1

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Services des thèses canadiennes

Ottawa, Canada
K1A 0N4

CANADIAN THESES

THÈSES CANADIENNES

NOTICE

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

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

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

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

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

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

AVIS

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

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

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

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

ADAPTIVE FILTERING IN A TIME-VARYING ENVIRONMENT

WITH APPLICATIONS TO EQUALIZATION OF

FADING HF CHANNELS

by

Evángelos Eleftheriou, Dipl. Eng., M.Eng.

A thesis submitted to the

Faculty of Graduate Studies

in partial fulfilment for the degree of

Doctor of Philosophy

Department of Systems and Computer Engineering

Faculty of Engineering

Carleton University

Ottawa, Ontario

October 1985

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-29369-3

The undersigned recommends to the Faculty of Graduate Studies
and Research, acceptance of the thesis

"ADAPTIVE FILTERING IN A TIME-VARYING ENVIRONMENT

WITH APPLICATIONS TO EQUALIZATION OF FADING HF CHANNELS"

submitted by Evangelos Eleftheriou, Dipl. Eng., M.Eng., in partial
fulfilment of the requirements for the degree of Doctor of Philosophy.

External Examiner

Thesis Supervisor

Chairman, Department of Systems
and Computer Engineering

Carleton University

November 1985

ABSTRACT

Adaptive signal processing algorithms derived from LS (Least Squares) cost functions are known to converge extremely fast and have excellent capabilities to "track" an unknown parameter vector. This work treats analytically and experimentally the steady-state operation of RLS (Recursive Least Squares) adaptive filters with exponential windows for stationary and nonstationary inputs. A new formula for the "estimation-noise" has been derived involving second- and fourth-order statistics of the filter input as well as the exponential windowing factor and filter length. Furthermore, it is shown that the adaptation process associated with "lag effects" depends solely on the exponential weighting parameter λ . In addition, the calculation of the excess mean square error due to the lag for an assumed Markov channel provides the necessary information about tradeoffs between speed of adaptation and steady-state error. It is also the basis for comparison with the simple LMS (Least-Mean-Square) algorithm. In the simple case of channel identification it is shown that the LMS and RLS adaptive filters could have the same tracking behavior. Also, we present new RLS restart procedures applied to transversal structures for mitigating the disastrous results of the third source of noise, namely, finite precision arithmetic.

A detailed investigation of the performance of fractional spaced FRLS (Fast Recursive Least Squares) decision feedback equalization receives on 3 KHz-wide HF fading dispersive channels has been conducted.

Our study is based not only on simulated data but also on real recorded HF data. It is shown that under high fade rate conditions the RLS adaptation algorithms reach their limitations as far as tracking capability is concerned.

Finally, we introduce a new decision feedback (DFE) structure. Using an off-line predictor concept we manage to split the joint adaptation process ~~associating~~ with the forward and feedback sections. This approach leads to a flexibility regarding the adaptation method adopted in the different sections of the DFE and, therefore, to an overall computational complexity reduction.

ACKNOWLEDGEMENTS

I wish to express my deep gratitude to my thesis supervisor, Prof. David D. Falconer. His continued guidance and innumerable suggestions and contributions to this dissertation are greatly appreciated.

I would like also to offer my appreciation to the members of the defense committee for their effort in reading this thesis.

Very special thanks are due to Dr. J.M. Cioffi for his very helpful comments and suggestions on the first part of this work.

Most importantly, my wife Maria and my daughter Katerina both deserve special mention for their patience and support during all those years of my study.

Finally, I thank Diane Dodds for typing the manuscript.

This work was supported financially in part by the Natural Sciences and Engineering Research Council and by the Department of Communications.

TABLE OF CONTENTS

ABSTRACT	iii
ACKNOWLEDGEMENTS	v
LIST OF FIGURES	ix
LIST OF SYMBOLS	xi
CHAPTER 1 -- INTRODUCTION	1-1
1.1 Adaptive Filter Principles	1-2
1.2 Organization of the Thesis	1-6
References - 1	R-1-1
CHAPTER 2 -- TRACKING PROPERTIES AND STEADY-STATE PERFORMANCE OF RLS ADAPTIVE FILTER ALGORITHMS	2-1
2.1 Recursive Least Squares Algorithms	2-4
2.2 RLS Adaptive Filters and the "Estimation-Noise"	2-8
2.3 RLS Adaptive Filter Response to Time-Varying Inputs	2-16
2.3.1 Tracking Behaviour	2-16
2.3.2 Calculation of the "Excess Lag" MSE	2-19
2.4 Finite Precision Arithmetic Considerations	2-26
2.4.1 Steady-State Performance	2-26
2.4.2 Periodic Restart for FRLS Algorithms	2-24
2.5 Simulation Results	2-33
2.5.1 Estimation-Noise	2-34
2.5.2 Lag Effects	2-37
2.5.3 Restarting Procedures	2-42
2.6 Summary	2-49
APPENDIX A-2	A-2-1
APPENDIX B-2	B-2-1
APPENDIX C-2	C-2-1
REFERENCES - 2	R-2-1

CHAPTER 3 --	ADAPTIVE DECISION-FEEDBACK-EQUALIZERS AND THEIR PERFORMANCE ON FADING DISPERSIVE CHANNELS	3-1
3.1.	HF Channel Characteristics	3-3
3.1.1	Multiple Propagation	3-3
3.1.2	Fading	3-4
3.1.3	Delay Distortion	3-5
3.2	Serial Data Transmission Over HF Channels	3-6
3.3	HF Channel Simulator	3-12
3.4	Adaptation Methods for a Decision Feedback Equalizer	3-14
3.4.1	The LMS Decision Feedback Equalization Algorithm	3-17
3.4.2	RLS Algorithms in Decision Feedback Equalization	3-18
3.4.2.1	FRLS Transversal DFE With An Efficient Periodic Restart	3-22
3.5.	DFE Receiver Performance Over Simulated HF Channels	3-26
3.5.1	Ideal Reference Mode	3-26
3.5.2	Decision Directed Mode	3-30
3.6	A Whitening Predictor As A Front-End For An LMS DFE	3-34
3.7	DFE Receiver Performance On Real HF Channels	3-37
3.7.1	Transmitted Signal and Characteristics of the Recorded Data	3-37
3.7.2	Performance Comparison of FRLS and LMS Adaptation Algorithms	3-41
3.8	Summary	3-52
	APPENDIX A-3	A-3-1
	REFERENCES - 3	R-3-1
CHAPTER 4 --	AN ADAPTIVE MMSE PREDICTOR BASED DECISION FEEDBACK EQUALIZER	4-1
4.1	MMSE Prediction and Channel Autocorrelation Sequence Factorization	4-5

4.1.1	Definitions	4-5
4.1.2	MMSE Prediction and Decision Feedback Equalization	4-8
4.2	Adaptation of the Prediction Based DFE Scheme	4-12
4.3	Simulation Results	4-19
4.4	Fractional-Spaced Implementation Considerations	4-22
4.5	Summary	4-25
REFERENCES - 4		R-4-1

CHAPTER 5 -- CONCLUSIONS

5.1	Concluding Remarks	5-1
5.2	Suggestions for Future Work	5-2

LIST OF FIGURES

CHAPTER 1

1.1	Adaptive Filtering System	1-1
1.2	Adaptive Filter Structures	
	(a) Transversal Filter	1-7
	(b) Lattice Filter	1-7

CHAPTER 2

2.1	Adaptive Filtering System	2-5
2.2	Total Excess MSE, ϵ_{tot} , Versus Exponential Windowing Factor λ	2-24
2.3	Misadjustment versus λ	2-36
2.4	SNR Versus λ	2-38
2.5	SNR Versus λ	2-39
2.6	SNR Versus λ	2-40
2.7	SNR Versus λ	2-41
2.8	SNR Versus Time	2-43
2.9	SNR Versus Time	2-44
2.10	Example of Restart Procedures for FRLS DFE (Channel MDA010) Restart Interval = 30	2-46
2.11	SNR for FRLS Algorithms for Channel MDA010 During the Restarting Interval (Restart at 300), Restart Interval = 30	2-47
2.12	Example of Restart Procedures for FRLS DFE (Channel MDA010), Restart Interval = 40	2-48

CHAPTER 3

3.1	Receiver Front-End	3-8
3.2	HF Channel Model	3-13
3.3	Decision Feedback Equalization Receiver	3-16
3.4	P_e Versus SNR for a 3-Path, 2T Spaced Rayleigh Fading Channel, MPS = 4T (Ideal Reference Mode)	3-28
3.5	P_e Versus SNR for a 4-Path Rayleigh Fading Channel MPS = 7T (Ideal Reference Mode)	3-31
3.6	P_e Versus SNR in Decision Directed Mode for a 3-Path, 2T Spaced Rayleigh Fading Channel, MPS = 4T, Fade Bandwidth 0.15 Hz	3-33

3.7	Lattice Whitening Filter Followed by a LMS DFE	3-36
3.8	P_e Versus SNR for a 3-Path, 2T Spaced Rayleigh Fading Channel, MPS = 4T (Ideal Reference Mode)	3-38
3.9	Error Statistics for FRLS and LMS DFE on Channel MDA011 (Ideal Reference Mode)	3-42
3.10	Error Statistics for FRLS and LMS DFE on Channel MDA010 (Ideal Reference Mode)	3-43
3.11	Error Statistics for FRLS and LMS DFE on Channel MDA011 (Decision-Directed Mode)	3-46
3.12	Error Statistics for FRLS and LMS DFE on Channel MDA010 (Decision-Directed Mode)	3-47
3.13	Blocks Containing One or More Errors. Ideal Reference Cases Shown Above, Decision-Directed Cases Shown Below	3-49
3.14	SNR Versus Time for Channel MDA010 (SNR Averaged With 500-Symbol Interval Time Constant)	3-50
3.15	SNR Versus Time for Channel MDA011 (SNR Averaged With 500-Symbol Interval Time Constant)	3-51

CHAPTER 4

4.1	PAM Communications Model Followed By A Symbol Rate Sampler	4-6
4.2	z-Transform of a MMSE DFE Based on a MMSE Predictor	4-13
4.3	DFE Structure Based on a Finite-Length Off-Line MMSE Predictor	4-15
4.4	An Adaptive Pred-DFE Receiver	4-17
4.5	MSE, in dB, Versus Time (Initial Convergence)	4-21
4.6	1/MSE, in dB, Versus Time (Steady-State Behaviour)	4-23

LIST OF SYMBOLS

$e(n)$	Error for filter adaptation at time nT .
$d(n)$	Desired output of adaptive filter at time nT .
$y(n)$	Output of the adaptive filter at time nT .
$J(\cdot)$	Performance criterion functional.
$\underline{C}(n)$	Adaptive filter tap-coefficient vector.
μ_n	Time-varying LMS algorithm step-size.
∇	Gradient operator.
$J(n)$	Weighted cumulative squared error at time nT .
λ	Exponential weighting factor in RLS algorithms.
$\underline{X}(n)$	Adaptive filter input signal vector at time nT .
$\hat{R}(n)$	Estimated autocorrelation matrix of $\underline{X}(n)$.
$\hat{V}(n)$	Estimated cross-correlation vector of $\underline{X}(n)$ and $d(n)$.
$\underline{K}(n)$	Kalman gain vector.
$\hat{\underline{C}}$	Parameter target vector to be tracked by the adaptive filter.
$E\{\cdot\}$	Expected value of the quantity in brackets.
R	Autocorrelation matrix
r_{ij}	The ij th element of R .
$\underline{V}$	Cross-correlation vector.
$\Delta \underline{C}(n)$	Tap-coefficient error vector.
$\underline{C}_{opt}$	Optimum Wiener solution.
δ	Initialization parameter in RLS algorithms.
$e_{opt}(n)$	The optimum error at time nT .

ϵ_{est}	Excess MSE due to estimation-noise.
$\sigma_{ x }^2$	Variance of the quantity $ x(n) ^2$.
$(\sigma_x^2)^2$	The squared value of the variance of $x(n)$.
M	Misadjustment.
z	z-transform variable.
$\Delta C(z)$	The z-transform of $\Delta C(n)$.
τ_i	Time constant associating with the lag error.
α	Nonstationary coefficient.
$\underline{W}(n)$	Zero mean white Gaussian random vector driving the nonstationarity model.
$v(n)$	Zero mean additive white noise.
σ_w^2	Variance of the components of the random vector $\underline{W}(n)$.
ϵ_{lag}	Excess MSE due to lag.
ϵ_{tot}	Total excess MSE.
N	Length of adaptive filter.
$\tilde{\underline{X}}(n)$	Adaptive filter input signal vector during restart.
$\tilde{d}(n)$	Modified desired outputs during restart.
$\underline{C}_{\text{aux}}(n)$	Tap-coefficient vector from auxiliary LMS algorithm.
σ_v^2	Additive noise variance.
MSE(n)	Measured SNR at time nT.
$r(t)$	Bandpass received waveform.
$h(\tau, t)$	Low-pass equivalent time varying impulse response.
T	Symbol interval duration.
$x(t)$	Complex baseband demodulated waveform.

$H(z)$	z-transform of the 2-pole Butterworth type filter.
$\hat{a}_n$	Data symbol decision.
$z(n)$	Complex DFE quantizer input.
$2N_1$	Length of forward filter in symbol intervals.
N_2	Length of feedback filter in symbol intervals.
$\{c_i^f\}_{i=0}^{2N_1-1}$	Forward complex-valued tap coefficients.
$\{c_i^b\}_{i=1}^{N_2}$	Feedback complex-valued tap coefficients.
$\xi(n)$	Latest input to DFE.
$\tilde{\xi}(n)$	Latest input to DFE during restart.
$\tilde{a}_n$	Modified desired response applied to a DFE during restart.
$C_{aux}(n)$	Auxiliary LMS DFE tap-coefficients at time nT .
$h(t)$	Impulse response of a stationary channel model.
r_i	Channel autocorrelation sequence.
f_i	Factored causal part of the channel autocorrelation sequence.
$R(z)$	z-transform of the channel autocorrelation sequence.
$F(z)$	z-transform of the factored causal part of the channel autocorrelation sequence.
$e_p(nT)$	Predictor error at time nT .
$\delta_{n,0}$	Kronecker delta.
$B(n)$	Feedback filter coefficients in the pred-DFE arrangement.
$e_{DFE}(nT)$	Error sequence used for adaptation in the traditional DFE.
$e_{pred-DFE}$	Error sequence used for adaptation in the prediction based DFE.
$q(n)$	Output of the feedback section at time nT .
$Q(z)$	z-transform of the output of the feedback section sequence.

CHAPTER 1

INTRODUCTION

The classical notion of a filter as a device exhibiting frequency selective properties for processing continuous or discrete-time signals has been extended by the application of statistical ideas. A filter can be defined to be any system that processes incoming signals or data in such a way to eliminate noise or smooth the signals or to predict the next input signal from time to time [1].

One special category of such systems is known as adaptive filters. They are termed adaptive because they possess self-optimizing behaviour and base their own characteristics upon estimated statistical measurements of the operating environment. These filters may suffer a performance loss when no a priori statistics of the input signal are known because estimation errors may prevent the adaptive system from reaching the optimal performance obtainable for known a priori statistics [1], [3].

The loss in performance is directly related to properties of the particular recursive algorithm used for automatic adjustment of the filter coefficients and the statistical properties of the environment in which the adaptive filter is operating in, as well. This loss in performance is of great importance in many applications such as equalization, system identification, adaptive antenna arrays, spectral estimation, adaptive prediction in speech coding, adaptive noise cancelling etc. [5], [14], [15], [18], [19], [20], [21], [22] and it will be our main concern in the first part of this thesis.

Although the field of adaptive filtering has grown enormously in the past five years or so embracing a variety of diverse areas, it is probably the area of communications which attracted major consideration as far as adaptive filtering applications are concerned. Furthermore, the recent introduction of the fast adapting least squares algorithms [10], [12], [18], [19], [21], [22] opened new perspectives for applications. It is a later concern of this thesis to examine the behaviour of the fast adapting Recursive Least Squares algorithms applied to Decision Feedback Equalization of a highly dispersive and nonstationary environment as the High Frequency Ionospheric Channel.

Another important consideration besides the demand for fast adaptation is the issue of computational complexity. Computational requirements are related not only to the adaptation method itself but also to the underlying receiver structure. Trying to meet the strict demands for low computational burden, a new predictor-based decision feedback equalization receiver is presented.

1.1 Adaptive Filter Principles

It is worthwhile to start by discussing very briefly the basic principles governing the operation of an adaptive filter. Figure 1.1 depicts the general adaptive system. It is considered to be of the transversal filter form of length N . The variable coefficients or tap-gains of the structure are indicated by the $N \times 1$ complex vector $\underline{C}(n)$, where n represents the discrete-time index. At each time instant the adaptive filter acts on the input signal $x(n)$ while the second input

Fig.1.1 Adaptive Filtering System

$d(n)$ represents the "desired response" sequence.

The parameters of the filter are to be adjusted according to some algorithm so as to minimize some performance functional $\mathcal{F}(\underline{C}(n), e(n))$ which depends jointly on the coefficient vector $\underline{C}(n)$ and on the error sequence

$$e(n) = d(n) - y(n) \quad (1.1-1)$$

which represents the difference between the desired response $d(n)$ and the output $y(n)$ and time n . For a perfect knowledge of the operating environment, the optimal tap-gains could be obtained by solving the system of equations that result by setting the gradient vector $\nabla \mathcal{F}$, with respect $\underline{C}(n)$, equal to zero. Existence and uniqueness of a solution depends solely on the selection of the functional $\mathcal{F}$. A common performance function is the mean square error, (MSE), $E\{|e(n)|^2\}$. For stationary inputs the MSE can be shown to be a second order function of the tap-coefficient vector. The quadratic nature of the MSE cost function guarantees the existence of a unique solution.

