invariant systems; and that there is an optimum step size parameter to use when trading off speed of response for larger steady state error.

This thesis has also presented some original results for tracking time varying systems. At present there is very little literature available on the design of adaptive systems for time varying environments. Here again for simplicity gradient algorithms with fixed step size parameters were selected. It was shown that these algorithms are still stable when used with time varying systems and that again optimum values for the fixed step size parameters can be found. However, the equations are somewhat more difficult to use.

In most sections the analytical results have been compared with the results of simple computer simulations and very good agreements were observed.

A section on the adaptive adjustment of the step size parameters was included for completeness;

but the techniques given are not, in general, practical approaches to use. The extra improvement to be gained from using adaptive step size parameters is probably very small due to the time required to estimate the proper corrections and due to the approximations that must be introduced to implement the corrections. This method is probably at best a second order correction process.

One problem given particular consideration in this thesis was the PAM communication system in which the received pulses were disturbed by both additive noise and intersymbol interference. These systems, using receivers which consist of an analog filter followed by a digital transversal equalizer, have had considerable success in adapting the equalizer with gradient algorithms.^(9-12;32;33) In this thesis a method was presented for optimizing the analog filter parameters as well. A hybrid computer simulation of this system was performed which confirmed the proper operation of the dual adaptive receiver.

While the results obtained for the particular example chosen indicated no significant improvement in performance over the receiver with a fixed analog

filter; simulation work using a two section sampled data adaptive receiver with a linear equalizer recently carried out by Mark and Budihardjo^(55,56) demonstrated a performance improvement for their example. Only time invariant received signals were tested, but one may conjecture that their system would also perform well in time varying situations. The discussion in Sections 5.1 and 5.5.1 indicates that in some cases significant improvement can be expected while in other cases it cannot. Time did not permit the testing of more complex transmitted pulse shapes or more complex receive filters which were difficult to implement on the available analog computers.

It seems appropriate to make one more general comment on the study of complex adaptive systems. For problems such as the adaptive PAM receiver it is obvious that a general theoretical treatment is too complex to be useful. Constrained optimizations, (such as was used here by fixing the form of the analog filter before deriving an adaptive mechanism) give simpler answers to some problems. However it seems always to be necessary to resort to a computer simulation in order to check the actual system performance. It is hoped, by using a few specific simulations,

to gain insight into the performance to be expected when the system is actually built. The results of this study indicate that the simulation examples should be carefully chosen, and that several examples should be tried since the results of specific simulations may be peculiar to that particular system.

In practice if a problem of this complexity is to be solved, various techniques of gaining improvement should be individually investigated to see if they warrant incorporation into the adaptive system. For PAM receivers some of these techniques would include selection of the adaptive prefilter, adaptive adjustment of the step size parameters in the algorithms, equalizer tap gain adjustment by correlating the error with the sample values as opposed to the decisions, selection of special fixed sampling times rather than adaptive adjustment, and selection of the proper main tap on the forward equalizer.

It would appear best to approach each problem with ingenuity and some specific simulations rather than a general theoretical technique.

7.1 Topics for Further Study

This section gives a brief summary of areas in which further research may extend the knowledge of practical adaptive techniques. Some of these points were discussed in greater detail at appropriate places in the main text.

It would perhaps have been useful to do more computer simulation examples to better establish some of the techniques used. However, this may be of limited value unless a specific problem is to be solved.

A comparison of algorithms, especially with the Lucky algorithm, is suggested. The Lucky algorithm appears to work well on acquisition problems and it may also be very efficient for tracking problems. It has the additional advantage of being very easy to implement with digital circuitry.

More work could be done with PAM systems using different prefilter/equalizer combinations; and more study is needed to enable determination of the main forward equalizer tap position when mismatched prefilters are used. Some discussion

was also given in Section 5.5.1.1 on the effects of correlating the error with the decisions as opposed to the delay line sample values. Techniques of this type, which actually select modified optima, deserve further investigation since they do not always result in a performance improvement.

It was pointed out in Section 6.4 that recent advances in digital circuit technology will soon make more complex adaptive procedures practical to use in an on-line mode. It would seem appropriate to pursue some of these techniques through simulation with a view to implementation using the new micro processors.

Section 4.5 briefly discussed some methods of adaptively estimating the step size parameter values. Some further work in this area may yield useful results especially as more complex procedures become practical.