A well known iterative approach for adjusting the tap-coefficients towards their optimal settings is the "steepest-descent" algorithm [1], [3]. This method is described by the following relation

$$\underline{C}(n) = \underline{C}(n-1) + \mu_n \nabla [E\{|e(n)|^2\}] \quad (1.1-2)$$

where μ_n is a possibly time-varying step size.

In the absence of a-priori knowledge of ensemble averages the minimization of the MSE cannot be performed by numerical algorithms such as the steepest descent. Instead, one must resort to stochastic approximation algorithms [4]. With these algorithms noisy estimates of the gradient components are used in the search procedure for a global minimum. Such an algorithm is the gradient-type or LMS (Least Mean Square) [1], [3], adaptation algorithm which uses the gradient of the instantaneous squared error rather than the gradient of the MSE. Under these conditions the tap gains are updated as

$$\underline{C}(n) = \underline{C}(n-1) + \mu_n \nabla[|e(n)|^2] \quad (1.1-3)$$

If μ_n is sufficiently small, the sequence of coefficient vectors $\underline{C}(n)$ should gradually evolve in the direction of decreasing the average value of the $|e(n)|^2$, towards a value for which the gradient is zero.

An alternative optimization criterion is the exponentially-weighted sum of squared errors [12], [18], [19], [21], [22]

$$J(n) = \sum_{k=0}^n \lambda^{n-k} |e(k,n)|^2 \quad (1.1-4)$$

where $0 < \lambda \leq 1$, and $e(k,n)$ denotes the error signal at time k due to a tap-coefficient vector at time n . Minimization of the above cost function with respect to the filter's parameter $\underline{C}(n)$ involves an explicit least squares (LS) minimization at each time instant. Because least squares algorithms make better use of all the past available information than the stochastic gradient algorithms, they are expected to converge and track faster.

The most common adaptive signal processor is the linear transversal filter shown in Figure 1.2a. That filter comprises a tapped delay-line with coefficients $\{c_i\}_{i=0}^N$ so that the output $y(n)$ at time n is a linear combination of the $N+1$ previous inputs $x(n)$. The Recursive Least Squares (RLS) adaptation algorithm applied to a transversal structure will be the adaptive filter we shall use almost exclusively throughout this thesis.

Besides the transversal filter more recently the lattice filter structure has been proposed [13], [19], [21]. The finite impulse response lattice filter consists of say N stages. The outputs of the i th stage correspond to the i th order "forward" and "backward" prediction errors. Figure 1.2b shows a lattice structure with "partial correlation" coefficients $\{k_i\}_{i=1}^N$ and gain coefficients $\{g_i\}_{i=0}^N$ which are multiplying the outputs of the successive stages.

Implementation of the RLS adaptation algorithms in an infinite precision environment, can be achieved equivalently by a transversal or lattice filter. The fact that transversal RLS algorithms exhibit a computational complexity much lower than that of their lattice counterparts make them preferable especially in adaptive equalization and communications channel identification.

1.2 Organization of the Thesis

The rest of this thesis is divided into four chapters. Each of the following three chapters forms an entity in itself and a

(a) Transversal filter (boxes are one sample delays)

(b) Lattice filter

Fig. 1.2 Adaptive filter structures

presentation of novel developments. A literature review on publications related to this work, is made at the introduction of each chapter. Also a list of the appropriate references is provided at the end of the appropriate chapter. Chapter 2 presents an analytic and experimental investigation on the steady-state operation of RLS adaptive filters with exponential windows for stationary and non-stationary inputs. Special emphasis has been given to the finite precision error problem of RLS transversal structures. Subsequently, restart methods to improve their numerical stability have been discussed. Chapter 3 concentrates on communications applications. In particular, it considers the application of the fast recursive least squares adaptation algorithm to decision feedback equalization of the HF channel. Chapter 4 introduces a formulation for the decision feedback equalizer (DFE) different from the traditional. It also discusses adaption issues, in a RLS context, associated with this new DFE structure.

Chapter 5 includes concluding remarks and extensions for future work.

REFERENCES - 1

- [1] B. Widrow: "Adaptive Filters", in Aspects of Network and Systems Theory, ed. R. Kalman and N. DeClaris, New York, Holt Rinehart and Winston, 1971.
- [2] T. Kailath: "A View of Three Decades of Linear Filtering Theory", IEEE Trans. on Inform. Theory, Vol.IT-20, pp.145-181, March 1974.
- [3] B. Widrow, et al.: "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter", Proceedings of the IEEE, pp.1151-1162, August 1976.
- [4] D.J. Sakrison: "Stochastic Approximation: A Recursive Method for Solving Regression Problems", in Advances in Communication Theory, Vol.5, A.V. Balakrishnan, Ed., Academic Press, 1966.
- [5] A. Gersho, "Adaptive Equalization of Highly Dispersive Channels", BSTJ, Vol.48, pp.55-70, January 1969.
- [6] J.G. Proakis: "Advances in Equalization for Intersymbol Interference" in Advances in Communication Systems, Vol.4, pp.123-198, Academic Press, 1975.
- [7] R.W. Lucky, J. Salz, E.J. Weldon: "Principles of Data Communication", McGraw-Hill Book Company, 1968.
- [8] M. Schwartz, W.R. Bennett and S. Stein: "Communication Systems and Techniques", McGraw-Hill Book Company, 1966.
- [9] P. Monsen: "Feedback Equalization for Fading Dispersive Channels", IEEE Trans. on Inform. Theory, Vol.IT-17, pp.56-64, January 1971.
- [10] D. Godard: "Channel Equalization Using a Kalman Filter for Fast Data Transmission", IBM Journ. Research and Develop., pp.267-273, May 1974.
- [11] D.G. Messerschmitt: "A Geometric Theory of Intersymbol Interference: Part I", BSTJ, Vol.59, pp.1483-1518, November 1973.
- [12] D.D. Falconer and L. Ljung: "Application of Fast Kalman Estimation to Adaptive Equalization", IEEE Trans. on Commun., Vol.COM-26, pp.1439-1440, October 1978.
- [13] J. Makhoul: "A Class of All-Zero Lattice Digital Filters: Properties and Applications", IEEE Trans. on Acoust. Speech and Signal Processing, pp.304-314, August 1978.

- [14] G.C. Goodwin and R.L. Payne: "Dynamic System Identification: Experiment Design and Data Analysis", Academic Press, 1977.
- [15] D.D. Falconer: "Adaptive Filter Theory and Applications", Fourth Intern. Conf. on Analysis and Optimization of Systems, Versailles, France, December 1980.
- [16] B.D.A. Anderson, and J.B. Moore: "Optimal Filtering", Prentice-Hall, Inc., 1979.
- [17] J.E. Proakis: "Digital Communications", McGraw-Hill Book Company, 1983.
- [18] J.M. Cioffi: "Fast Transversal Filters for Communications Applications", Ph.D. Dissertation, Stanford University, 1984.
- [19] M.L. Honig and D.G. Messerschmitt: "Adaptive Filters: Structures, Algorithms and Applications", Kluwer Academic Publishers, Hingham, 1984.
- [20] L. Ljung and T. Söderström: "Theory and Practice of Recursive Identification", MIT Press, Cambridge, Massachusetts, 1983.
- [21] S. Haykin: "Introduction to Adaptive Filters", McMillan, 1984.
- [22] A.A. Giordano and F.M. Hsu: "Least-Squares Estimation With Applications to Digital Signal Processing", John Wiley and Sons, 1985.

TRACKING PROPERTIES AND STEADY-STATE PERFORMANCE
OF RLS ADAPTIVE FILTER ALGORITHMS

An area which is of strong current practical importance and research interest is the adaptive channel estimation and equalization of rapidly time-varying channels.

Adjustment algorithms for adaptive filtering derived from LS (Least Squares) cost functions are known to converge extremely fast and have excellent capabilities working in a time-varying environment. Although various simulation results confirming these facts can be found in the literature there is little theoretical work published, describing the steady-state performance characteristics of the RLS (Recursive Least Squares) adaptive filter.

It is known that all adaptive filters capable of adapting at real-time rates experience losses in performance because their adjustments are based on statistical averages taken with limited sample sizes [1]. For exponentially-windowed RLS algorithms these losses expressed through the excess MSE (Mean Square Error) are a result of two main sources of error. The first source of error is attributed to the exponential weighting of the squared error sequence and therefore to the exponential nature of the estimators used to estimate the correlation matrix and the cross-correlation vector [3], i.e., finite window effect. This error, which we shall call "estimation-noise" results in a misadjustment

of the coefficient vector of the adaptive filter from its optimal setting. An analogy to the previous situation is encountered with the LMS (Least Mean Square) adaptive filter because of the gradient noise [5]. The second source of error is associated with filtering nonstationary signals. This error which has been called "lag error" is caused by the attempt of the adaptive system to track variations of the input signal. At this point we should emphasize that finite-precision arithmetic also contributes to the excess MSE and in that sense could be considered as a third source of noise. More important though is the fact that roundoff errors trigger numerical instabilities posing potential problems in implementing the RLS adaptive algorithms.

The analysis of the stationary and nonstationary characteristics of the LMS algorithm can be found in the pioneering work of Widrow et al in [5]. In reference [4] the tracking ability of a wide class of adaptive signal processing algorithms has been studied. That work develops an upper bound on the squared error between the parameter vector being tracked and the value obtained by the algorithm. Also in references [6], [7] a more or less qualitative analysis of a preliminary experimental examination of the response of an adaptive lattice predictor and a Kalman estimator to nonstationary inputs has been presented respectively. Very recently the analysis of the behaviour of an orthogonalized LMS adaptive filter in a time varying environment appeared in [8]. A discussion of the steady-state operation limitations of RLS adaptive algorithms because of the finite precision arithmetic and possible solutions can be found in the work of Cioffi and Kailath [10] and Lin [9].

Finally, work more closely related to the subject of the present chapter and especially to the part dealing with the calculation of the "lag error" time constant has been presented by Ling and Proakis in [11].

In this chapter we present an attempt at a quantitative understanding of the steady-state performance characteristics of adaptive filters driven by RLS algorithms. A new more accurate expression for the "estimation-noise" than the one given in [3] and [11] has been derived. Furthermore, the general tracking ability and the explicit calculation of the excess MSE due to the lag effects for an assumed Markov channel provides the necessary information about the tradeoffs between speed of adaptation and steady-state performance. A number of simplifying assumptions are made in order to obtain simple results which give insight into the adaptation process. Finally, we describe new restart procedures which guarantee smooth re-initialization and steady-state operation of the adaptive algorithm without introducing any significant performance-degrading transients. These restart procedures appear to make the RLS or FRLS (Fast Recursive Least Squares) transversal filter algorithms viable for adaptive equalization or channel estimation of time-varying channels.

Throughout our analysis for the estimation and lag error noises we use the mathematically more tractable transversal filter realization, as opposed to a Lattice filter realization, without any loss of generality, since it has been shown, e.g., [12], that these algorithms perform similarly; neglecting errors due to finite word length, and they are all equally-fast converging.

The remainder of this chapter is organized as follows:

Section 2.1 is a brief review of RLS adaptive filters. This section cites the general concepts of RLS adaptive algorithms and establishes the relevant notation that will be used in the sequel. Section 2.2 derives a new formula for the excess error due to the "estimation-noise". Section 2.3 presents analytical results concerning the behaviour of the RLS adaptive filters in a time-varying environment. Section 2.4 discusses restart procedures for RLS transversal algorithms as a remedy to numerical roundoff errors. Section 2.5 presents experimental results of computer simulations that verify the developed theory. Finally, Section 2.6 is a brief conclusion of the chapter.

2.1 Recursive Least Squares Algorithms

In this section we describe briefly the basic principles of the LS adaptation problem. The general adaptive filter is shown in Figure 2.1. The system is considered to be of the tapped delay line form of length N . The impulse response of the system is denoted by the $N \times 1$ complex vector $\underline{C}(n)$ (where n is the discrete time index). The adaptive system is acting on the input signal samples $x(n)$, while the other input $d(n)$ is the so-called "desired response". If at time n the adaptive filter has the N latest inputs stored denoted by the vector

$$\underline{X}(n) = (x(n)^*, x(n-1)^*, \dots, x(n-N+1)^*)^* \quad (2.1-1)$$

then the adaptive filter output at time n , $\underline{C}(n-1)^* \underline{X}(n)$, may differ from the ideal output $d(n)$ by an error $e(n)$

* The * denotes conjugate scalars and conjugate complex transposed vectors or matrices.

Fig.2.1 Adaptive Filtering System.

$$e(n) = d(n) - \underline{C}(n-1)^* \underline{X}(n) \quad (2.1-2)$$

The objective of exponentially-windowed LS algorithms is to generate that tap-coefficient vector $\underline{C}(n)$ at time n which minimizes the weighted cumulative squared error [12], [13],

$$J(n) = \sum_{k=0}^n \lambda^{n-k} |d(k) - \underline{C}(n)^* \underline{X}(k)|^2 \quad (2.1-3)$$

In the sense of this error minimization criterion then, the recursive least squares estimation algorithm makes the best possible use of all data $\{X(k), d(k)\}$ up to time n ; therefore in this sense it converges and tracks "as fast as possible". The parameter λ is some positive constant close to, but less than 1 used for exponential weighting of the past. Roughly speaking $1/(1-\lambda)$ represents the memory of the algorithm.

The minimizing vector is the solution of the discrete time, Wiener-Hopf equation

$$\underline{C}(n) = \hat{R}(n)^{-1} \hat{V}(n) \quad (2.1-4)$$

where

$$\hat{R}(n) = \sum_{k=0}^n \lambda^{n-k} \underline{X}(k) \underline{X}(k)^* \quad (2.1-5)$$

and

$$\hat{V}(n) = \sum_{k=0}^n \lambda^{n-k} \underline{X}(k) d(k)^* \quad (2.1-6)$$

It can be seen that for $\lambda=1$ and large n , $\frac{1}{n} \hat{R}(n)$ is a consistent estimate of the input signal vector autocorrelation matrix and $\frac{1}{n} \hat{V}(n)$ is a consistent estimate of the cross-correlation between the input

signal vector and the desired response [12].

At this point it should be noted that the solution of equation (2.1-4) could follow basically along one of two possible directions. The first procedure is based on the fact that equations (2.1-5) and (2.1-6) can be written recursively in time and therefore that the sequence $\underline{C}(n)$ obeys the following time update relation [12], [13].

$$\underline{C}(n) = \underline{C}(n-1) + \underline{K}(n) e(n)^* \quad (2.1-7)$$

where $\underline{K}(n)$ is the Kalman gain defined as

$$\underline{K}(n) = \hat{R}(n)^{-1} \underline{X}(n) \quad (2.1-8)$$

The inverse estimated correlation matrix $\hat{R}(n)^{-1}$ displayed in (2.1-8) greatly accelerates the adaptation process since it performs the appropriate orthogonalization of the autocorrelation matrix as a means to very fast adaptation [14]. This approach, i.e., equations (2.1-7) and (2.1-8) for solving (2.1-4) leads to the transversal adaptive filter structure. That is the adaptive filter model we are going to use throughout our analysis.

A second approach for solving (2.1-4) is by Lattice methods [15], [18]. In this method certain properties of the estimated autocorrelation matrix are used in order to obtain order-updated equations. We should emphasize that the RLS algorithms of Lattice or Transversal filter form perform the same minimization of equation (2.1-3) and therefore offer the same convergence and tracking ability. The difference lies in the manner and the complexity with which the minimization is achieved [12]

and the numerical stability properties as well.

Both approaches may be implemented with Fast Recursive Least Squares (FRLS) algorithms which exploit the shift relationship between successive input vectors $\underline{X}(n)$ to limit the computational complexity of the order of N [10], [12], [13], [15], [18].

2.2 RLS Adaptive Filters and the "Estimation-Noise"

During the adaptation process the filter is trying to match its coefficients $\underline{C}(n)$ with those of some unknown parameter vector. For instance, in a channel estimation situation this parameter vector will be the unknown impulse response to be identified while in the equalization case it will be the impulse response that minimizes the overall mean squared error.

A discussion of the LS estimation procedures and analysis of some of its statistical properties for the special case of $\lambda=1$ can be found in [10] and [20]. The input signal vector $\underline{X}(n)$ is assumed to be given up to time n in [20]. The desired response $d(n)$ is considered to be a stochastic process with mean value conditioned on $\underline{X}(n)$ a linear function of the parameter target vector $\underline{C}$, i.e.,

$$E\{d(n)/\underline{X}(n)\} = \underline{\hat{C}}^* \underline{X}(n) \quad (2.2-1)$$

Taking thus ensemble averages with respect to the random sequence $\{d(n)\}$, $\underline{C}(n)$ is proved to be the best linear unbiased estimator of $\underline{C}$. Furthermore, the covariance matrix of $\underline{C}(n) - \underline{\hat{C}}$ is given in [10], [20] as

$$\text{Cov} \{ (\underline{C}(n) - \hat{\underline{C}}) / \underline{X}(n) \} = \left(\sum_{k=0}^n \underline{X}(k) \underline{X}(k)^* \right)^{-1} \epsilon \quad (2.2-2)$$

where ϵ is the expected mean squared error conditional on $\underline{X}(n)$,

$$\epsilon = E \{ |d(n) - \hat{\underline{C}}^* \underline{X}(n)|^2 / \underline{X}(n) \} \quad (2.2-3)$$

The same expression for the above covariance matrix appears in [21] where Kalman filtering techniques for adaptation of an equalizer are used. Clearly the quantity $\frac{1}{n} \sum_{k=0}^n \underline{X}(k) \underline{X}(k)^*$ converges to the autocorrelation matrix of the input signal vector when n goes to infinity. Therefore asymptotically

$$\lim_{n \rightarrow \infty} \text{Cov}(\underline{C}(n) - \hat{\underline{C}} / \underline{X}(n)) = 0 \quad (2.2-4)$$

In other words, the estimator $\underline{C}(n)$ converges in probability to the vector $\hat{\underline{C}}$.

It can easily be seen that, if we relax the conditional expectations of (2.2-1) to (2.2-4) on $\underline{X}(n)$, for $\lambda=1$ and $n \rightarrow \infty$, LS adaptation algorithms realize the optimum Wiener Solution $\underline{C}_{\text{opt}}$. Unfortunately, things change drastically if one uses exponentially windowed RLS algorithms in order to track possible variations in the unknown channel. In that case the inconsistency of the exponential weighted estimators and generally any finite window estimators causes noise in the tap coefficients and makes them to be on the average misadjusted from their optimal values. This noise in the tap coefficient vector $\underline{C}(n)$ is the source of an excess error (the error above the MSE of the Wiener solution) namely; "estimation-noise" or "misadjustment-noise".

We recall [2], [5], [14] that the excess MSE is $E\{\Delta \underline{C}(n)^* \cdot R \cdot \Delta \underline{C}(n)\}$. Here R is a positive definite matrix with elements

$$r_{ij} = E\{x(n-i+1) x(n-j+1)^*\} \quad 1 \leq i, j \leq N \quad (2.2-5)$$

and $\Delta \underline{C}(n)$ is a $N \times 1$ dimensional tap-coefficient error vector, i.e.,

$$\Delta \underline{C}(n) = \underline{C}(n) - \underline{C}_{\text{opt}} \quad (2.2-6)$$

where $\underline{C}_{\text{opt}}$ is the solution of the following matrix equation

$$\underline{C}_{\text{opt}} = R^{-1} \underline{V} \quad (2.2-7)$$

and $\underline{V}$ denotes another $N \times 1$ dimensional vector with elements

$$v_i = E\{x(n-i+1)d(n)^*\} \quad (2.2-8)$$

Using the adaptation equation (2.1-7) along with the definition of the Kalman gain and the error sequence $e(n)$ we can easily show that:

$$\hat{R}(n) \underline{C}(n) = \lambda \hat{R}(n-1) \underline{C}(n-1) + \underline{X}(n) d(n)^* \quad (2.2-9)$$

Defining the optimum error at time n as

$$e_{\text{opt}}(n) \triangleq d(n) - \underline{C}_{\text{opt}}^* \underline{X}(n) \quad (2.2-10)$$

and using the fact that

$$\hat{R}(n) = \lambda \hat{R}(n-1) + \underline{X}(n) \underline{X}(n)^* \quad (2.2-11)$$

equation (2.2-9) yields the following recursive relation for the tap-coefficient error vector

$$\hat{R}(n) \Delta \underline{C}(n) = \lambda \hat{R}(n-1) \Delta \underline{C}(n-1) + \underline{X}(n) e_{\text{opt}}^*(n) \quad (2.2-12)$$

The solution to this equation is given by

$$\Delta \underline{C}(n) = \lambda^n \hat{R}(n)^{-1} \hat{R}(0) \Delta \underline{C}(0) + \hat{R}(n)^{-1} \sum_{k=1}^n \lambda^{n-k} \underline{X}(k) e_{\text{opt}}^*(k) \quad (2.2-13)$$

where $\Delta \underline{C}(0)$ and $\hat{R}(0)$ are iterations starting fixed quantities. In particular, for the exponential windowed RLS algorithm $\hat{R}(0)$ is a diagonal matrix such that

$$\hat{R}(0) = \delta I_{NN} \quad (2.2-14)$$

where I_{NN} indicates the $N \times N$ unit matrix and δ is a small constant which ensures positive definiteness of $\hat{R}(n)$ for all n .

Although our ultimate aim is an expression for the excess mean-square error, we shall begin with the average tap-coefficient error vector, a simpler quantity, but one which gives some insight on the initial convergence and the parameters involved in it. Taking expected values on both sides of (2.2-13) we get

$$E\{\Delta \underline{C}(n)\} = \lambda^n \delta E\{\hat{R}(n)^{-1}\} \Delta \underline{C}(0) + E\{\hat{R}(n)^{-1} \sum_{k=1}^n \lambda^{n-k} \underline{X}(k) e_{\text{opt}}^*(k)\} \quad (2.2-15)$$

The optimum error sequence $e_{\text{opt}}(n)$ can be shown to be orthogonal with $\underline{X}(n)$. We also assume that $\hat{R}(n)$ is independent of $\underline{X}(k)$ and $e_{\text{opt}}^*(k)$ [†]. Therefore the second term in (2.2-15) is approximately zero, leaving only the first term as the dominant one during initial convergence. The fact that $\lambda^n \rightarrow 0$ for $|\lambda| < 1$ (for the case of $\lambda = 1$, see [10]) and the

[†] $\hat{R}(n)$ (estimate of the autocorrelation matrix) is a function of all past history of the input signal vector.

asymptotic stationarity of the estimated correlation matrix $\hat{R}(n)$, indicate that the expected tap-coefficient vector converges always to zero, as long as $\hat{R}(n)$ retains its positive definite nature. We also observe that RLS tap-coefficient convergence is independent of the eigenvalue distribution. Only the product $E\{\hat{R}(n)^{-1} \hat{R}(0)\}$ which in a worst case scenario has a norm on the same order as the product of δ and the inverse of the smallest eigenvalue of R affects in a proportional manner initial convergence. An excellent discussion on the choice of δ can be found in [10]. The above mentioned behaviour is in contrast to the initial convergence of the LMS algorithm. In the latter case the eigenvalues themselves form the convergence time constants and determine the value of the adaptation size μ which ensures stability.

In the sequel the asymptotic behaviour of the mean square error will be studied. In order to facilitate mathematical treatment $\Delta C(n-1)$ and $X(n)$ are assumed independent of each other. If $E\{\cdot\}$ is defined as the mean value over all possible input sequences $\{X(n)\}$ and $\{d(n)\}$, then our major concern will be directed to the steady-state analysis of the expected excess noise in RLS adaptive filters denoted by $\langle E\{\Delta C(n)^* R \Delta C(n)\} \rangle$. This quantity represents the ensemble mean value of $E\{\Delta C(n)^* R \Delta C(n)\}$ averaged over $\Delta C(n)$. In that sense $E\{\cdot\}$ can be interpreted as the mean value conditional to a particular $\Delta C(n)$.

Let q_n be defined as

$$q_n \triangleq E\{\Delta C(n)^* R \Delta C(n)\}$$

(2.2-16)

In Appendix A-2 we show that at the steady-state $\langle q_n \rangle$ satisfies the following relation

$$\langle q_n \rangle = \lambda^2 \langle q_{n-1} \rangle + (1-\lambda)^2 \text{trace} \{ I_{NN} + E\{P(n)^2\} \} \epsilon_{\text{opt}} \quad (2.2-17)$$

where $P(n)$ is a zero mean fluctuation matrix. The matrix $S = E\{P(n)^2\}$ has entries which depend on some fourth order statistics of the received signal (see equation (B-2.10)). Under these conditions the asymptotic value of $\langle q_n \rangle$ will be

$$\epsilon_{\text{est}} \frac{\Delta \langle q_\infty \rangle}{1+\lambda} = \frac{1-\lambda}{1+\lambda} \text{trace} \{ I_{NN} + E\{P(n)^2\} \} \epsilon_{\text{opt}} \quad (2.2-18)$$

The above formula is a general expression for the excess MSE due to "estimation-noise in exponential weighted RLS adaptive algorithms".

A simplification of equation (2.2-18) which retains the essence and emphasizes more the quantitative characterization of the "estimation-noise", can be achieved if we further assume that in the steady-state the diagonal elements only of R and $\tilde{R}(n)$ contribute to the estimation error (see Appendix B-2, equation (B-2.8)). Thus asymptotically S will be of diagonal form with elements

$$s_{ij} = \begin{cases} \frac{1-\lambda}{1+\lambda} \frac{\sigma_{|x|}^2}{(\sigma_x^2)^2} = \frac{1-\lambda}{1+\lambda} \gamma & i = j \\ 0 & i \neq j \end{cases} \quad (2.2-19)$$

$$\text{where } \gamma = \frac{\sigma_{|x|}^2}{(\sigma_x^2)^2} \quad (2.2-20a)$$

$$\text{and } \sigma_{|x|}^2 \triangleq \text{var}(|x(n)|^2) \quad (2.2-20b)$$

$$(\sigma_x^2)^2 \triangleq (\text{var}(x(n)))^2 \quad (2.2-20c)$$

The statistical parameter γ can be computed in a straightforward manner for given channel and signal statistics. In that case, it can easily be shown that the steady-state excess mean-square error will be

$$\epsilon_{\text{est}} = \frac{1-\lambda}{1+\lambda} \left(1 + \frac{1-\lambda}{1+\lambda} \gamma\right) N \epsilon_{\text{opt}} \quad (2.2-21)$$

This formula is an expression for the "estimation-noise", it gives as a quantitative insight of the excess MSE due to exponential nature of the RLS estimators. If we define the "misadjustment" as the dimensionless ratio of the excess MSE to the minimum MSE [5], i.e.,

$$M = \frac{\text{average excess MSE}}{\epsilon_{\text{opt}}} \quad (2.2-22)$$

then the misadjustment for LS algorithms will be

$$M = \frac{1-\lambda}{1+\lambda} \left(1 + \frac{1-\lambda}{1+\lambda} \gamma\right) N \quad (2.2-23)$$

The misadjustment is a useful measure of the cost of the adaptability [5]. A value of $M=0.15$ means that the adaptive filter has a MSE only 15% greater than ϵ_{opt} .

For values of λ very close to 1 and γ reasonably small (for instance if the input samples are assumed to be Gaussian distributed then $\gamma=2$) it can be shown easily that

$$1 + \frac{1-\lambda}{1+\lambda} \gamma \approx 1 \quad (2.2-24)$$

Therefore (2.2-21) becomes

$$\epsilon_{\text{est}} \approx \frac{1-\lambda}{1+\lambda} N \epsilon_{\text{opt}} \quad (2.2-25)$$

and the corresponding misadjustment formula will be

$$M = \frac{1-\lambda}{1+\lambda} N \quad (2.2-26)$$

Since we can approximate $1+\lambda=2$ equation (2.2-26) leads to the following "rule of thumb": The misadjustment due to estimation noise equals the half number of tap-coefficients multiplied by the inverse of the system memory. The result of equation (2.2-25) had first appeared in [3]. A different proof was also given in [11]. That expression is less general than the ones we derived (equations (2.2-18) and (2.2-21)) because its applicability is restricted to values of λ very close to 1. The key assumption in its derivation is (2.2-24), which essentially is equivalent to modelling the estimation auto-correlation matrix as follows

$$\hat{R}(n)^{-1} R = (1-\lambda) I_{NN} \text{ as } n \rightarrow \infty \quad (2.2-27)$$

This model is a degenerate form of the relation (B-2.4) in Appendix B-2, where the asymptotic variance of the elements of the perturbation matrix is considered to be very small or zero. Apparently as we decrease λ the above assumption can no longer be justified. Thus the influence of the statistical parameter γ becomes noticeable.

In any case, we see that as the memory of the system decreases the misadjustment increases. In general, fast adaptation leads to a more noisy adaptation process. In a stationary environment the best steady-state performance results from slow adaptation which corresponds to $\lambda=1$.

Concluding we should emphasize again that the formulas (2.2-18), and (2.2-21) we derived for the excess MSE indicate that the level of the estimation noise does not depend solely on the nature of the "windowing" but on some fourth-order statistics of the input signal as well. The latter is more noticeable in fast adaptation where λ may take values as low as 0.9.

2.3 RLS Adaptive Filter Response to Time-Varying Inputs

2.3.1 Tracking Behaviour

Recursive Least Squares algorithms are known to have much faster tracking capabilities than any other of the existing algorithms. Although the usefulness of fast adaptation for tracking rapidly time-varying systems has been demonstrated by many simulations and in various applications, there is little theoretical work published on this subject. Our aim therefore, is to shed some light on the tracking behavior of the RLS adaptive algorithms.

The problem we examine here involves adaptive filtering of an unknown time-varying system using a RLS transversal filter of length N . As a result of the time variations of the environment the optimum solution will be a N -dimensional vector dependent on the discrete-time index, n , $C_{opt}(n)$, reflecting variations of the autocorrelation matrix and cross-correlation vector respectively.

The degree of mathematical difficulty in analyzing nonstationary cases is directly related to the particular application. In a general situation such as channel equalization, both the autocorrelation matrix $R(n)$ and the cross-correlation vector $\underline{V}(n)$ will be functions of time. In the system identification case though, only the cross-correlation vector will depend on time.

For a RLS transversal filter the adaptation equation (2.1-7) yields

$$\begin{aligned}\underline{C}(n) &= \underline{C}(n-1) + \underline{K}(n) e(n)^* \\ &= \underline{C}(n-1) + \hat{R}(n)^{-1} \underline{X}(n) \underline{X}(n)^* (\underline{C}_{opt}(n-1) - \underline{C}(n-1)) + \underline{K}(n) e_{opt}(n)^*\end{aligned}\quad (2.3-1)$$

where we have defined

$$d(n) \triangleq \underline{C}_{opt}(n-1)^* \underline{X}(n) + e_{opt}(n) \quad (2.3-2)$$

and $e_{opt}(n)$ represents the minimum (in a mean square sense) error at time n . In studying tracking behaviour we may exclude the influence of the estimation noise, since the deviation of $E\{\underline{C}(n)\}$ from $\underline{C}_{opt}(n)$ determines the response of the adaptive algorithm to the nonstationarities of the environment. Taking expected values on both sides of (2.3-1) and assuming that $\hat{R}(n)$ is independent of $\underline{X}(n)$ and $e_{opt}(n)$, using also the fact [19] that $\underline{X}(n)$ and $e_{opt}(n)$ are orthogonal we get

$$E\{\underline{C}(n)\} = E\{\underline{C}(n-1)\} + E\{\hat{R}(n)^{-1} \underline{X}(n) \underline{X}(n)^*\} (\underline{C}_{opt}(n-1) - E\{\underline{C}(n-1)\}) \quad (2.3-3)$$

In Appendix C-2 we show that if the variations of the environment are slow with respect to the memory of the adaptive algorithm then independently of the adaptive system structure (i.e., equalizer or channel estimator) we approximately have

$$E\{R(n)^{-1} \underline{X}(n) \underline{X}(n)^* \} = (1-\lambda) I_{NN} \quad (2.3-4)$$

Substituting (2.3-4) into (2.3-3) we finally get

$$E\{\underline{C}(n)\} = E\{\underline{C}(n-1)\} + (1-\lambda) [\underline{C}_{opt}(n-1) - E\{\underline{C}(n-1)\}] \quad (2.3-5)$$

If $\Delta \underline{C}(z)$ is the z-transform of the lag error vector $\Delta \underline{C}(n) \triangleq E\{\underline{C}(n)\} - \underline{C}_{opt}(n)$ then

$$\Delta \underline{C}(z) = \frac{z^{-1} - 1}{1 - \lambda z^{-1}} \underline{C}_{opt}(z) \quad (2.3-6)$$

The above transfer function has a zero at $z=1$ and a pole at $z=\lambda$.

Thus under general time varying conditions the lag error vector is a geometric converging process. The time constant of the process given by

$$\tau_1 = \frac{1}{1-\lambda} \quad (2.3-7)$$

is the same for all coordinates i , $1 \leq i \leq N$. It is also independent of the spread of the eigenvalues of the autocorrelation matrix. This is in contrast to the LMS algorithm where the time constant is in general different for each component and corresponds to a particular eigenvalue of $R(n)$ [5]. The independence of the geometrically converging process from the eigenvalue distribution causes the tracking ability of the RLS algorithms to be at least as good as that of the LMS algorithm.

2.3.2 Calculation of the "Excess Lag" MSE

In this section we are going to derive an expression for the "excess lag" MSE concerning a particular application, namely estimation or identification of a time varying impulse response. This description exactly fits real applications as echo-cancellation and adaptive Maximum-Likelihood-Sequence-Estimation (MLSE). In the latter application an adaptive channel estimator is necessary in combination with Forney's receiver in order to communicate over unknown time-varying dispersive channels [27].

Identification of an unknown time-variance system is the simplest and mathematically the most tractable situation one may encounter. The simplicity of this particular nonstationary adaptation lies in the fact that the autocorrelation matrix of the stationary input signal is constant.

Widrow et al. in [5] studied the LMS adaptive filter as a means of modelling an unknown time-varying system. Although the nonstationary situations they examined were highly simplified, they retained the essence of the problem that is common to more complicated and realistic applications [5]. Based on concepts and techniques developed in this work our aim is to extend their results to another class of filters namely, the RLS adaptive filters.