REFERENCES

1. M. Schwartz; W.R. Bennett and S. Stein, "Communication Systems and Techniques", McGraw Hill Book Co., New York; 1966.
2. J.E. Gibson, "Nonlinear Automatic Control", McGraw Hill Inc., New York; 1963.
3. T.H. Lee; G.E. Adams; and W.M. Gaines, "Computer Process Control: Modeling and Optimization", John Wiley and Sons Ltd., New York; 1968.
4. J.A. Aseltine; A.R. Mancini, and C. Sarture, "A Survey of Adaptive Control Systems", Trans. IRE PGAC-6; pp. 102-108; Dec. 1958.
5. P.R. Stromer, "Adaptive or Self-Optimizing Control Systems - A Bibliography", Trans. IRE PGAC-4 No. 1; pp. 65-68, May 1959.
6. T. Kailath; "Adaptive Matched Filters", Ch. 6 in "Mathematical Optimization Techniques", edited by Richard Bellman, University of California Press; 1963.
7. D.J. Sakrison; "Iterative Design of Optimum Filters for Non Mean Square Error Criteria", IEEE Trans. on Information Theory, pp. 161-167; 1963.
8. 1970 IEEE Symposium on Adaptive Processes.
9. R.W. Lucky, "Techniques for Adaptive Equalization of Digital Communication Systems", Bell Syst. Tech. J.; vol. 45; pp. 255-268, Feb. 1966.
10. J.G. Proakis and J.H. Miller; "An Adaptive Receiver for Digital Signaling through Channels with Intersymbol Interference", IEEE Trans. Inform. Theory; vol. IT-15; pp. 484-497; July 1969.
11. D.A. George; R.R. Bowen, and J.R. Storey, "An Adaptive Decision Feedback Equalizer", IEEE Trans. on Commun. Technol.; vol. com-19, pp. 281-293; June 1971.

12. H. Kobayashi, "Simultaneous Adaptive Estimation and Decision Algorithm for Carrier Modulated Data Transmission Systems", IEEE Trans. Commun. Technol., vol. com-19, pp. 268-280, June 1971.
13. G.S.G. Beveridge and R.S. Schechter, "Optimization: Theory and Practice", McGraw Hill Inc., New York, 1970.
14. D.J. Wilde, "Optimum Seeking Methods", Prentice Hall Inc., 1964.
15. J. Kowalik and M.R. Osborne, "Methods for Unconstrained Optimization Problems", American Elsevier Publishing Co., Inc. New York, 1968.
16. D.J. Sakrison, "Stochastic Approximation: A Recursive Method for Solving Regression Problems", in "Advances in Communication Systems", vol. 2, edited by A.V. Balakrishnan, Academic Press, 1966.
17. A. Gersho, "Linear Adaption", Proceedings of the 1969 Symp. on Comp. Proc. in Commn., vol. XIX, pp. 653-664, Polytechnic Press, P.I.B., 1970.
18. H. Robbins and S. Monro, "A Stochastic Approximation Method", Annals. Math. Statist., vol. 22, 1951, pp. 400-407.
19. J. Kiefer and J. Wolfowitz, "Stochastic Estimation of the Maximum of a Regression Function", Annals. Math. Statist., vol. 23, 1952, pp. 462-466.
20. R. Fletcher and C.M. Reeves, "Function Minimization by Conjugate Gradients", Computer J., vol. 7, 1964, pp. 149-154.
21. M.R. Hestenes and E.L. Stiefel, "Methods of Conjugate Gradients for Solving Linear Systems", J.Res. Nat. Bur. Standards, Sect. B, vol. 49, 1952, p. 409.
22. R. Fletcher and M.J.D. Powell, "A Rapidly Converging Descent Method for Minimization", Computer J., vol. 6, 1963, pp. 163-168.