The model for the unknown system to be used in this investigation is a tapped delay line filter structure with time variant coefficients and length N . The time dependent frequency response of the model will be

$$H(f,n) = \sum_{i=1}^N c_{opt,i}(n) \exp(-j2\pi if) \quad (2.3-8)$$

Thus the $N \times 1$ dimensional complex vector

$$\underline{C}_{opt}(n) = (c_{opt,1}(n), c_{opt,2}(n), \dots, c_{opt,N}(n)) \quad (2.3-9)$$

represents the true time varying unknown impulse response at time n , to be tracked by the RLS adaptive filter. It is assumed furthermore that the unknown coefficients obey a first-order Markov difference equation, i.e.,

$$\underline{C}_{opt}(n) = \alpha \underline{C}_{opt}(n-1) + \underline{W}(n) \quad (2.3-10)$$

where $|\alpha| < 1$ and $\underline{W}(n)$ is a zero mean Gaussian random vector with covariance matrix

$$E\{\underline{W}(k) \underline{W}(l)^*\} = \sigma_W^2 \cdot I_{NN} \delta_{k,l} \quad (2.3-11)$$

for all k, l where $\delta_{k,l}$ is the Konecker delta, which is 1 for $k=l$ and 0 otherwise, and I_{NN} denotes the $N \times N$ unit matrix. At the output of the system a zero-mean white Gaussian noise $v(n)$, with variance σ_v^2 , is added. Under these conditions the desired response sequence will be

$$d(n) = \underline{C}_{opt}(n-1)^* \underline{X}(n) + v(n) \quad (2.3-12)$$

Such a model has been used in [5] and [8] for the study of the nonstationary characteristics of the LMS and orthogonalized LMS algorithm respectively.

At this point it is worthwhile to note that, since the unknown system is considered to be of the same length as the adaptive filter used for its identification, it is clear that only the zero mean additive noise $v(n)$ will prevent a perfect match between them. Therefore the minimum MSE that can be achieved is

$$\epsilon_{opt} = E\{|v(n)|^2\} = \sigma_v^2 \quad (2.3-13)$$

Apparently the use of a longer or shorter adaptive filter than the channel impulse response will cause an increase to the minimum MSE. This increase reflects the truncation error involved in the first case and the noise picked up by the excess tap-coefficients in the latter one.

Clearly a typical region of operation for the adaptive algorithm will be $\frac{1}{1-\alpha} \gg \frac{1}{1-\lambda}$. Therefore in cases of interest, α will be very close to but less than 1. Using this assumption along with (2.3-6), (2.3-10) and (2.3-11) the covariance matrix $\text{cov}\{\Delta C(n)\}$ of the lag error vector can be calculated as

$$\text{cov}\{\Delta C(n)\} = \sum_{j=0}^{\infty} (\lambda^2)^j \sigma_w^2 I_{NN} = \frac{1}{1-\lambda^2} \sigma_w^2 I_{NN} \quad (2.3-14)$$

We observe that the lag error covariance matrix and subsequently the excess lag MSE, ϵ_{lag} , are geometrical converging processes with time constant $\frac{1}{1-\lambda^2}$. Furthermore, since in most practical situations λ usually takes values no lower than 0.9 we can use the approximation

$$\frac{1}{1-\lambda^2} \approx \frac{1}{2(1-\lambda)} \quad (2.3-15)$$

The excess MSE due to the lag, ϵ_{lag} , may be expressed as

$$\epsilon_{\text{lag}} = E\{\Delta \underline{C}(n)^* R \Delta \underline{C}(n)\} \quad (2.3-16)$$

Therefore using the fact that

$$E\{\Delta \underline{C}(n)^* R \Delta \underline{C}(n)\} = E\{\text{tr}(R \Delta \underline{C}(n) \Delta \underline{C}(n)^*)\} \quad (2.3-17)$$

where $\text{tr}(\cdot)$ stands for the trace of the matrix in the brackets, we get

$$\epsilon_{\text{lag}} = \frac{\sigma_w^2}{2(1-\lambda)} \text{tr}(R) \quad (2.3-18)$$

The misadjustment due to the lag error will be

$$M = \frac{\epsilon_{\text{lag}}}{\epsilon_{\text{opt}}} = \frac{1}{2(1-\lambda)} \frac{\sigma_w^2}{\sigma_v^2} \text{tr}(R) \quad (2.3-19)$$

The expression given by (2.3-18) indicates that the excess MSE due to the lag, ϵ_{lag} , is directly related to the variance of the source of non-stationarity, to the power of the input signal and to the memory of the RLS algorithm as well. As λ approaches 1 the lag error approaches infinity. The extreme situation of $\lambda=1$ corresponds to an equal weighting of all past information in calculating an updated coefficient vector. For $\lambda < 1$ the past is attenuated geometrically and therefore the algorithm can track variations of the input statistics.

The total excess MSE is the sum of the excess MSE due to estimation noise and due to lag. Combining (2.2-25) and (2.3-18) we get

$$\epsilon_{\text{tot}} = \frac{1-\lambda}{1+\lambda} N \epsilon_{\text{opt}} + \frac{1}{2(1-\lambda)} \text{tr}(R) \sigma_W^2 \quad (2.3-20)$$

The formula derived for the total excess MSE can be optimized with respect to the parameter λ . Differentiating (2.3-20) and setting the result equal to zero, also noticing that

$$N \sigma_x^2 = \text{tr}(R) \quad (2.3-21)$$

we get

$$\lambda_{\text{opt}} = \frac{1-\beta}{1+\beta} \quad (2.3-22)$$

where

$$\beta = \left(\frac{\sigma_x^2 \sigma_W^2}{4\epsilon_{\text{opt}}} \right)^{1/2} \quad (2.3-23)$$

with the constraint $\beta < 1$. It is interesting to note that the optimum λ_{opt} is independent of the length of the adaptive filter. It depends on the statistics of the input signal, on the statistics of the nonstationarity driving noise and on the minimum MSE as well. In Figure 2.2 the total excess MSE has been plotted for a hypothetical RLS adaptive filter of length $N=10$ and $\epsilon_{\text{opt}}=10^{-3}$. The two curves shown in Figure 2.2 correspond to two different levels of σ_W^2 , i.e., $\sigma_W^2 = 10^{-5}$ and $\sigma_W^2 = 10^{-6}$. It appears that the higher the level of the σ_W^2 the smaller the λ_{opt} should be.

It is of great importance to compare the above results with those derived by Widrow et al. for the LMS algorithm in [5]. We recall that for the LMS adaptive filter the time constant $\tau_{1,\text{LMS}}$ and the total

Fig.2.2. Total excess MSE, ϵ_{tot} , versus exponential windowing factor λ .

excess MSE are given by

$$\tau_{i,LMS} = \frac{1}{2\mu\rho_i} \quad 1 \leq i \leq N \quad (2.3-24)$$

and

$$\epsilon_{tot,LMS} = \mu \text{tr}(R) \epsilon_{opt} + \frac{N\sigma_w^2}{4\mu} \quad (2.3-25)$$

respectively, where μ is the adaptation step size and ρ_i is the eigenvalue corresponding to the i th component of the tap-coefficient vector. We observe that if all eigenvalues of R are equal and $\text{tr}(R) = N$, then by choosing $\mu = \frac{1-\lambda}{2}$ the time constants and the total excess MSE corresponding to the LMS and RLS adaptive algorithm become equal.[†] Hence, one would expect the steady-state performance of the two algorithms for white inputs to be exactly the same at least when the unknown coefficients obey a first order Markov stochastic difference equation. In many cases the value of μ which predicts similar tracking behavior with that of the RLS algorithm, might exceed the critical value for a stable operation. For a LMS adaptive algorithm with white inputs such that $\text{tr}(R) = N$, it can be shown [16], [35] that the step size μ must always satisfy the following inequalities:

$$0 < \mu < \frac{1}{N} \quad (2.3-26)$$

For example, in the hypothetical situation we discussed previously, we have seen that if $\sigma_w^2 = 10^{-5}$, then $\lambda_{opt} = .904$. This value suggests $\mu = .048$.

[†] Equality holds also for the covariance of the lag error vector, $\text{cov}\{\Delta \underline{C}(n)\}$, of the two algorithms.

If an adaptive filter with length $N \geq 21$ is to be used, then obviously the optimum value of μ which gives the same tracking performance violates the stability bound established in (2.3-26). When the maximum stable value of the step size is inserted instead then there is a significant tracking advantage of the RLS algorithms. Thus, for relatively long filters used in time-varying environments, RLS algorithms yield superior tracking performance. However, for fairly short system-identification filters, where the LMS step size μ which guarantees similar tracking behaviour does not violate the stability bound of equation (2.3-26), there is no tracking advantage in using the more complex RLS algorithm.

2.4 Finite Precision Arithmetic Considerations

2.4.1 Steady-State Performance

A crucial factor in addition to the tracking ability and complexity of an adaptive filter algorithm is its numerical stability. Numerical instabilities or finite precision problems pose potential problems for any recursive least squares adaptation algorithm transversal or lattice, which minimizes (2.1-3) since according to (2.1-3) this minimization explicitly or implicitly forms and inverts a N by N matrix whose elements are found by accumulating weighted sums of all previous filter inputs. Previous simulation studies have reported that the transversal FRLS adaptation algorithm tend to become unstable (the coefficients "blew up"), especially for λ less than unity (fast adaptation) [10], [12]. The same phenomenon has also been observed with the much more computation-intensive RLS algorithm.

Preliminary simulations of the FRLS adaptation algorithm used for identification and Decision Feedback Equalization (DFE), on real HF recorded data verified the previously discussed behaviour. Eventually the algorithm became unstable, and the tap-coefficient assumed unreasonable values. Using single precision on a 36-bit Honeywell Level 66 Computer, (28-bit mantissa) a FRLS 10 tap-coefficient adaptive channel estimation algorithm typically became unstable after about 800 symbol intervals with $\lambda=0.93$. By using double precision, the onset of instability was delayed to about 2200 symbol intervals. A single-precision FRLS DFE with 30 $T/2$ -spaced forward tap coefficients and 10 feedback coefficients also went unstable, after about 700 symbol intervals with $\lambda=0.97$. Decreasing the value of the exponential parameter to $\lambda=0.93$ the DFE typically became unstable after about 300 symbol intervals. It is interesting to note the accelerating influence of small values of λ on the onset of numerical instabilities.

The most stable reported recursive least squares transversal adaptive filter is the "square-root Kalman" filter [22], but even in this case, it was considered necessary to re-initialize certain variables of the algorithm about once every 100 symbol intervals to avoid instabilities due to roundoff errors. It is worth noting that the relatively slow converging simple LMS adaptive transversal filter does not fall prey to numerical stability resulting from roundoff error. However, digital implementation of the LMS as a fractionally spaced equalizer must be carefully designed to avoid problems due to bias in some types of digital arithmetic [23].

Among lattice adaptive filters the FRLS lattice has been found to exhibit better numerical stability than its FRLS transversal counterpart [12]. Still more stable is a normalized version of the FRLS lattice [24], which unfortunately is more computation-intensive, requiring square roots as well as multiplications and divisions. We should emphasize that adaptive lattice algorithms all require many more divisions than do the transversal filter algorithms; thus their complexity if implemented with most existing or contemplated digital signal processing devices is even greater.

It is unfortunate that the FRLS or "fast-Kalman" adaptation algorithm, which is the least computation-intensive of all the equally-fast-converging RLS and FRLS algorithms is also the most numerically unstable. A recent study of this type of algorithm suggests modifications to it which may improve its numerical stability and even decrease its complexity somewhat [10].

Our remedy for numerical stability problems of FRLS transversal filter is periodic re-initialization. We describe new restart procedures which guarantee smooth re-initialization of the adaptive algorithm without introducing any significant performance-degrading transients. These restart procedures appear to make the FRLS transversal filter algorithm viable for adaptive equalization or channel estimation of time-varying channels. The new restart procedures have some essential differences from those reported in [9] and [10]. We shall illuminate these differences in the next paragraph.

2.4.2 Periodic Restart for FRLS Algorithms

To avoid build-up of roundoff errors with time we propose interrupting and restarting the FRLS with critical internal variables cleared to zero, at periodic intervals, say every n_0 symbol intervals. Immediately following such a restart at time n_0 , a simple LMS adaptation algorithm, which has been initialized with the tap coefficient vector $\underline{C}(n_0-1)$ that the FRLS algorithm has obtained just before the restart was initiated, provides the filter output temporarily until the restarted FRLS takes over again after a short time.

FRLS methods have been developed which, aided by the LMS algorithm's tap coefficients, converge faster during restart than the "unaided" FRLS algorithm does during initial start-up at $n=0$. The transition period is so short (N to $1.5N$ time units) and smooth that the momentary reliance on the slower adapting LMS algorithm causes little or no performance degradation.

It can be shown [10] that following a restart at time n_0T the adaptive FRLS algorithm minimizes for $n \geq n_0$

$$J(n) = \sum_{k=n_0}^n \lambda^{n-k} |d(k) - \underline{C}(n) * \underline{X}(k)|^2 + \delta \lambda^{n-n_0} (\underline{C}(n) - \underline{C}(n_0))^* \Lambda_N (\underline{C}(n) - \underline{C}(n_0)) \quad (2.4-1)$$

where $\underline{C}(n_0)$ is the value of the tap-coefficients updated just prior to restart and $\delta (\delta > 0)$ is a factor used to weight the effect of the initial

+ Λ_N is the $N \times N$ Diagonal matrix $\Lambda_N = \text{Diag}(1, \lambda^{-1}, \lambda^{-2}, \dots, \lambda^{-N+1})$.

condition upon the cost criterion. In that sense $\underline{C}(n_0)$ can be set to zero or not. All other relevant internal variables will be initialized as during the start-up at time $n=0$. The input vector $\underline{\tilde{X}}(n)$ seen by the FRLS algorithm for adaptation purposes will have components

$$\underline{\tilde{X}}(n) = \begin{cases} x(n) & n > n_0 \\ 0 & n \leq n_0 \end{cases} \quad (2.4-2)$$

Thus for instance, in the simple situation of channel estimation where the new components entering the adaptive filter are complex scalars, the filter output at times $n_0 < n \leq n_0 + N$ is

$$\underline{C}(n)^* \underline{\tilde{X}}(n) \quad (2.4-3)$$

where

$$\underline{\tilde{X}}(n)^T = (x(n), x(n-1), \dots, x(n_0+1), 0, 0, \dots, 0) \quad (2.4-4)$$

is a $N \times 1$ dimensional vector.

The key idea behind the restarting methods is to remove from the desired response $\{d(n)\}$ the influence of those components of the input vector $\underline{X}(n)$ which are not available to the restarted FRLS. This modification on the desired response will mitigate the effect of the truncation of the input vector $\underline{X}(n)$ on the tracking ability of the algorithm. This is the major feature distinguishing our restart procedures from those of [9], [10].

To see how the $\{d(n)\}$ should be modified we note that if there has been no restart just prior to time nT , then we could express $d(n)$ as

$$d(n) = \underline{C}'(n) \underline{X}(n) + u(n) \quad (2.4-5)$$

where $u(n)$ is an error due to noise and random data, and where $\underline{C}(n)$ is the value $\underline{C}(n)$ would have in the absence of restart. Now define the N -dimensional vector

$$\underline{X}(n)^{(-)} \triangleq (0, 0, \dots, 0, x(n_0), x(n_0-1), \dots, x(n-N+1))^T \quad (2.4-6)$$

which leads to

$$\underline{X}(n) = \underline{X}(n) + \underline{X}(n)^{(-)} \quad (2.4-7)$$

i.e., $\underline{X}(n)^{(-)}$ contains only the components of $\underline{X}(n)$ which occur prior to $n=n_0$. Thus

$$d(n) = \underline{C}(n)' * \underline{X}(n)^{(-)} = \underline{C}(n)' * \underline{X}(n) + u(n) \quad (2.4-8)$$

Since the restarted FRLS algorithm only has the vectors $\underline{X}(n)$ available, and since the adapted tap coefficient vector $\underline{C}(n)$ should ideally approach $\underline{C}(n)'$, equation (2.4-8) suggests that during the restart period, the desired outputs used for FRLS adaptation should be

$$\tilde{d}(n) = d(n) - \underline{C}_{aux}(n)' * \underline{X}(n)^{(-)} \quad (2.4-9)$$

where $\underline{C}_{aux}(n)$ is the current set of tap coefficients from the auxiliary LMS algorithm. Note that when n exceeds n_0 by a number of symbol intervals equal to the filter's memory, $\underline{X}(n)^{(-)}$ is zero and $\tilde{d}(n) = d(n)$.

Another type of modification also suggested by (2.4-8) is the following:

$$d(n) = \begin{cases} c_{aux}(n) & \text{During restart interval} \\ d(n) & \text{After restart interval} \end{cases} \quad (2.4-10)$$

The two suggested modifications on the desired response along with the fact that $\underline{C}(n)$ can be set to zero or not at the initialization of the restart, give four possible variations of the periodic restart procedures. These procedures are based on the modified desired response methods we described previously and all rely on an auxiliary LMS adaptation algorithm.

Restart Procedure 1.

The FRLS adaptive filter tap coefficients $\underline{C}(n)$ were not reinitialized at the initiation of each restart ($n=n_0$). The desired filter output for this procedure was that of equation (2.4-9).

Restart Procedure 2.

In this restart procedure, the adaptive filter's tap coefficient vector $\underline{C}(n)$ was set to zero at the initialization of the restart. Again equation (2.4-9) was used to provide the desired response.

Restart Procedure 3.

Like procedure 1, $\underline{C}(n)$ were not re-initialized, but equation (2.4-10) was used for the desired output.

Restart Procedure 4.

In this procedure $\underline{C(n)}$ were set to all zeros. Equation (2.4-10) provides the desired response.

As suggested in [9], another way to continue the FRLS algorithm is to use the unwindowed versions which do not assume that $\underline{X(n)}$ is zero for negative time arguments, since $\underline{X(n)}$ is not zero when restarting in steady-state operation. In such a case modification of the desired response $d(n)$ is not necessary. Unfortunately, unwindowed algorithms are more computation intensive than the exponential windowed ones requiring approximately $3N$ more multiplications per iteration.

2.5- Simulation Results

The theory developed in the previous sections has been tested through computer simulations. The adaptive filters used for this investigation were driven by FRLS transversal algorithms as reformulated in [10].

For the evaluation of the estimation-noise formula a DFE (Decision Feedback Equalizer) was implemented. The second part of our work concerning the lag effects has been tested by running a channel estimator. In both cases double-precision on a 36-bit Honeywell Level 66 Computer (28-bit mantissa) has been used in order to delay as much as possible the onset of numerical instabilities.

Finally, the restarting methods have been tested in specific applications such as decision feedback equalization, using real recorded HF data [26].

2.5.1 Estimation-Noise

In this simulation we used a FRLS DFE with $N_1=8$ forward and $N_2=2$ feedback taps. The number of coefficients in the finite impulse response discrete channel is three with a maximum to minimum eigenvalue ratio of $\rho_{\max}/\rho_{\min}=56$. Binary signalling was also assumed, i.e., $d(n) = \pm 1$, while the level of the variance of the zero-mean additive white Gaussian noise was fixed to $\sigma_v^2 = 0.01$. For the given channel and signal statistics the statistical parameter γ has been computed as $\gamma = 0.8$.

Table 2.1 shows the results of our investigation comparing theoretical and experimental values as a function of the exponential windowing factor λ . The first column corresponds to the misadjustment M estimated over a sample of 1,600 iterations. Measurements were taken only after initial transients had faded away. Second and third columns correspond to the predicted theoretical values by using the new formula (e.g., (2.2-23)) and the approximate one (equation (2.2-26)) respectively.

Figure 2.3 illustrates the same results in a dB scale. The two solid curves represent theoretical values while the dashed curve represents experimental ones. The small vertical lines on top of each experimental point indicate the minimum and maximum range of variation of the estimated misadjustment M over a longer sample of 4,000 iterations.

The above simulation results indicate that the experimental misadjustment values are closer to the theoretical ones predicted by the new formula, equation (2.2-23). This conclusion is also in agreement to the experimental results reported in [11] for similar kinds of simulation.

Table 2.1

λ	$\hat{M}$	M, Equation (2.2-23)	M, Equation (2.2-26)
0.99	0.059	0.0504	0.0503
0.98	0.109	0.1018	0.1010
0.97	0.164	0.1541	0.1523
0.96	0.220	0.2074	0.2041
0.95	0.279	0.2616	0.2564
0.94	0.340	0.3169	0.3093
0.93	0.403	0.3739	0.3630
0.92	0.473	0.4305	0.4170

Fig. 2.3 Misadjustment versus λ

2.5.2 Lag Effects

In the second part of our investigation we simulated a FRLS transversal channel estimator. The number of tap coefficients used for the unknown time varying systems and the adaptive filter was $N=3$. As input signal we used a white sequence taking values ± 1 so that $R=I_{33}$. The additive noise variance was fixed to $\sigma_v^2 = \epsilon_{opt} = 0.01$.

Figures 2.4, 2.5, 2.6 and 2.7 summarize the results of the simulations, comparing theory and experiments for different values of the nonstationarity parameters λ and σ_w^2 . They illustrate the signal-to-noise ratio as a function of the adaptation parameter λ . The SNR on the vertical axis is the ratio of 1 (mean squared input sample) to the observed mean square value or the ensemble mean squared value of the error $e(n)$ between the output of the adaptive filter and the desired response.

Experimental values on those figures correspond to statistical time averages taken over 2,000 iterations in two different ways. Points on the dashed lines indicate time averaging with equal weighting of all past information while points on the dotted line indicate exponential weighting of the past information. The solid curves in all these figures represent theoretical results as have been predicted by using equation (2.3-20).

Experimental results came very close to what was theoretically expected with the exception of the case when λ approaches α . However, this outcome is not surprising since when λ approaches α , $\frac{1}{1-\alpha}$ is not significantly greater than $\frac{1}{1-\lambda}$ and therefore equation (2.3-18) is inaccurate.

Fig.2.4 SNR versus λ

Fig.2.5 SNR versus λ

Fig. 2.6 SNR versus λ

Fig. 2.7 SNR versus λ

Figures 2.8 and 2.9 show the evolution of the SNR with time for different levels of the parameter σ_w^2 . The solid straight line represents the theoretical predicted value. Again the agreement between theoretical and experimental values is apparent.

2.5.3 Restarting Procedures

The different restarting methods have been tested by implementing a decision feedback equalization (DFE) receiver. For this investigation we used complex baseband 2.4 kbps QPSK demodulated signal recorded from HF channels [25]. Preliminary simulations established the choice of the number of equalizer tap coefficients: $N_1=20$ T/2-spaced forward tap coefficients and $N_2=7$ feedback tap coefficients, where T is the duration of one symbol. The exponential weighting factor and the auxiliary LMS adaptation step size were set to $\lambda=0.97$ and $\mu=0.035$ respectively.

The four variations of the periodic restart procedure of the previous section for the FRLS algorithms were tried. All relied on an auxiliary LMS adaptation algorithm to provide the equalizer output during each restart interval.

Figure 2.10 shows an example of the application of restart methods initiated at $n=300$ th symbol intervals for HF channel designated MDA010. The SNR on the vertical axis is the ratio of 1 (the mean squared data symbol) to the observed mean squared value of the error $e(n)$ between the decision-making quantizer's input and its output (decision-directed mode). The averaging of the mean squared error (MSE) at the n th interval is as follows:

Fig. 2.8 SNR versus time

Fig. 2.9 SNR versus time

$$\text{MSE}(n) = 0.9 \text{MSE}(n-1) + 0.1|e(n)|^2 \quad (2.5-1)$$

and the SNR at this time is

$$\text{SNR}(n) = \frac{1}{\text{MSE}(n)} \text{ (dB)} \quad (2.5-2)$$

In the case shown in Figure 2.10 the restart interval is 30 symbol intervals i.e., approximately equal to the length of the adaptive filter. During this period the useful equalizer output (the input to the decision-making quantizer) is taken from an auxiliary DFE driven by an LMS algorithm, whose tap-coefficients were initialized to the tap-coefficients of the FRLS DFE at the beginning of the restart period. In Figure 2.11 the SNR obtained from this LMS filter is shown by the solid line between $n=300$ and $n=330$. During the 30-symbol interval restart interval, the FRLS algorithm adapted from an initial state restarting at $n=300$. After $n=330$ the FRLS-adapting DFE took over from LMS-adapting DFE. Figure 2.10 shows the SNR for the four restart procedures beyond $n=330$. In this example, Restart Procedure 1 gave slightly better performance than Restart Procedure 2, which in turn was somewhat better than procedures 3 and 4. At this point we should emphasize that the SNR's big dip just after the FRLS DFE took over from the auxiliary LMS DFE (approximately at $n=335$), is irrelevant to the restarting procedure and it has been caused by a sudden fade. This is clear in Figure 2.12, where the restart interval has been increased to 40 data symbol intervals and thus the SNR at $n=335$ has been calculated using the output of the auxiliary LMS DFE.

Fig.2.10 Examples of restart procedures for FRLS DFE
(channel MDA010) Restart interval = 30

Fig.2.11 SNR for FRLS algorithms for Channel MDA010 during the restarting interval. (Restart at 300) Restart interval = 30

Fig. 2.12 Example of restart procedures for FRLS DFE
(channel MDA010) Restart interval = 40

The differences among these restart methods in the same example are emphasized in Figure 2.11 which shows the measured SNR between $n=300$ and $n=330$, where the SNR is now defined as the reciprocal of the MSE, defined as in equation (2.5-2), for the still restarting FRLS algorithm, instead of the LMS algorithm. The MSE was initialized to zero at $n=300$, which accounts for the large values of SNR near $n=300$. It is clear that Restart Procedures 1 and 3 in which the equalizer tap coefficients were not zeroed, caused little disruption to the FRLS adaptation during the restart period, while the procedures involving zeroed coefficients required at least 40 iterations to restore a reasonably low mean square error.

On the basis of simulation results like these we conclude that there is no noticeable disruption as a result of the periodically initiated restarts.

2.6 Summary

In this chapter we have been concerned with the steady-state performance characteristics of the RLS adaptive algorithms. We have examined quantitatively the estimation-noise inherent in all exponential windowed RLS algorithms as well as their tracking ability in a time varying environment. We have also studied the steady-state performance limitations due to roundoff noise effects.

We have shown that the level of the estimation-noise does not depend only on the nature of the "windowing" but also on some fourth-

order statistics of the input signal. The latter is more noticeable in fast adaptation where λ may take values as low as 0.9.

Additionally, it has been shown that the mechanism associated with the tracking ability of the RLS is a geometrical converging process. Furthermore, independent of the kind of the nonstationary model and the adaptive filter application (e.g., equalizer or channel estimator) the time constant of this process is the same for all components of the tap-coefficient vector and equal to the memory of the algorithm. It is also independent of the eigenvalues of the autocorrelation matrix. In that sense the tracking ability of the RLS algorithms is superior than that of the LMS algorithm for the case of disparate eigenvalues.

Finally, assuming a first-order Markov time-varying model, the excess lag MSE for a channel estimator has been calculated. For this simple nonstationary situation where there is direct control on the eigenvalues of the constant autocorrelation matrix (all eigenvalues can be equal) it was shown that the simple LMS tracks as well as any of the more complex RLS algorithms provided that the judiciously chosen step size does not violate the stability bound. Tradeoffs involved in choosing the optimum exponential weighting parameter λ_{opt} in order to obtain simultaneously acceptable low excess error and tracking ability are clearly illustrated in (2.3-22).

In the last part of our work we have been concerned with the problem of numerical instability due to roundoff-noise in FRLS transversal algorithms. By restarting (zeroing) only the internal

variables and modifying the desired adaptive filter outputs during restart to account for the step function in the filter inputs during restart disruptions due to restart were minimized. The results indicate that by applying any one of these efficient restart procedures we can circumvent the numerical instability problem to which FRLS algorithms are subject, without significantly compromising its ability to track rapidly time-varying channels.

APPENDIX A-2Derivation of Equation (2.2-17)

We have already shown that the tap-coefficient error satisfies the following relation

$$\Delta \underline{C}(n) = \lambda \hat{R}(n)^{-1} \hat{R}(n-1) \Delta \underline{C}(n-1) + \hat{R}(n)^{-1} \underline{X}(n) e_{opt}(n)^* \quad (A-2.1)$$

Therefore

$$\begin{aligned} \langle E\{\Delta \underline{C}(n)^* \Delta \underline{C}(n)\} \rangle &= \lambda^2 \langle E\{\Delta \underline{C}(n-1)^* \hat{R}(n-1) \hat{R}(n)^{-1} \hat{R}(n)^{-1} \hat{R}(n-1) \Delta \underline{C}(n-1)\} \rangle \\ &+ \lambda \langle E\{\Delta \underline{C}(n-1)^* \hat{R}(n-1) \hat{R}(n)^{-1} \hat{R}(n)^{-1} \underline{X}(n) e_{opt}(n)^*\} \rangle \\ &+ \lambda \langle E\{e_{opt}(n) \underline{X}(n)^* \hat{R}(n)^{-1} \hat{R}(n)^{-1} \hat{R}(n-1) \Delta \underline{C}(n-1)\} \rangle \\ &+ \langle E\{\underline{X}(n)^* \hat{R}(n)^{-1} \hat{R}(n)^{-1} \underline{X}(n) |e_{opt}(n)|^2\} \rangle \end{aligned} \quad (A-2.2)$$

At the steady-state we can assume that $\hat{R}(n)$ and $\hat{R}(n-1)$ are almost equal.

Thus approximately

$$\hat{R}(n)^{-1} \hat{R}(n-1) \approx I_{NN} \quad \text{as } n \rightarrow \infty \quad (A-2.3)$$

In Appendix B-2 we have also shown, that asymptotically the inverse estimated autocorrelation matrix $\hat{R}(n)^{-1}$ and the true autocorrelation matrix R satisfy approximately the following relation

$$\hat{R}(n)^{-1} R \approx (1-\lambda) [I_{NN} - P(n)] \quad (A-2.4)$$

where $P(n)$ is a zero-mean fluctuation matrix.

Using (A-2.3) and (A-2.4) and referring also to the independence assumption of $P(n)$ from $\underline{X}(n)$ as well as to the fact that $E\{P(n)\} = 0$, the second term in (A-2.2) yields

$$\begin{aligned} & \lambda \langle E\{\Delta \underline{C}(n-1)^* \hat{R}(n-1) \hat{R}(n)^{-1} R \hat{R}(n)^{-1} \underline{X}(n) e_{\text{opt}}(n)^*\} \rangle \\ &= \lambda(1-\lambda) \langle E\{\Delta \underline{C}(n-1)^* (I-P(n))^* \underline{X}(n) e_{\text{opt}}(n)^*\} \rangle \\ &= \lambda(1-\lambda) \langle \Delta \underline{C}(n-1)^* \rangle E\{\underline{X}(n) e_{\text{opt}}(n)^*\} = 0 \end{aligned} \quad (\text{A-2.5})$$

Note that at the final step in deriving (A-2.5), the statistical independence assumption of $\Delta \underline{C}(n-1)$ with $\underline{X}(n)$, and, the orthogonality property, [19], which holds between $\underline{X}(n)$ and $e_{\text{opt}}(n)$ have been used. We can similarly show that the third term of (A-2.2) is zero. Combining (A-2.2), (A-2.3), (A-2.4) and (A-2.5) we get

$$\begin{aligned} \langle E\{\Delta \underline{C}(n)^* R \Delta \underline{C}(n)\} \rangle &= \lambda^2 \langle E\{\Delta \underline{C}(n-1)^* R \Delta \underline{C}(n-1)\} \rangle + \\ &+ (1-\lambda)^2 E\{\underline{X}(n)^* [I-P(n)]^2 R^{-1} \underline{X}(n) |e_{\text{opt}}(n)|^2\} \end{aligned} \quad (\text{A-2.6})$$

In general $e_{\text{opt}}(n)$ is uncorrelated with $\underline{X}(n)$. If we assume that they are also independent, (A-2.6) yields

$$\begin{aligned} \langle E\{\Delta \underline{C}(n)^* R \Delta \underline{C}(n)\} \rangle &= \lambda^2 \langle E\{\Delta \underline{C}(n-1)^* R \Delta \underline{C}(n-1)\} \rangle \\ &+ (1-\lambda)^2 E\{\underline{X}(n)^* [I-P(n)]^2 R^{-1} \underline{X}(n)\} \epsilon_{\text{opt}} \end{aligned} \quad (\text{A-2.7})$$

where

$$\epsilon_{\text{opt}} \triangleq E\{|e_{\text{opt}}(n)|^2\} \quad (\text{A-2.8})$$

or in a more compact form

$$\langle E\{\Delta C(n)^* R \Delta C(n)\} \rangle = \lambda^2 \langle E\{\Delta C(n-1)^* R \Delta C(n-1)\} \rangle +$$

$$(1-\lambda)^2 \text{trace} \{I_{NN} + E\{P(n)^2\}\} \cdot \epsilon_{\text{opt}}$$

(A-2.9)

APPENDIX B-2Derivation of Equation (A-2.4)

The estimated autocorrelation matrix $\hat{R}(n)$ has been defined as

$$\hat{R}(n) = \sum_{k=0}^n \lambda^{n-k} \underline{X}(k) \underline{X}(k)^* \quad (\text{B-2.1})$$

with elements

$$\hat{r}_{ij}(n) = \sum_{k=0}^n \lambda^{n-k} x(k-i+1) x(k-j+1)^* \quad 1 \leq i, j \leq N \quad (\text{B-2.2})$$

We observe that asymptotically and for stationary inputs

$$E\{\hat{R}(n)\} = \frac{1}{1-\lambda} E\{\underline{X}(n) \underline{X}(n)^*\} = \frac{1}{1-\lambda} R \quad (\text{B-2.3})$$

Thus a reasonable model for $\hat{R}(n)$ will be

$$\hat{R}(n) = \frac{1}{1-\lambda} R + \tilde{R}(n) \quad (\text{B-2.4})$$

where $\tilde{R}(n)$ is assumed to be a perturbation matrix with entries the zero mean random variables $\tilde{r}_{ij}(n)$ $1 \leq i, j \leq N$ independent from $x(n)$. Combining (B-2.2), (B-2.3) and (B-2.4) and referring to the independence assumption between $x(n)$ and $\tilde{r}_{ij}(n)$ we get

$$\text{var}\{\tilde{r}_{ij}(n)\} = \lambda^2 \text{var}\{\tilde{r}_{ij}(n-1)\} + \text{var}\{x(n-i+1) x(n-j+1)^*\} \quad (\text{B-2.5})$$

Thus asymptotically

$$\text{var}\{\tilde{r}_{ij}(n)\} = \frac{1}{1-\lambda^2} \text{var}\{x(n-i+1) x(n-j+1)^*\} \quad (\text{B-2.6})$$

It can easily be seen from (B-2.4) that

$$\hat{R}(n)^{-1} R = (1-\lambda) [I_{NN} + (1-\lambda) \hat{R}^{-1}(n)]^{-1} \quad (B-2.7)$$

Assuming that $\| (1-\lambda) \hat{R}^{-1}(n) \| < 1$ where $\| \cdot \|$ denotes some kind of consistent vector or matrix norm, we get

$$\hat{R}(n)^{-1} R \approx (1-\lambda) [I_{NN} - (1-\lambda) \hat{R}^{-1}(n)] \quad (B-2.8)$$

The randomness of (B-2.8) is embodied in the existence of the perturbation matrix $\hat{R}(n)$. The zero mean matrix

$$P(n) \triangleq (1-\lambda) \hat{R}^{-1}(n) \quad (B-2.9)$$

manifests the fluctuation of the product $\hat{R}(n)^{-1} R$ around the unit matrix I_{NN} . Therefore it is an expression for the estimation error involved in using exponential estimates for the autocorrelation matrix.