23. J.C. Hancock and P.A. Wintz; "Signal Detection Theory", McGraw Hill Inc., New York, 1966.
24. B. Widrow, P.E. Mantey, L.J. Griffiths and B.B. Goode, "Adaptive Antenna Systems", Proc. IEEE, vol. 55, p. 2143, Dec. 1967.
25. B. Widrow, "Adaptive Sampled-Data Systems", Tech. Report No. 2104-1, July 15, 1960, Stanford Electronics Laboratories, Stanford University, Stanford, California.
26. R.W. McLaren, "A Stochastic Automaton Model for the Synthesis of Learning Systems", IEEE Trans. on Systems Science and Cybernetics, vol. SSC-2, No. 2, Dec. 1966, pp. 109-114.
27. G.J. McMurtry and K.S. Fu, "A Variable Structure Automaton Used as a Multimodal Searching Technique", IEEE Trans. Auto. Control, vol. AC-11, No. 3, July 1966, pp. 379-387.
28. J.S. Riordon, "Dual Control Strategies for Discrete State Markov Processes", Part I, Int. J. Control, 1967, vol. 6, No. 3, pp. 249-261.
29. J.S. Riordon, "Dual Control Strategies for Discrete State Markov Processes", Part II, Int. J. Control, vol. 6, No. 4, pp. 317-330.
30. C. Devieux and R. Pickholtz, "Adaptive Equalization with a Second Order Gradient Algorithm", Proceedings of the 1969 Symp. on Comp. Proc. in Commn., vol. XIX, pp. 665-681, Polytechnic Press, P.I.B., 1970.
31. T.J. Schonfeld and M. Schwartz, "Rapidly Converging Second-Order Tracking Algorithms for Adaptive Equalization", IEEE Trans. Inform. Theory, vol. IT-17, pp. 572-579, Sept. 1971.
32. R.W. Lucky, "Automatic Equalization for Digital Communication", B.S.T.J., vol. 44, No. 4, pp. 547-588, April 1965.
33. A. Gersho, "Adaptive Equalization of Highly Dispersive Channels for Data Transmission", B.S.T.J., vol. 48, No. 1, pp. 55-70, January 1969.

34. T.P. Daniell, "Stochastic Approximation Procedures for Engineering Applications", Proc. IEEE (Letters), vol. 57, No. 4, pp. 733-734, April, 1969.
35. R. De Figueiredo, "Convergent Algorithms for Pattern Recognition in Nonlinearly Evolving Nonstationary Environment", Proc. IEEE (Letters), vol. 56, No. 2, pp. 188-189, Feb. 1968.
36. V. Dupac, "A Dynamic Stochastic Approximation Method", Ann. Math. Statist., vol. 36, pp. 1695-1702, 1965.
37. A.H. Jazwinski, "Limited Memory Optimal Filtering", IEEE Trans. Auto Control, pp. 558-563, October 1968.
38. A.H. Jazwinski, "Adaptive Filtering", Automatica, vol. 5, pp. 475-485, Pergamon Press, 1969.
39. R.E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", Trans. ASME J. Basic Eng., Ser. D, vol. 82, Mar. 1960, pp. 34-45.
40. R.E. Kalman and R.S. Bucy, "New Results in Linear Filtering and Prediction Theory", Trans. ASME J. Basic Eng. Ser. D, vol. 83, Dec. 1961, pp. 95-108.
41. R.E. Kalman, "Design of a Self-Optimizing Control System", Trans. ASME J. Basic Eng. Ser. D, vol. 80, 1958, pp. 468-478.
42. D.T. Magill, "Optimal Adaptive Estimation of Sampled Stochastic Processes", IEEE Trans. on Auto. Control, AC-10, No. 4, October 1965, pp. 434-439.
43. H.J. Kushner, "A Simple Iterative Procedure for the Identification of the Unknown Parameters of a Linear Time Varying Discrete System", Trans. ASME, Journal of Basic Engineering, vol. 85, pp. 227-235, June 1963.
44. A. Gersho, "Adaption in a Quantized Parameter Space", Sixth Annual Allerton Conference on Circuit and System Theory, Monticello, Illinois, Oct. 1968.

45. R.W. Chang, "Analysis of Convergence Rate of Automatic Equalizers for Fast Start-up Digital Communications", 1971 Int'l. Conf. Commun. Record, pp. 21.1-21.6.
46. R.W. Lucky and H.R. Rudin "An Automatic Equalizer for General Purpose Communication Channels", B.S.T.J. vol. 46, pp. 2179-2208, November 1967.
47. G. Ungerboeck, "Nonlinear Equalization of Binary Bipolar Signals in Gaussian Noise", 1971 Int'l. Conf. Commun. Record, pp. 21.31-21.36.
48. J.W. Mark and S.S. Haykin, "Adaptive Equalisation for Digital Communication", Proc. IEEE, vol. 118, No. 12, pp. 1711-1720, December 1971.
49. R.J. Keeler, "Construction and Evaluation of a Decision Feedback Equalizer", 1971 Int'l. Conf. Commun. Record, pp. 21.8-21.13.
50. D.A. George, "Matched Filters for Interfering Symbols", IEEE Trans. Inform. Theory, IT-11, No. 1, pp. 153-154, January 1965.
51. D.W. Tufts, "Nyquist's Problem - The Joint Optimization of Transmitter and Receiver in Pulse Amplitude Modulation", Proc. IEEE, vol. 53, No. 3, pp. 248-259, March 1965.
52. M.E. Austin, "Decision Feedback Equalization for Digital Communication over Dispersive Channels", M.I.T./R.L.E. Tech. Rep. 461, August 11, 1967.
53. D.C. Coll and D.A. George, "A Receiver for Time-Dispersed Pulses", Conf. Rec. 1965 IEEE Int. Conf. Commns., pp. 753-758.
54. M.C.J. Cowpland and D.C. Coll, "A Quantitative Comparison of Several Adaptive Equalizer Algorithms", Proceedings of the 1969 Symp. on Comp. Proc. in Commn., vol. XIX, pp. 683-694, Polytechnic Press, P.I.B., 1970.
55. J.W. Mark and P.S. Budihardjo, "Joint Optimization of Receiver Filter and Equalizer", IEEE Trans. on Commns. vol. 21, No. 3, Mar. 1973, pp. 264-266.

56. J.W. Mark and P.S. Budihardjo; "Performance of Jointly Optimized Prefilter-Equalizer Receivers"; IEEE Trans. on Commns.; vol. 21; No. 8; pp. 941-945; August 1973.
57. W.C. Lindsey and R.C. Tausworthe; "A Survey of Phase-Locked Loop Theory"; Jet Propulsion Laboratory; Pasadena; California; 1967. Report for Information only.
58. "Transmission Systems for Communications"; Bell Telephone Laboratories; Inc.
59. C.A. Baird, G.G. Rassweiler, C.L. Zahm, and G.P. Martin; "Adaptive Processing for Antenna Arrays"; Final Report; Radiation Inc., Melbourne, Florida; July 1972.
60. S.U.H. Qureshi and E.E. Newhall; "An Adaptive Receiver for Data Transmission over Time-Dispersive Channels"; IEEE Trans. on Inform. Theory; IT-19; No. 4; pp. 448-457; July 1973.

APPENDIX A

This Appendix shows that the noise term z_A of Eqn. 3.16 satisfies the requirements placed on z_i of Eqn. 2.21.

These requirements are:

- (i) z_A^k is zero mean
- (ii) z_A^k is white
- (iii) z_A^k is independent of $e_A^k = A^k - A^*$

$$(i) \quad E\{z_A^k\} = 2E\{A^*(x^k + y^k)^2 - \hat{\theta}^k(x^k + y^k)\}$$

$$A^* = \frac{|x|}{\overline{x^2} + \sigma_y^2} \quad \text{from Eqn. 3.13}$$

$$\therefore E\{z_A^k\} = \frac{2E\{|x|(x^k + y^k)^2\}}{\overline{x^2} + \sigma_y^2} - 2|\overline{x^k} + \overline{y^k}|$$

$$\text{Now } E\{y^k\} = \overline{y^k} = 0 \text{ and } E\{x^k y^k\} = 0$$

$$\therefore E\{z_A^k\} = \frac{2|x|(\overline{x^2} + \sigma_y^2)}{\overline{x^2} + \sigma_y^2} - 2|x| = 0$$

Thus z_A^k is zero mean

The MSE is given by

$$E\{(z_A^k)^2\} = 4 E\{[A^*(x^k + y^k)^2 - \hat{\theta}^k(x^k + y^k)]^2\} .$$

Further reduction of this expression is given in Appendix B.

- (ii) It is assumed that, if necessary, the data samples for the adaptive loop are taken far enough apart to ensure that z_A is independent from sample to sample. This assumption of whiteness usually holds in practice and simplifies the analysis.
- (iii) Since z_A^k is a function of A^* , a fixed quantity, and not a function of A^k then $E\{z_A^k e_A^k\} = 0$ (because z_A^k is zero mean).