Let s_{ij} be the entries of the matrix $S = E\{P(n)^2\}$ and r_{ij}' be the entries of the inverse autocorrelation matrix R^{-1} . Then by using (B-2.6) and (B-2.9) and assuming that $\tilde{r}_{ij}$ are uncorrelated for all i, j it can easily be seen that

$$s_{jj} = \frac{1-\lambda}{1+\lambda} r_{ij}' \sum_k \text{var}\{x(n-j+1)x(n-k+1)\} r_{kk}' + \sum_{k \neq j} r_{ik}' E\{x(n-k+1)x(n-j+1)\} - r_{kj}'^2 r_{jk}' \quad (B-2.10)$$

Observe that for a given channel and input signal statistics, the elements of S can explicitly be determined.

APPENDIX C-2Derivation of Equation (2.3-4)

In the general case of nonstationary inputs the autocorrelation matrix of the input vector $\underline{X}(n)$ will be functions of the discrete time index n , i.e.

$$R(n) = E\{\underline{X}(n) \underline{X}(n)^*\} \quad (C-2.1)$$

Then according to (B-2.4) we can write

$$\hat{R}(n) = E\{\hat{R}(n)\} + \tilde{R}(n) \quad (C-2.2)$$

or

$$\hat{R}(n) = \sum_{k=1}^n \lambda^{n-k} R(k) + \tilde{R}(n) \quad (C-2.3)$$

If the variations of the environment are slow with respect to the memory of the adaptive algorithm then asymptotically

$$\hat{R}(n) = \frac{1}{1-\lambda} R(n) + \tilde{R}(n) \quad (C-2.4)$$

Note that in a system identification situation the above expression is valid with strict equality since in such a case the autocorrelation matrix does not depend on the time. Assuming that $||\cdot|| (1-\lambda) R(n)^{-1} \tilde{R}(n) || < 1$ where $||\cdot||$ denotes some kind of consistent norm it can be shown easily that

$$\begin{aligned}
 E\{\hat{R}(n)^{-1} \underline{X}(n) \underline{X}(n)^*\} &= E\left\{\left(\frac{1}{1-\lambda} R(n) + \tilde{R}(n)\right)^{-1} \underline{X}(n) \underline{X}(n)^*\right\} \\
 &= (1-\lambda) E\{R(n)^{-1} \underline{X}(n) \underline{X}(n)^*\} - (1-\lambda)^2 E\{R(n)^{-1} \tilde{R}(n) R(n)^{-1} \underline{X}(n) \underline{X}(n)^*\}
 \end{aligned}$$

(C-2.5)

Finally, using the independence assumption of $\tilde{R}(n)$ from $X(n)$ along with the fact that $E\{R(n)\} = 0$ we get

$$E\{\hat{R}(n)^{-1} \underline{X}(n) \underline{X}(n)^*\} = (1-\lambda) I_{NN} \quad (C-2.6)$$

REFERENCES - 2

- [1] B. Widrow: "Adaptive Filters", in Aspects of Network and System Theory, (R.E. Kalman, N. De Claris), Holt Rinehart Winston, 1971.
- [2] J.G. Proakis and J.H. Miller: "An Adaptive Receiver for Digital Signalling Through Channels With Intersymbol Interference", IEEE Trans. Inform. Theory, pp.487-497, July 1969.
- [3] R.S. Medaugh and L.J. Griffiths: "A Comparison of Two Linear Predictors", Proc. IEEE, ICASSP '81, Atlanta, pp.293-296, April 1981.
- [4] D.C. Farden and K. Sayood: "Tracking Properties of Adaptive Signal Processing Algorithms", Proc. IEEE, ICASSP '80, pp.466-469.
- [5] B. Widrow, J.M. McCool, M.G. Larimore, C.R. Johnson: "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter", IEEE Proceedings, pp.1151-1162, August 1976.
- [6] C. Gibson and S. Haykin: "Non-Stationary Learning Characteristics of Adaptive Lattice Filters", Proc. IEEE, ICASSP '82, Paris, April 1982.
- [7] Q.T. Zhang and S. Haykin: "Tracking Characteristics of the Kalman Filter in a Nonstationary Environment for Adaptive Filter Applications", Proc. IEEE, ICASSP '83, pp.671-674, Boston.
- [8] B. Widrow and E. Walach: "On the Statistical Efficiency of the LMS Algorithm With Nonstationary Inputs", IEEE Trans. Inform. Theory, March 1984, pp.211-221.
- [9] D.W. Lin: "On Digital Implementation of the Fast Kalman Algorithm", IEEE Trans. Acoust. Speech, Signal Processing, October 1984, pp.998-1005.
- [10] J.M. Cioffi and T. Kailath: "Fast Recursive-Least-Squares, Transversal Filters for Adaptive Filtering", IEEE Trans. on Acoust. Speech, Signal Processing, April 1984, pp.304-337.
- [11] F. Ling and J.G. Proakis: "Nonstationary Learning Characteristics of Least Squares Adaptive Estimation Algorithms", Proc. IEEE, ICASSP '84, March 1984, San Diego.
- [12] M.S. Mueller: "Least Square Algorithms for Adaptive Equalizers", BSTJ, pp.1905-1925, October 1981.

- [13] D.D. Falconer and L. Ljung: "Application of Fast Kalman Estimation to Adaptive Equalization", IEEE Trans. Comm., pp.1439-1446, October 1978.
- [14] R.D. Gitlin and F.R. Magee: "Self-Orthogonalizing Adaptive Equalization Algorithms", IEEE Trans. on Comm., pp.666-672, July 1977.
- [15] F. Ling and J.G. Proakis: "Adaptive Lattice Decision Feedback Equalizers and Their Applications on Fading Dispersive HF-Channels", Proc. IEEE, ICC '83, June 1983, Boston.
- [16] J.E. Mäzo: "On the Independence Theory of Equalizer Convergence", BSTJ, Vol.58, 963-993, May-June 1979.
- [17] D.J. Sakrison; "Stochastic Approximation: A Recursive Method for Solving Regression Problems", in Advances in Communication Theory, Vol.2, A.V. Balakrishnan, Ed., New York, Academic Press, 1966.
- [18] E.H. Satorious and J.D. Pack: "Application of Least Squares Lattice Algorithms to Adaptive Equalization", IEEE Trans. Comm., pp.136-142, February 1981.
- [19] G. Ungerboeck: "Theory on the Speed of Convergence in Adaptive Equalizers for Digital Communications", IBM J. Res. Develop., pp.546-555, November 1972.
- [20] G.C. Goodwin and R.L. Payne: "Dynamic System Identification: Experiment Design and Data Analysis", Academic Press, 1977.
- [21] D. Godard: "Channel Equalization Using A Kalman For Fast Data Transmission", IBM J. Research and Development, May 1974, pp.267-273.
- [22] F.M. Hsu: "Square Root Kalman Filtering For High-Speed Data Received Over Fading Dispersive HF Channels", IEEE Trans. Inform. Theory, pp.753-763, September 1982.
- [23] R.D. Gitlin, H.C. Meadors, and S.B. Weinstein: "The Tap-Leakage Algorithms: An Algorithm for the Stable Operation of a Digitally Implemented Fractionally Spaced Adaptive Equalizer", BSTJ, pp.1817-1839, October 1982.
- [24] B. Friedlander: "Lattice Filters for Adaptive Processing", Proc. IEEE, pp.829-867, August 1982.
- [25] D.D. Falconer, A.U.H. Sheikh, E. Eleftheriou and M. Tobis: "Comparison of DFE and MLSE Receiver Performance on HF Channels", IEEE Trans. Communications, May 1985, pp.484-486.

- [26] E. Eleftheriou and D.D. Falconer: "Restart Methods for Stabilizing FRLS Adaptive Equalizers in Digital HF Transmission", Proc. IEEE, Globecom '84, Atlanta, November 1984.
- [27] R.F. Magee and J.G. Proakis: "Adaptive Maximum Likelihood Sequence Estimation for Digital Signalling in the Presence of ISI", IEEE Trans. Inform. Theory, pp.120-124, January 1973.
- [28] D.A. George, R.R. Bowen, and J.R. Storey: "An Adaptive Decision Feedback Equalizer", IEEE Trans. Comm. Tech., pp.281-293, June 1971.
- [29] P. Monsen: "Feedback Equalization for Fading Dispersive Channels", IEEE Trans. Inform. Theory, pp.56-64, January 1971.
- [30] M.L. Honig and D.E. Messerschmitt: "Convergence Properties of an Adaptive Digital Lattice Filter", IEEE Trans. on Circuits and Systems, Vol.CAS-28, pp.482-493, June 1981.
- [31] M.L. Honig: "Convergence Models for Lattice Joint Process Estimators and Least Squares Algorithms", IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol.ASSP-31, pp.415-425, April 1983.
- [32] R.D. Gitlin, J.I. Mazo, and M.G. Taylor: "On the Design of Gradient Algorithms for Digitally Implemented Adaptive Filters", IEEE Trans. on Circuit Theory, pp.125-136, March 1973.
- [33] G.W. Stewart: "Introduction to Matrix Computations", Academic Press, 1973.
- [34] E. Eleftheriou and D.D. Falconer: "Steady-State Behaviour of RLS Adaptive Algorithms", Proc. IEEE, ICASSP '85, Tampa, Florida, March 1985.
- [35] L.L. Horowitz and K.D. Senne: "Performance Advantage of Complex LMS for Controlling Narrow-Band Adaptive Arrays", IEEE Trans. on Acoust., Speech and Signal Processing, pp.722-736, June 1981.

CHAPTER 3

ADAPTIVE DECISION-FEEDBACK-EQUALIZERS AND THEIR PERFORMANCE ON FADING DISPERSIVE CHANNELS

In many important radio applications, the medium connecting the transmitter to the receiver, contains several propagation "paths", so that some fraction of the total received energy unavoidably arrives over each path. It is also very common, that the individual path characteristics exhibit strong nonstationarity, causing time and frequency dependent variations of the amplitude and phase of the received signal. Such channels are usually referred to as multipath fading channels.

Historically, it was the high frequency, nominally 2-30 MHz, portion of the spectrum which attracted major considerations for long distance radio communications. In HF transmission, point-to-point propagation over a fairly long distance is achieved by using the reflections (or refractions) of the ionosphere. The availability and low cost were the main factors for such an early tremendous development of communication systems and techniques, suitable for the HF band.

Although we could call the last decade the satellite era, HF will continue to be extensively used in many civilian applications because of its low cost. Its main use, though, is for military purposes. The major limitation in satellite communications is the transmitted power or equivalently the establishment of an adequate signal-to-noise ratio. In contrast, in HF communications apart from an acceptable signal-to-noise

ratio, limitations are posed by the highly dispersive and fading nature of the transmission environment. In spite of these constraints, which had rendered HF receiver structures complicated during the past, the new spectacular advances in VLSI technology and contemplated signal processing devices are very promising for high frequency, high data rate communication links with moderate to high reliability.

This chapter is concerned with reception techniques for relatively high speed digital communications over 3 KHz-wide HF channels. Our specific approach has two essential aims. The first aim is to test different adaptive receiver schemes in terms of their capability to cope with the highly time-varying characteristics of the transmission medium. For this reason, a quite sophisticated HF channel simulator with various options has been developed. The second aim is to simulate receiver processing for recorded digitized signals that resulted from transmitting 2400 bps QPSK-modulated data over actual HF channels. These recordings, made and provided by the Communications Research Center of Canada, gave us the unique opportunity to work with real life data and compare the performance of the resulting receiver structures to the one we derive using the channel simulator.

The remainder of the chapter is organized as follows: Section 3.1 is a brief characterization of the transmission environment. Section 3.2 presents the inherent problems in transmitting data over high frequency propagation links. Section 3.3 discusses the development of a channel simulator. Section 3.4 presents the advantages and disadvantages of the different adaptation schemes for a decision feedback equalizer.

Section 3.5 deals with the DFE receiver performance over simulated HF channels. Section 3.6 discusses a whitening prefilter concept as a front-end to a LMS DFE. Section 3.7 presents the DFE receiver performance on real HF recorded data. Finally, Section 3.8 is a brief summary of the chapter.

3.1 HF Channel Characteristics

3.1.1 Multipath Propagation

One of the major sources of problems in HF communications is the multipath effect. The meaning of multipath is implicit in its name: The propagation medium contains several distinguishable paths such that echoes from the transmitted signal arrive at the receiver with possible different amplitudes and time varying, relative phase difference. Thus, the overall impulse response of the transmission medium is the superposition of the different path impulse responses with variable differential delays. This type of effect may arise due to the presence of ground wave, tropospheric wave and skywave paths and is usually due to the skywave returns from different ionospheric layers [1], [2].

Although in HF transmission many propagation modes are possible, depending on operational requirements, most of them experience a large attenuation, so that the surviving number of effective modes is in general small. Thus, for instance, as the transmitter-receiver distance is increased only the skywave mode (resulting from reflection from the ionosphere) contributes to the overall time-varying channel response,

while the tropospheric and ground wave modes experience significant attenuation.

In any case, multiplicity of the path in the transmission medium implies elongation or time dispersion in the unit impulse response of the medium. This time dispersion gives rise to intersymbol interference. The maximum serial data transmission is limited to the reciprocal of the maximum multipath spread time. For a 2500 Km path the maximum multipath spread has been shown to be about 3 msec [2]; for 1,000 Km it increases to 5 msec and for 200 Km it is about 8 msec. Sometimes even a single skywave mode of propagation gives rise to an approximately 100 μ sec spread due to roughness of the ionospheric layers [2].

3.1.2 Fading

The distortion revealed as a variation upon the received signal strength has been called fading [3], [4]. Fading can occur in many different ways. One common cause of fading on the received radio signals is due to the multiple propagation modes, each with time variant differential delays. Such time-varying multipath characteristics of the channel lead to a received signal which varies in amplitude and phase.

Another possible cause of fading in HF radio channels is due to ionospheric layer height variations. In this case the relative path length changes, inducing amplitude and phase variations upon the received signal.

Regardless of how regular or irregular the ionosphere is, the amplitude of the signal at the receiver would remain steady if the environment were stationary. Therefore, a rough measure of the changes of the ionosphere will be related to the width of the received power spectrum. Such a gross indication of the rapidity of the fading has been called fade rate. The inverse of the fade rate is a time span characteristic of the period in which changes occur in the medium. For HF channels, the fade rate is usually less than 1 Hz but on transauroral paths it can be 10 Hz or more [5].

3.1.3 Delay Distortion.

If the group delay introduced by the different ionospheric paths is a function of frequency, the various spectral components of the input signal suffer different amounts of time delay, resulting in phase or delay distortion. A complete picture concerning delay versus frequency characteristics for a given path is provided by the so-called ionograms [2]. Such diagrams reveal that the delay distortion is highly nonlinear when working close to the Maximum Useable Frequency (MUF) [2], [8]. Therefore, delay distortion can be significantly reduced by judicious choice of operating frequency. However, due to spectrum congestion, a favourable choice is not always possible.

Summarizing, in this section we have been briefly concerned, with the time and frequency characteristics of the high frequency propagation medium, which covers the 2-30 MHz band of the radio spectrum. The HF channel is a multipath time-varying environment producing time and

frequency spreads. The typical time dispersion is measured to be between 3-8 msec and depends highly on the range (physical path length). The frequency spreads are usually less than 1 Hz, without excluding the possibility of higher values.

3.2 Serial Data Transmission Over HF Channels

The major limitations on high data rate HF transmission are a result of the non-ideal characteristics of the medium, such as linear distortion, rapid channel variations and severe fading as well as bandwidth constraints. Due to the extent of a typical HF channel's time dispersion, efficient use of the available bandwidth is limited adversely by the presence of intersymbol interference. Therefore, transmission of a bit rate on the same order as or greater than the channel bandwidth is considered "high speed" and generally requires powerful equalization techniques.

In this study we have been concerned with 2400 bps transmission of QPSK modulated signals over 3 KHz-bandwidth HF channels. Using the complex representation of passband and baseband signals [9] the channel's real bandpass output waveform will have the form

$$r(t) = \sqrt{2} \operatorname{Re} \left\{ \sum_n a_n h(t-nT; t) e^{j2\pi f_0 t} + v_c(t) \right\} \quad (3.2-1)$$

where $h(\tau; t)$ represents a complex value time-varying equivalent low-pass impulse response accounting for transmitter and channel filtering. It is apparent that $h(\tau; t)$ in its dependence on the age variable τ , includes and describes the time variability of the transmission medium. For our

purposes, such representation will suffice [3], [4]. The data symbols $\{a_n\}$ are statistically independent discrete-valued complex random variables from a finite set; for the case of QPSK each a_n takes one of the four possible values $\frac{1}{\sqrt{2}} e^{j\theta_n}$, $\theta_n = \frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}$. The symbol interval is T seconds and the parameter f_0 represents the carrier frequency. Finally, the complex waveform $v_c(t)$ represents channel additive noise. After demodulation and low pass filtering, as shown in Figure 3.1, the complex valued baseband output waveform will be

$$x(t) = \sum_n a_n h(t-nT; t)^* + v(t) \quad (3.2-2)$$

where $v(t)$ is a lowpass filtered version of the additive noise $v_c(t)$.