APPENDIX B

In this appendix the expressions for $E\{(z_A)^2\}$ and $E\{e_A^2\}$ of Section 3.2.1 are reduced to simpler forms. The numerator of Eqn. 3.19(a) is

$$\begin{aligned}\epsilon^2 E\{z_A^2\} &= 4\epsilon^2 \overline{[A^*(x+y)^2 - \hat{\theta}(x+y)]^2} \\ &= 4\epsilon^2 \left[\overline{(A^*)^2 (x+y)^4} + \overline{\hat{\theta}^2 (x+y)^2} - 2\overline{A^* \hat{\theta} (x+y)^2} \right]\end{aligned}$$

$$\text{Now } \overline{(x+y)^4} = x^4 + 6x^2\sigma_y^2 + 3\sigma_y^4$$

Note that $\bar{x} = x$; $\overline{x^2} = x^2$, etc. since x is deterministic for the problem.

$$\overline{(x+y)^2} = x^2 + \sigma_y^2$$

$$\text{and } \overline{\hat{\theta}(x+y)^3} = |x|^3 + 3|x|\sigma_y^2$$

$$\text{Now } A^* = \frac{|x|}{x^2 + \sigma_y^2}$$

Thus the numerator of Eqn. 3.19(a) is

$$\begin{aligned}4\epsilon^2 \left[\frac{x^2}{x^2 + \sigma_y^2} |x|^4 + 6x^2\sigma_y^2 + 3\sigma_y^4 + x^2 + \sigma_y^2 \right. \\ \left. - \frac{2|x|}{x^2 + \sigma_y^2} (|x|^3 + 3|x|\sigma_y^2) \right]\end{aligned}$$

The denominator is

$$\begin{aligned}
 & 1 - \overline{(1 + 2\epsilon(x + y)^2)^2} \\
 &= -4\epsilon^2 \overline{(x + y)^4} - \overline{4\epsilon(x + y)^2} \\
 &= -4\epsilon^2(x^4 + 6x^2\sigma_y^2 + 3\sigma_y^4) - 4\epsilon(x^2 + \sigma_y^2)
 \end{aligned}$$

Dividing the denominator into the numerator and performing the required algebra results in

$$E\{(e_A)^2\} = \frac{-\epsilon\sigma_y^2(x^2 + \sigma_y^4)}{(x^2 + \sigma_y^2)^2(x^4 + \sigma_y^2 + \epsilon(x^4 + 6x^2\sigma_y^2 + 3\sigma_y^4))}$$

This is the required Eqn. 3.19(b).

APPENDIX C

In this appendix it is shown that the noise term z_T^k of Eqn. 3.24 satisfies the requirements placed on z_i^k of Eqn. 2.21.

These requirements are:

- (i) z_T^k is zero mean
- (ii) z_T^k is white
- (iii) z_T^k is independent of $e_T^k = T_S^k - T_S^*$
- (iv) z_T^k is independent of $\left(1 + \beta \frac{\partial^2 x(T_S^k)}{\partial T_S^2} \right) \bigg|_{T_S^k = T_S^*}$

$$\begin{aligned}
 \text{(i)} \quad E\{z_T^k\} &= \frac{1}{2\delta} E\{[y(T_S^k + \delta) - y(T_S^k - \delta)] \hat{\theta}^k\} \\
 &= \frac{1}{2\delta} [E\{y(T_S^k + \delta)\} - E\{y(T_S^k - \delta)\}] E\{\hat{\theta}^k\} \\
 &= 0
 \end{aligned}$$

since y is a stationary process.

- (ii) It was assumed that y is a white process; and that if necessary samples would be taken far enough apart to make this true in practice. Thus z_T^k is a white process.

(iii)

$$E\{z_T^k e_T^k\} = E\{(T_S^k - T_S^*)z_T^k\} = 0$$

since z_T is white it is not dependent upon the sampling time T_S^k , or upon the constant T_S^* .

Thus the above independence holds.

(iv) The derivative is not a function of the random variable T_S^k since it is evaluated for $T_S = T_S^*$, a constant. Thus the random variable z_T^k is not dependent, in the statistical sense, on the term

$$\left(1 + \beta \frac{\partial^2 x(T_S^k)}{\partial T_S^2} \right) \bigg|_{T_S = T_S^*}$$

Evaluation of $\sigma_{z_T}^2 = E\{(z_T^k)^2\}$

$$\begin{aligned} E\{(z_T)^2\} &= \left(\frac{\hat{\theta}}{2\delta}\right)^2 E\{[y(T_S^k + \delta) - y(T_S^k - \delta)]^2\} \\ &= \frac{1}{4\delta} \left[2E\{y^2\} - 2E\{y(T_S^k + \delta) - y(T_S^k - \delta)\} \right] . \end{aligned}$$

$E\{y^2\} = N_0/2$ using Eqn. 3.7, and it can also be shown that

$$E\{y(T_S + \delta)y(T_S - \delta)\} = \frac{N_0}{2} e^{-2\alpha\delta}.$$

$$\therefore E\{(z_T)^2\} = \frac{N_0}{4\delta^2} (1 - e^{-2\alpha\delta})$$

For small values of δ this can be approximated as

$$\begin{aligned} E\{(z_T)^2\} &\approx \frac{N_0}{4\delta^2} \left[1 - 1 + 2\alpha\delta - \frac{4\alpha^2\delta^2}{2} \right] \\ &= \frac{N_0}{2} \left[\frac{\alpha}{\delta} - \alpha^2 \right] \end{aligned}$$

APPENDIX D

This appendix computes the values of $\rho_{j,i}$ and $\sigma_{p_i}^2$ used in Section 4.1. From Section 2.3

$$\begin{aligned}\rho_{1,i} &= \frac{\overline{(d_i^{m+1} d_i^m)}}{\sigma_{d_i}^2} \\ &= \frac{\overline{(p_i^{m+1} - p_i^m)(p_i^m - p_i^{m-1})}}{\overline{(p_i^{m+1} - p_i^m)^2}} = \frac{-\overline{p_i^2}}{2\overline{p_i^2}} = -\frac{1}{2}\end{aligned}$$

Clearly $\rho_{j,i} = 0$ for $j > 1$ when p_i^* is a white process. Thus from Eqn. 2.55

$$\sigma_{p_i}^2 = \frac{\alpha_i^2 \sigma_{z_i}^2}{1 - \overline{H_i^2}} + \frac{\sigma_{d_i}^2}{1 - \overline{H_i^2}} (1 - \overline{H_i})$$

$$\text{where } \overline{H_i} = 1 + \alpha_i \frac{\partial^2 V}{\partial p_i^2} \text{ and } \overline{H_i^2} = \overline{\left(1 + \alpha_i \frac{\partial^2 V}{\partial p_i^2}\right)^2}$$

$$\therefore \sigma_{p_i}^2 = \frac{\alpha_i^2 \sigma_{z_i}^2}{1 - \overline{H_i^2}} + \frac{\sigma_{d_i}^2}{1 - \overline{H_i^2}} \left(-\alpha_i \frac{\partial^2 V}{\partial p_i^2} \right)$$

$$= \frac{\sigma_{d_i}^2(Y) - \alpha_i \sigma_{z_i}^2}{2Y + \alpha_i Z}$$

where $Y = \frac{\overline{\partial^2 V}}{\partial p_i^2}$

and $Z = \left[\frac{\overline{\partial^2 V}}{\partial p_i^2} \right]^2$

APPENDIX E

This appendix derives the optimum fixed step size parameter, α_i^* , for the conditions of Section 4.1.

Eqn. 4.6 gives the mean square parameter error as

$$\sigma_{p_i}^2 = \frac{\alpha_i^2 \sigma_{z_i}^2}{1 - \overline{H_i^2}} + \frac{\sigma_{d_i}^2}{1 - \overline{H_i^2}} \left[1 + 2\rho_{1,i} \overline{H_i} + 2\rho_{2,i} (\overline{H_i})^2 \right]$$

$$\text{where } \overline{H_i} = 1 + \alpha_i \frac{\overline{\partial^2 V}}{\partial p_i^2}$$

$$\text{and } \overline{H_i^2} = 1 + 2\alpha_i \frac{\overline{\partial^2 V}}{\partial p_i^2} + \alpha_i^2 \left[\frac{\overline{\partial^2 V}}{\partial p_i^2} \right]^2$$