The existence of several distinct more or less independent propagation paths on an HF channel reduces the likelihood that a deep fade occurs in which all paths fade simultaneously. However, this inherent diversity advantage of multipath, may introduce severe frequency selectivity. Thus, for high data rate serial transmission through 3 KHz HF channels, one is forced to consider adaptive equalization techniques.

The linear equalizer is extensively used for equalization of telephone channels. Attempts to compensate for severe frequency selective fading (which can cause nulls in the channel's frequency response) with a linear equalizer yield a very poor error rate performance. The reason is that a linear equalizer tends to enhance the channel noise at those frequencies most attenuated by the channel, in an attempt to produce a flat overall frequency response [10]. This indicates that a

Fig. 3.1 Receiver front-end

nonlinear equalization technique such as Decision Feedback Equalization (DFE) or Maximum Likelihood Sequence Estimation (MLSE, also known as the Viterbi algorithm) must be employed at the receiver.

In spite of its effectiveness, the computational and storage requirements of the MLSE receiver limit its application. The computational complexity of the MLSE receiver increases exponentially with the duration of the channel's impulse response. For a QPSK system in which the overall impulse response (including transmitting, channel, and receiving filters) spans L symbol intervals the number of operations required per output data symbol is 4^L . If in particular we consider the effective duration of the complex baseband HF channel impulse response, which is typically in the range of 2 to 8 msec, then for a 2400 bps QPSK system with a symbol rate of 1200 Hz it will amount to an impulse response span between 2 and 10 symbol intervals. The worst case numbers of the above realistic situation leads to an unacceptable complexity-performance tradeoff as far as the Viterbi receiver is concerned. Furthermore, a comparison study between DFE and MLSE receiver performance on real recorded HF channels [11] reveal that there is no spectacular advantage by using Viterbi decoder rather than a DFE. These results along with previous similar conclusions [13] argue for the adoption of a Decision Feedback Equalizer. Its relatively small complexity (similar to that of the linear equalizer) and its ability to perform much better on channels with spectral nulls, make the DFE appropriate for equalization of a time-variant multipath channel such as the HF.

The efficiency of a DFE receiver in demodulating single carrier bandwidth efficient modulation schemes, for high-speed data transmission over 3 KHz ionospheric channels has been demonstrated in [17]. There are a number of papers [18] - [22] which while attempting to study the performance of DFE receivers for communication over fading dispersive HF channels, tend to concentrate on the tracking problem in a time-varying environment presenting and testing different fast adapting RLS schemes. Apart from the high complexity issue concerning the adaptation algorithms themselves (Kalman, Square-Root Kalman, and Lattice), the simulated channel models used are oversimplified. The transmission medium is modelled as a T-spaced tapped delay line with usually 2 to 3 T-spaced time varying coefficients, followed by additive zero-mean white Gaussian noise. Such a model encompasses a matched filter and a symbol rate sampler and is referred to as the equivalent discrete-time white-noise filter model [10]. As a result, the effect of greater dispersed in time impulse responses, (not necessarily with T-spaced components), the existence of a fixed low-pass filter instead of a whitening matched filter, and the subsequent sensitivity to symbol clock sampling phase of T-spaced equalizer structures have not been considered. A considerably more realistic and flexible HF channel simulator has been developed in [23]. In this reference the performance of a T/2 fractional spaced DFE operating with the most computational intensive Square Root Kalman adaptation algorithm has been studied.

In the following sections we have focused on the performance of DFE receivers for communication over 3 KHz HF channels. For our

investigation a very sophisticated channel simulator similar in concept to the one presented in [23] has been developed. The behaviour of the DFE receiver has been tested not only on simulated data but also on real recorded HF data. In our approach the decision feedback equalization structures, utilizes a $T/2$ - rather a T -spaced forward transversal filter. In effect, it is adaptive in accordance with the most computational efficient Fast Recursive Least Squares (FRLS) or Fast Kalman algorithm [25], [26] employing a novel restart procedure [24], as has been demonstrated in the previous chapter. At this point we should emphasize that sampling and subsequent equalization at a rate $2/T$ (2 samples per symbol clock are shifted in the forward section) is appropriate not only as a means to eliminate the sensitivity to sample timing phase but also to effectively compensate for an extensively wide range of delay distortion and considerably improve severe amplitude distortion [16]. The disadvantages of using longer structures in terms of tap-coefficients, implying increased computational and storage requirements (and probably slower tracking behaviour) are more than compensated for by the superior performance of fractional spaced equalizers [16].

Finally, a further novel scheme has been tested which involves the cascade of a $T/2$ adaptive predictor followed by a $T/2$ -forward spaced LMS decision feedback equalizer. The predictor operates as a whitening filter and aims to produce an approximate equal eigenvalue autocorrelation matrix. The whitening of the input signal seen by the LMS DFE causes it to converge and track faster.

2 OF/DE 2

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

3.3 HF Channel Simulator

An essential part of this work is the development of a flexible baseband HF channel simulator. The implementation is mainly based on the general guidelines described in [4], [7]. Our specific approach however is closer to the one followed in [23].

The propagation medium usually contains a limited number of relatively distinguishable paths. In addition, the almost stationary nature of the medium, if we restrict attention, for a sufficiently short time (10-15 seconds), to bandlimited channels, allows for the use of a variable coefficient tap delay-line model [7]. Such a general simulation structure is shown in Figure 3.2. The input signal feeds a tapped delay line with differential delays $\{\tau_i\}_{i=1}^L$ where L is the total number of paths. A typical value of L is 2-4 paths. The delayed versions of the signal at each tap are multiplied by complex baseband suitable selected functions $\{h_i(t)\}_{i=1}^L$ and all products are summed together. At the output of the summer a complex valued, infinite bandwidth, zero-mean, white Gaussian noise, is added, with the resulting signal passing through a fixed bandwidth low-pass filter to form an output signal.

For the purpose of our simulation $T/8$ is the smallest permitted spacing in the delay line model. This resolution is adequate for all kinds of signal processing which takes place and allows a greater flexibility as far as selection and location of multipath components, and sample timing are concerned. Eight replications of the data symbols, before shifting them into the tapped-delay line, preserves consistency with the 9600 Hz operational sampling rate requirements of the model.

Fig. 3.2 HF Channel model

and maintains the rectangular pulse shape [23]. The low pass filter which essentially limits the noise and signal bandwidths has cutoff frequency at 1200 Hz.

The channel model we described is a general one until the time-varying tap-coefficients are specified.

The tap-coefficients, which represent the variable path amplitudes, are assumed to be independent complex zero-mean Gaussian random variables with fixed variance. They have been generated by passing white Gaussian noise generated at 9600 samples/sec, through low pass filters with 3 dB bandwidth at the order of the fade rate. The filters used to generate the Rayleigh fading path components are 2-pole Butterworth type filters with selectable bandwidth, as described in [5], [18]. The z-transform is given by

$$H(z) = \frac{1 + 2z^{-1} + z^{-2}}{1 + C_1 z^{-1} + C_2 z^{-2}} \quad (3.3-1)$$

where C_1 , C_2 determine the filter's bandwidth and subsequently the channel's fade rate. The 9600 samples/sec stream generated in this way is bandlimited by the LPF and then subsampled with the specified sampling phase resulting to a T or T/2 sample stream for equalization.

3.4 Adaptation Methods for a Decision Feedback Equalizer

A typical DFE consists of a forward and a feedback transversal filter. The inputs at the forward section are channel output samples at T/2 or T-seconds intervals, for a symbol rate of 1/T, and the inputs

at the feedback filter are previous data symbol decisions $\hat{a}(n)$.

A schematic representation of the transversal decision feedback equalizer is shown in Figure 3.3. Subject to the assumption of correct decisions, the feedback section essentially subtracts the intersymbol interference from the current symbol estimate caused by previously detected symbols, without suffering a significant noise penalty [4], [29], [30]. Referring to Figure 3.3 the input $z(n) \triangleq z(nT)$ at time nT into the quantizer is

$$z(n) = \sum_{i=0}^{2N_1-1} c_i^{f*} x(nT - iT/2) - \sum_{i=1}^{N_2} c_i^{b*} \hat{a}_{n-i} \quad (3.4-1)$$

where c_i^f are $2N_1$ complex forward tap-coefficients and c_i^b are N_2 complex feedback tap-coefficients. If we define the following complex partitioned vectors

$$\underline{C} \triangleq (c_1^{f*}, c_2^{f*}, \dots, c_{2N_1-1}^{f*}, c_1^{b*}, c_2^{b*}, \dots, c_{N_2}^{b*})^* \quad (3.4-2)$$

and

$$\underline{X}(n) \triangleq (x(nT)^*, x(nT - T/2)^*, \dots, x(nT - N_1 T + T/2)^*, \hat{a}_{n-1}^*, \dots, \hat{a}_{n-N_2}^*)^* \quad (3.4-3)$$

Then the equalizer output can be written in a more compact form as

$$z(n) = \underline{C}^* \underline{X}(n) \quad (3.4-4)$$

Normally, the tap coefficients must be computed adaptively since the channel is not known in advance at the receiver. Especially, in a nonstationary environment previous studies [2], [17], [18] have indicated that the ability of the adaptive equalization algorithm to

Fig. 3.3 Decision feedback equalization receiver

track the time-varying channel response is at least as important as the choice of the equalization method itself.

3.4.1 The LMS Decision Feedback Equalization Algorithm

The Least-Mean Squared (LMS) algorithm is the most commonly employed adaptation algorithm in many different signal processing applications. Its simplicity and ease of implementation make this algorithm an attractive solution for many practical problems. If $\underline{C}(n)$ represents the DFE receiver tap-coefficient vector at the n th iteration, and $\underline{X}(n)$ is the partitioned vector consisting of both sampled channel outputs and previous receiver decisions, as in equation (3.4-3), then the coefficient vector is updated in the LMS algorithm as follows:

$$\underline{C}(n+1) = \underline{C}(n) + \mu \underline{X}(n) e(n)^* \quad (3.4-5)$$

where

$$e(n) = \hat{a}_n - \underline{C}(n)^* \underline{X}(n) \quad (3.4-6)$$

and $\hat{a}_n$ is the receiver's most recently-decided data symbol. The positive scalar μ is the algorithm step size and must be chosen judiciously in order to guarantee convergence.

The adaptive transversal DFE shown in Figure 3.3 then adjusts itself, to approximately create an overall system response with no intersymbol interference. That is, the DFE receiver optimizes its coefficients iteratively, trying to reconstruct a delayed version of the data

symbol at time nT , denoted a_n , using $2N_1$ past samples and N_2 previous symbol estimates. For stationary inputs and suitable selected value of μ , this algorithm can be shown to converge eventually to the coefficient vector $\underline{C}_{opt}$ that minimizes the mean-square error $E\{|e(n)|^2\}$. The speed of convergence as well as the tracking properties of the LMS algorithm depend on μ but mainly on the spread in the eigenvalues of the positive definite autocorrelation matrix $R = E\{\underline{X}(n) \underline{X}(n)^*\}$ [10], [12].

3.4.2 RLS Algorithms in Decision Feedback Equalization

Filter adaptation algorithms with faster tracking capability than that of the LMS algorithm are available for channel equalization of rapidly time-varying channels such as HF radio channels. These recently developed faster algorithms require roughly an order-of-magnitude increase in the number of arithmetic operations per processed sample relative to the LMS algorithm. However, they appear to be within the capabilities of existing or near-future VLSI signal processing devices for signal bandwidths limited to 3 or 4 KHz. They are applicable to both transversal and lattice filter structures.

Here we briefly restate the RLS criterion applied to a decision feedback equalizer, which results in superior convergence and tracking properties in comparison to the typical LMS algorithm. In particular, a recursive least squares adaptive DFE yields at any time nT that coefficient vector $\underline{C}(n)$ which minimizes the weighted cumulative square error

$$J(n) = \sum_{k=0}^n \lambda^{n-k} |\hat{a}_k - \underline{C}(n)^* \underline{x}(k)|^2 + \delta \lambda^n \underline{C}(n)^* \underline{\Lambda}_N \underline{C}(n) \quad (3.4-7)$$

The exponential weighting factor λ is a positive constant close to but less than 1, while the parameter $\delta (\delta > 0)$ ensures stability at the start-up of the algorithm. Finally, $\underline{\Lambda}_N$ is a $N \times N$ matrix defined in Section 2.4.2.

In general, the superior tracking properties of the RLS algorithms compared to those of the LMS algorithms have been studied in depth in Chapter 2. We have shown that regardless of adaptive system structure and independently of the variations of the environment, the convergence of the lag error vector is governed by a unique time constant. This time constant does not depend on the eigenvalue ratio of the autocorrelation matrix and is given by

$$\tau_i = \frac{1}{1-\lambda_i}, \quad 1 \leq i \leq N \quad (3.4-8)$$

where N indicates the total number of tap-coefficients. In decision feedback equalization of a time varying channel where the eigenvalue ratio can take on any realistic value, such behavior explains the performance advantage we obtain by using RLS algorithms instead of the simple LMS algorithm.

Efficient methods which minimize recursively in time and order $J(n)$ in single channel applications are available. Such algorithms are discussed in [19], [25], [26], [33]. In the case of a decision feedback equalizer and especially when a $T/2$ -spaced forward section is employed as opposed to the T -spaced feedback section, one must resort

to the so-called "multichannel" versions of the RLS algorithms with different sampling rates. For a general multichannel RLS estimator, processing M parameters in each channel such that the dimensionality of the input signal vector to be $N = Mp$, the number of computation and storage requirements rise in proportion to Mp^2 or Mp^3 , where p denotes the number of new entries at each iteration [26], [27], [33], [34]. Such an unusually high increase in complexity, in proportion to p^2 or p^3 shouldn't be surprising, since in a p -channel RLS algorithm we are dealing at each time instant with p -dimensional vectors and pxp matrices. Although a fractional spaced RLS DFE is essentially a 2-input estimator, it is classified as a p -channel one, where $p > 2$ in order to accomodate the higher sampling rate of the signals in the forward part. Thus, the shifting property crucial for computation reduction in all "fast" RLS schemes is exploited only at the symbol rate $1/T$, resulting in an increase of the algorithm's complexity.

The square-root Kalman [19], [23] is the most stable (in a floating-point arithmetic environment) recursive least squares transversal filter. It has the disadvantage of requiring matrix operations. Therefore, the number of calculations is proportional to N^2 and grows very fast with increasing N . Especially in a fractional spaced implementation, one would expect an additional increase in computational burden and memory requirements.

The FRLS multichannel transversal schemes possess the lowest complexity of all least squares algorithms [25], [26], [33]. Their main problem is numerical instability due to the propagation of roundoff

errors associated with the adaptation process.

Very recently normalized versions of the FRLS transversal algorithms which greatly improve performance in a finite precision environment have been introduced [33]. However, rescue operations may still be required in certain applications in order to avoid the disastrous effects of roundoff errors, although not as frequently as in the traditional unnormalized FRLS adaptation algorithms [33].-

An alternative to transversal filter type algorithms is to employ a lattice filter approach towards minimizing the least squares cost function given by equation (3.4-7), [22], [27]. Multichannel lattice filters are known to exhibit very good numerical stability properties. Their complexity, though proportional to M_p^2 or M_p^3 , is still quite intensive in comparison with that of the FRLS transversal multichannel algorithms because of the large coefficient of p^3 or p^2 . We should emphasize also that both the adaptive lattice and square-root Kalman algorithms require divisions proportional to N , thus their complexity is even greater.

While the square-root Kalman transversal adaptive filter or the various lattice adaptive filters have been advocated for equalization of fading HF channels [18], [19], [22], [23] primarily for their numerical stability properties, (note that formulation or implementation of a complex fractional spaced fast lattice DFE has not been reported) we have focussed our attention on the simpler fractional spaced FRLS transversal DFE. Numerically stable propagation of the algorithm is,

ensured by any one of the efficient restart techniques developed in the previous chapter.

3.4.2.1 FRLS Transversal DFE With An Efficient Periodic Restart

The fast transversal algorithm, with its derivation, for computing the Kalman gain vector sequences and subsequently for updating the decision feedback equalizer tap-coefficients can be found in [25]. Appendix A-3 contains the complex form of the algorithm and its extension to accommodate fractional input samples on the forward input section.

The general principles and basic ideas behind the restart procedures for FRLS transversal adaptive filters have been presented in detail in Chapter 2. Here we shall briefly restate the application of the most efficient one, Restart Procedure 1, in decision feedback equalization.

Following a restart at time n_0T all inputs $x(nT)$, $x(nT - \frac{T}{2})$ and $\hat{a}_{n-1}$ are taken to be zero for $n \leq n_0$. In other words, for $n > n_0$, the N-dimensional input vector for purposes of adaptation is taken to be $\underline{X}(n)$ whose components consist of

$$\underline{\tilde{X}}(n) = \begin{cases} \underline{\tilde{X}}(n) & \text{for } n > n_0 \\ 0 & \text{for } n \leq n_0 \end{cases} \quad (3.4-8)$$

Where $\underline{\tilde{X}}(n)$ is a 3-dimensional vector ($p=3$) with components

$$\underline{\tilde{X}}(n) = (x^*(nT), x^*(nT - \frac{T}{2}), \hat{a}_{n-1}^*)^* \quad (3.4-9)$$

According to restart procedure 1, equation (2.4-9) suggests that during the reinitialization period the desired outputs used for FRLS DFE

adaptation should be

$$\hat{\tilde{a}}_n = \hat{a}_n - \underline{C}_{aux}(n)^* \underline{X}(n)^{(-)} \quad (3.4-10)$$

where $\underline{C}_{aux}(n)$ is the current set of tap coefficients from the auxiliary LMS algorithm, and $\underline{X}(n)^{(-)}$ is a N-dimensional signal vector defined in (2.4-7).