As in Appendix D these are more conveniently written as

$$\overline{H_i} = 1 + \alpha_i Y$$

$$\overline{H_i^2} = 1 + 2\alpha_i Y + \alpha_i^2 Z$$

$$\therefore \sigma_{p_i}^2 = \frac{\alpha_i^2 \sigma_{z_i}^2}{-2\alpha_i Y - \alpha_i^2 Z} + \frac{\sigma_{d_i}^2}{-2\alpha_i Y - \alpha_i^2 Z} \left[1 + 2\rho_{1,i} + 2\rho_{2,i} \right]$$

$$+ \alpha_i (2\rho_{1,i}Y + 4\rho_{2,i}Y) + \alpha_i^2 (2\rho_{2,i}Z) \Big]$$

It is shown in Section 4.3 (Eqn. 4.24) that for good steady state performance

$$\alpha_i^2 Z \ll \alpha_i Y$$

Thus defining $A \triangleq 1 + 2\rho_{1,i} + 2\rho_{2,i}$

$$B \triangleq 2\rho_{1,i}Y + 4\rho_{2,i}Y$$

we may write

$$\sigma_{p_i}^2 \approx \frac{\alpha_i^2 \sigma_{z_i}^2 + \sigma_{d_i}^2 (A + \alpha_i B)}{-2\alpha_i Y}$$

Minimizing $\sigma_{p_i}^2$ with respect to α_i gives

$$\frac{\partial \sigma_{p_i}^2}{\partial \alpha_i} = 0 = \frac{(-2\alpha_i Y)(2\alpha_i \sigma_{z_i}^2 + \sigma_{d_i}^2 B) - (\alpha_i^2 \sigma_{z_i}^2 + \sigma_{d_i}^2 (A + \alpha_i B)) (-2Y)}{4\alpha_i^2 Y^2}$$

$$\therefore \alpha_i^2 \sigma_{z_i}^2 = \sigma_{d_i}^2 A$$

$$\therefore \alpha_i = \frac{\sigma_{d_i}}{\sigma_{z_i}} \sqrt{A}$$

This is the optimum fixed step size parameter.

APPENDIX F

Estimation of Second Partial Derivatives

It is possible to make an estimate of a second partial derivative as illustrated in Figure F.1 and the equations below.

$$V(x_B + \delta) \approx V(x_B) + \delta \left. \frac{\partial V(x)}{\partial x} \right|_{x=x_B} + \frac{\delta^2}{2} \left. \frac{\partial^2 V(x)}{\partial x^2} \right|_{x=x_B}$$

$$V(x_B - \delta) \approx V(x_B) - \delta \left. \frac{\partial V(x)}{\partial x} \right|_{x=x_B} + \frac{\delta^2}{2} \left. \frac{\partial^2 V(x)}{\partial x^2} \right|_{x=x_B}$$

$$\therefore \delta^2 \left. \frac{\partial^2 V(x)}{\partial x^2} \right|_{x=x_B} \approx V(x_B + \delta) - 2V(x_B) + V(x_B - \delta)$$

$$\therefore \left. \frac{\partial^2 V(x)}{\partial x^2} \right|_{x=x_B} \approx \frac{1}{\delta^2} [V(x_B + \delta) - 2V(x_B) + V(x_B - \delta)]$$

Thus by taking three samples of the function V an estimate of the second partial derivative can be made. If this estimate is used in Eqn. 4.27 an estimate of the optimum fixed step size for time varying systems of the type considered in Section 4.3 can be made.

Figure F.1 Estimation of Second Partial Derivatives

APPENDIX G

Details of the Hybrid Computer Simulation of a PAM Communications System

This appendix describes the computer simulation configuration that was used to obtain the results reported in Chapter 5.

A basic block diagram of the system is shown in Figure G.1. A hybrid simulation involving two analog computers and a small digital computer was used. This gives a fairly realistic simulation in that some practical problems such as filter inaccuracies, filter impulse approximation, d.c. voltage offsets and component tolerances also show up in some of the analog circuitry used. In addition the small digital computer can simulate all the digital system hardware much more economically than any breadboard setup, and it can be easily modified to try different techniques. The digital computer can also control the whole system and gather statistical data on the adaptive performance.

The computers used in this work were a TR-10 analog computer, an EAI-580 analog computer and a PDP-8 digital computer with A/D and D/A converters.

Figure G.1 Hybrid Computer Simulation of a PAM System

The two analog computers were required since neither had sufficient amplifiers and multipliers to simulate the total analog portion of the system. This created some problems with d.c. offsets and drifts between the separate computer power supplies.

G.1 The Transmitter and Channel

The functions of the transmit filter and channel were combined into one filter since the example considered second order pulses being received at the channel output. The form of these pulses is

$$\theta t e^{-\beta t} \quad t \geq 0$$

where θ is a binary random variable taking values of ± 1 ; and β is a parameter which could be time varying as well as unknown at the receiver. The value of β indicates the "bandwidth to pulse rate" ratio of the system for a given transmission rate and can be used as an indicator of the intersymbol interference at the receiver input. The TR-10 was used to form a filter having an impulse response of the form $te^{-\beta t}$. The ability to vary β at the transmitter was included by using multipliers to change the gain in the feedback paths of the two integrators forming the pulse shape. The value of β could then be changed by adjusting

a potentiometer on the TR-10 computer. By replacing the d.c. value at the output of the potentiometer with a varying voltage from some other source, the value of β could be made time varying.

The input to this computer was a pseudorandom sequence of very narrow pulses used to simulate impulses. These pulses were triggered by the digital computer via the D/A converter driving the Pulse Interface Circuitry which in turn triggered a Hewlett-Packard 8005A Pulse Generator to produce 10 volt pulses one millisecond wide to excite the TR-10 filter. Negligible error was introduced by this approximation to impulses. (This approximation is necessary in any physical system for obvious reasons.)

G.2 Analog Receiver Filter

The EAI-580 analog computer received the second order pulses and added the channel noise. The noise was Gaussian with a flat power spectral density over a bandwidth much greater than the receiver bandwidth. The noise therefore appeared white to the receiver. The noise source used was a Hewlett-Packard Model 3722A Pseudorandom Noise Generator. The advantage

of this device is that the actual noise power spectral density is very easily measured and adjusted to produce various signal to noise ratios.

The analog portion of the receiver consisted of two low-pass filters with bandwidths that could be adjusted by means of multipliers in the feedback paths as used in the transmitter/channel filter. The values of the bandwidths were adaptively adjusted, by the digital computer through its D/A converters, using the algorithms described in Chapter 5. The outputs of the filters were sampled, converted to digital values, and read into the PDP-8 where the adaptive algorithms and transversal equalizers were simulated. The sampling time at the analog filter outputs was also adaptively controlled by the PDP-8 computer.

G.3 Digital Receiver and Control

The PDP-8 computer was used for all the digital simulation required and for control of the whole simulation. The adaptive techniques used to control the analog filters were implemented by the PDP-8 as described in Chapter 5. The computer also simulated

both linear and decision feedback equalizers¹ which were made to adapt in a manner similar to that described in Reference (11). In addition the timing of the transmitted pulses and the receiver sampling were under control of this digital computer.

Since the transmitted pulses were generated by the PDP-8, it could easily measure the error rate by comparing the receiver decisions with a delayed version of the transmitted information. The computer was therefore used to gather statistical data on the receiver performance and print out the results. Several simulation runs (12-20) were averaged to get the final data.

The actual transmission rate in the simulated system was determined by the ability of the digital computer to handle all the simulation, control and bookkeeping tasks. The time required to do this was about 20 milliseconds. Therefore, allowing for a ± 5 millisecond adjustment of the sampling time ($\pm 20\%$) the nominal pulse rate was selected as 40 pulses per second ($T = 25$ msec.). In order to provide

¹ The simulations used here had 10 bit quantization of the tapped delay line sample values and 12 bit quantization of the delay line tap values since these were most easily simulated. However in practice, the tap values can usually be more coarsely quantized than the sample values(11).

severe intersymbol interference at the receiver input, the nominal value of $\beta = 32$ was chosen resulting in a received pulse shape te^{-32t} . This pulse has a nominal 3 dB bandwidth of about 3.3 Hz. If the pulses were transmitted at the Nyquist rate¹ this would imply a transmission rate of only 6.6 pulses per second instead of the 40 pulses per second actually used. Thus the system was transmitting at about six times the Nyquist rate. To ensure satisfactory operation without equalization most practical systems require a channel bandwidth of about 120% of the Nyquist bandwidth. This equalized system is therefore operating with about one-seventh of the bandwidth of an equivalent non equalized system and so considerable intersymbol interference occurs at the receiver input.

¹ The Nyquist rate is defined here as $2B$ where B is the received pulse 3 dB bandwidth.