A slightly different approach but within the framework of restart procedure 1, is to avoid the auxiliary LMS algorithm which adds into overall complexity of the structure. In this case in place of $\underline{C}_{aux}(n)$ we use the not zeroed set of tap coefficients $\underline{C}(n)$. Thus, the useful input into quantizer during restart will be

$$z(n) = \underline{C}(n)^* \underline{X}(n) \quad (3.4-11)$$

while equation (3.4-10) again provides the desired response during the same period, i.e.,

$$\hat{\tilde{a}}_n = \hat{a}_n - \underline{C}(n)^* \underline{X}(n)^{(-)} \quad (3.4-12)$$

In both cases following a restart at time n_0T , based on method 1, the FRLS DFE minimizes

$$\begin{aligned} J(n) &= \sum_{k=n_0}^n \lambda^{n-k} \left| \hat{\tilde{a}}_k - \underline{C}(n)^* \underline{X}(k) \right|^2 + \delta \lambda^{n-n_0} (\underline{C}(n) - \underline{C}(n_0))^* \Lambda_n (\underline{C}(n) - \underline{C}(n_0)) \\ &= \sum_{k=n_0}^n \lambda^{n-k} \left| \hat{a}_k - \underline{C}_{aux}^*(k) \underline{X}(k)^{(-)} - \underline{C}(n)^* \underline{X}(k) \right|^2 + \delta \lambda^{n-n_0} (\underline{C}(n) - \underline{C}(n_0))^* \Lambda_N (\underline{C}(n) - \underline{C}(n_0)) \end{aligned} \quad (3.4-13)$$

If the variations of the environment are slow with respect to the baud interval T , and the reinitialization period is sufficiently small, then approximately

$$J(n) = \sum_{k=n_0}^n \lambda^{n-k} \left| \hat{a}_{k-\underline{C}(n)}^* (\tilde{X}(k) + \tilde{X}(k)) \right|^2 + \delta \lambda^{n-n_0} (\underline{C}(n) - \underline{C}(n_0))^* \Lambda_N(\underline{C}(n) - \underline{C}(n_0)) \quad (3.4-14)$$

Finally using (2.4-7) we get

$$J(n) = \sum_{k=n_0}^n \lambda^{n-k} \left| \hat{a}_{k-\underline{C}(n)}^* \underline{X}(k) \right|^2 + \delta \lambda^{n-n_0} (\underline{C}(n) - \underline{C}(n_0))^* \Lambda_N(\underline{C}(n) - \underline{C}(n_0)) \quad (3.4-15)$$

Therefore, modification of the desired response minimizes the disruption associated with the step function effects on the filter's input signal vector. The FRLS restart interval is sufficiently short so that the momentary reliance on the slower adapting LMS algorithm or on the non-zeroed set of tap-coefficients of the restarted FRLS DFE, should cause little reduction in the adaptive filters tracking capability.

The performance of all, auxiliary-LMS-based, restart methods as far as mean square error is concerned has been presented in Chapter 2. Here we shall compare the error-rate performance of the two options in restart procedure 1 discussed previously. The decision feedback equalizer used in this investigation has $N_1 = 16$ $T/2$ -spaced forward complex tap-coefficients. The unknown time-varying channel is a 0.15 Hz Rayleigh fading channel with a $4T$ symbols interval multipath spread, while the signal-to-noise ratio is set to 20 dB. In all cases the restart interval is 30 symbols and the simulation runs are 200,000 symbols long. Table 3.1 lists the number of observed symbol errors, as a function of the

TABLE 3.1

Periodicity of Initiation of Restart ⁺	Number of Symbol Errors Observed	
	Restart Procedure 1 With Auxiliary LMS Alg.	Restart Procedure 1 No Auxiliary LMS Alg.
300	881	1431
500	911	1193
750	866	1141
1000	852	1081

⁺ Restart interval = 30 symbols

periodicity (in symbol intervals) of initiation of a restart. We can see that in the case where an auxiliary LMS algorithm is used in connection with restart procedure 1, then the number of symbol errors is almost insensitive to how frequently we invoke the restart. While, in the case where we do not use an auxiliary LMS DFE, then the error rate decreases as the periodicity on initiation of a restart increases.

On the basis of simulation results like these, we concluded that a valid and efficient procedure for periodically restarting the FRLS adaptation algorithm applied to a DFE receiver for HF channels is Restart Procedure 1 aided by an LMS algorithm. This is the restart procedure employed in all FRLS simulations described henceforth.

The second option which is less complex in terms of computation calls for a more stable FRLS algorithm (e.g., a square root formulation of the fast Kalman algorithm has been derived in [33]) since it would require less frequent restarting operations. Of course it can always be applied to virtually all stationary channels such as the telephone channel without loss in performance.

3.5 DFE Receiver Performance Over Simulated HF Channels

3.5.1 Ideal Reference Mode

In this section we shall study the bit error rate performance of the FRLS DFE receivers as a function of the signal-to-noise ratio in a variety of Rayleigh simulated HF channels; for different fade rates

and multipath spreads. In the simulations reported here, the data symbols which enter the feedback section of the DFE and which are used for adaptation purposes, were not receiver decisions, but were the known transmitted data symbols. This is the ideal reference mode. The results of these simulations shed light on the inherent capabilities of the DFE equalizer and its adaptation methods. The simulation results reported in Section 3.5.2, are for the realistic case, where the data symbols used by the adaptive receiver are its own decisions, and thus include the effects of error propagation or "crashes" caused during deep fades.

In our study the bit error probabilities have been estimated over a 200,000 data symbol long (400,000 bits) simulation runs. In all cases the SNR is defined as the average transmitted power at the input of the receiving low-pass-filter over the average additive-noise power at the same point.

Figure 3.4 shows the bit error rate as a function of the SNR for various fade rates. The channel used in this investigation has 3-equal strength (in mean-square sense) $2T$ -spaced paths, i.e., a multipath spread (MPS) of $MPS = 4T$. The length of the decision feedback equalizer has been fixed to the total $N = 20$ coefficients; 16 $T/2$ -spaced forward coefficients and 4 feedback coefficients. For the slowest fading channels, i.e., 0.08 Hz and 0.15 Hz, the exponential weighting factor λ was set to 0.97 while in the other two cases we choose $\lambda = 0.94$. In the same figure we present results for a 0.15 Hz fading channel, concerning a LMS DFE. The superiority of the FRLS DFE is evident. In general, the

Fig. 3.4 P_e versus SNR for a 3-path, $2T$ spaced Rayleigh fading channel, $MPS=4T$ (ideal reference model).

FRLS exhibits much better performance at smaller fade rates, while in the case of 1.1 Hz fading bandwidth or higher it bottoms out fast. These results are generally in close agreement with those presented in [23]. There under similar channel conditions and using an uninterrupted square-root Kalman algorithm, error-rate versus SNR curves have been obtained, which are approximately 5 - 10 times better in every case (even in the case of a LMS DFE). This leads us to conclude that this small difference in performance should be attributed to the different filter structure (equation (3.3-1)) used in limiting the bandwidth of the tap-coefficients driving noise processes of the HF channel, rather than to the slight suboptimality in tracking performance introduced by restarting the FRLS DFE.

The tracking performance of the FRLS algorithms under restarting conditions is an important issue and deserves further analytical and experimental investigation, especially at high SNRs. It is not that crucial though in an HF modem which will work in some kind of decision directed mode. In this case, as we shall discuss later, deep fades will induce crash situations and thus retraining the equalizer or even restarting it, would be a necessary measure regardless of numerical instability problems.

Another factor which contributes to slow tracking behavior and subsequently to high bit error rates, is the length, in terms of coefficients of the DFE structure. In our experiments in both simulated and real HF channels the total number of coefficients is in the range of 20-30 in order to cover all possible multipath spreads. Thus it is

in contrast to what has been used in [23] where the maximum number is 14 or 15 tap-coefficients.

Figure 3.5 depicts the bit error rate versus SNR for a much more dispersed in time channel and for various fade rates. In this case the number of paths is 4 with $MPS = 7T$. The length of the DFE has been increased to $N=25$ tap-coefficients (18 $T/2$ -spaced taps and 7 feedback taps). This is a good example which rules out any doubt with regard to the use of the Viterbi receiver because of the computational complexity[†]. Again the FRLS DFE receiver shows good performance especially at the low fade rates.

3.5.2 Decision Directed Mode

For a sufficiently deep fade, the receiver's error rate can become high enough that the sequence of decisions used for filter adaptation become unreliable; then, the adapted equalizer tap-coefficients may wander far from their optimal values, and catastrophic error propagation occurs. The onset of this disastrous error propagation or "crash" can only be hastened by the use of filter adaptation algorithms which rapidly adapt to changing channel conditions. Such crashes for decision-directed adaptation were observed during deep fades in our simulation; the decision-directed FRLS algorithm usually failed to reconverge within a reasonable time following a crash.

Therefore, subsequent simulations which used the receiver's decision $\{\hat{a}_n\}$ for equalizer adaptation and for decision feedback equalization were

[†] In this case the computational complexity of the Viterbi receiver would be proportional to 4^L operations per output data symbol.

modified in the following way called the decision-directed mode.

Periodically, the receiver's decisions were replaced by the portion of the known pseudo-random data sequence during that period.

Periodic resynchronization of the adaptive receiver with an ideal reference appears to be a practical measure for fading HF channels. An alternative approach to crash recovery is transmission of a training sequence on demand via a feedback channel as suggested in [18]. In our simulations the equalizer tap coefficient vector was never reset to zero; however, there may be some advantage to resetting this vector to all-zeros when the receiver's estimated signal-to-noise ratio has been below a threshold for more than some given time interval. In the case of the FRLS DFE receiver the training sequence coincided with each restart period. The decision directed mode has been tested in the case of a 0.15 Hz fading bandwidth for the 2T spaced 3-path Rayleigh fading channel. The results are shown in Figure 3.6. The pair (530, 30) indicates that an ideal reference training sequence of 30 data symbols was transmitted once every 530 symbol intervals. This is referred to as 5.6% ideal reference training sequence. Obviously by increasing the percentage of retraining, the equalizer will approach the ideal reference case. More simulation results on the decision-directed mode as well as discussion on the limitations posed by crashes will be presented in the next sections in connection with the real HF data.

3.6 A Whitening Predictor As A Front-End For An LMS DFE

It is well known that fast convergence and tracking with an LMS algorithm is not possible when the eigenvalues of R are very different in magnitude. Especially, in a dynamically changing environment, such as the MF multipath fading channel, the eigenvalue ratio of the autocorrelation matrix might take on a variety of values. This explains the very poor performance of the LMS DFE demonstrated in the previous sections. In order to improve convergence and tracking behavior, we should attempt to reduce the eigenvalue ratio. It has long been recognized that it is the correlation between the input samples which affects the difference between the eigenvalues. The eigenvalues can become equal if we apply an orthonormal transformation to the autocorrelation matrix R . This is essentially the fundamental principle behind the development of adaptive lattice equalizers. Basically, the adaptive lattice algorithms generate a set of orthogonal signals which can be used as inputs to equalizer tap-coefficients [28].

In a DFE case the inputs into the feedback section are usually uncorrelated, being past symbol decisions. Thus reduction of the overall eigenvalue spread, calls for decorrelation or whitening of the forward section input signal. This can be achieved either by using all order backward residuals of a Lattice predictor for orthogonalization of the feedforward input signal vector, or by using only the highest order residual for whitening the channel input signal stream. The second approach is the one we shall discuss in the sequel.

Figure 3.7 shows the combination of a lattice filter in cascade with a LMS adapted decision feedback equalizer. The lattice filter has as input the sampled channel output signal while its highest order normalized backward residual feeds the forward section of the DFE. Since the lattice structure operates as a prediction-error filter it essentially "whitens" the data stream seen by the LMS DFE. Such a whitening reduces the eigenvalue ratio and causes the LMS algorithm to track faster. Nevertheless, the use of the highest order residual rather than all order residuals in order to decorrelate the input signal vector elongates the overall impulse response to be equalized. So one would expect that in general a longer DFE structure might be necessary in order to achieve a specific mean square error. Simulation experience, however, indicates that a small number of lattice stages is more than adequate for the necessary whitening without the need of increasing the length of the DFE significantly.

Implementation of this simple concept requires a gradient or FRLS lattice predictor followed by a conventional LMS DFE, both adapted separately. The computational advantage of this approach is more apparent in a fractional spaced implementation of the receiver. In this case, the whitening predictor exploits the shifting property of the sampled channel output, since the desired training sequence is available at the same sampling rate with this of the input signal. Thus, a single channel predictor adapted at the same rate as the input sample stream, can be used.

Fig. 3.7 Lattice whitening filter followed by a LMS DFE

Figure 3.8 shows the bit error rate performance, in ideal reference mode, of the receiver with a 7-stage lattice predictor followed by a transversal LMS DFE, compared to a conventional LMS DFE and to the FRLS transversal DFE with periodic restart. The channel used is a 3-path 2T-spaced Rayleigh fading channel model. As we can see the simple whitening lattice approach offers a probability of error which is superior to the transversal LMS DFE but inferior to that of the FRLS DFE.

In conclusion, employing a whitening predictor as a front end to LMS equalizers, linear or decision feedback significantly improves convergence and tracking behaviour at the expense of a slight increase to complexity. Such an approach appears to be suitable for a variety of applications and especially to those where computational burden is a very critical factor.

3.7 DFE Receiver Performance On Real HF Channels

This part of our work deals with computer simulation of receivers which process digitally-recorded received waveforms from real HF channels. These recordings were made and provided to us by the Communication Research Centre.

3.7.1 Transmitted Signal and Characteristics of the Recorded Data

The transmitted signals were generated by QPSK-modulating a length 1023 pseudo-random pattern. The symbol rate, derived from a Cs beam standard, was 1201.739 Hz. Thus with 2 bits per symbol,

the QPSK carrier frequency was 1505.107 Hz. The approximately 2.4 KHz-bandwidth signal thus generated, was subsequently up-converted to the desired RF frequency in the HF band.

The modulated signals were transmitted over several HF channels. It is worthwhile to describe the behaviour of the channel when the data was being recorded. Such description will provide information about the average signal-to-noise ratio, the fade rate, the transmitted power, time of recording and mode of propagation.

As we have already mentioned, the data was obtained from measurements of 1200 baud QPSK signals transmitted from Cobden to the CRC laboratories at Shirley Bay, a distance of approximately 90 km. The signals were received via skywave without any significant groundwave component. The observed subjective characteristics of the recorded tapes are listed in Table 3.2. From Table 3.2, it seems that tapes MDA010 and MDA011 will provide sufficient information to cover a variety of impairment in the HF channel.

The received voiceband signals were sampled and digitized to 12 bits and were recorded in digital form by CRC using a PDP-11/23 minicomputer. The sampling rate used for recording was 9600.00 Hz.

The front end of the receivers included sine and cosine demodulation followed by low pass filtering as shown in Figure 3.1. The low pass filters shown in this figure are identical raised cosine filters with cutoff bandwidths at 1200 Hz. The output samples are further decimated by a factor of 4 so that the sampling frequency of the complex baseband signal used for equalization is $2/T$.

TABLE 3.2⁺SUBJECTIVE CHARACTERISTICS OF THE RECORDED TAPES

(The average signal-to-noise ratio at the receiver input = 13 - 16 dB).

Tape No.	Power (Watts)	Frequency (MHz)	Transmitted Time (Local Time)	Features/ Characteristics
MDA006	16	7.5555	11:40	Strong, steady and clear signal
MDA010	16	5.2675	16:00	Deep slow fading (> 20 dB fades) with very strong signal fade rate = 0.15 Hz
MDA011	16	5.2675	17:40	Strong signal with deep (down to noise level) rapid fades, fade rate = 0.5 Hz

⁺ All tapes with 1200 baud QPSK signals.

3.7.2 Performance Comparison of FRLS and LMS Adaptation Algorithms

A DFE receiver was simulated to test its performance by using the complex baseband 2.4 kbs QPSK-demodulated, recorded signals. The input signal in the forward part was presented as complex samples, $T/2$ seconds apart, where T is the duration of one symbol (1/1200 sec). The QPSK data symbols had the four possible values $(\pm 1, \pm j)/\sqrt{2}$.

Preliminary simulations established the choice of the number of equalizer tap-coefficients: 20 $T/2$ -spaced forward tap-coefficients and 7 feedback tap-coefficients, resulting to a total number of $N = 20+7 = 27$.

The simulated DFE equalizers, adapted according to the LMS algorithm and also according to the FRLS algorithm with periodic Restart Procedure 1, were compared with respect to the error rate. The error rates were measured in 2 ways:

- (1) Fraction of data bits in error
- (2) Probability distribution of the number of symbol errors occurring in blocks of 100 transmitted symbols.

The total number of transmitted symbols in each simulation was 30,000 (300 blocks).

Figure 3.9 shows the block error distribution - the probability that more than n errors occur in a block of 100 as a function of n - for the ideal reference mode on channel MDA011. The LMS adapting DFE exhibited on the order of twice the block error rate of the FRLS adapting DFE. Figure 3.10 shows the corresponding ideal reference block error statistics for channel MDA010. For this channel there was

Fig. 3.9 Error Statistics for FRLS and LMS
DFE on Channel MDA 011 (Ideal Reference Mode)

Fig. 3.10 . Error Statistics for FRLS and LMS DFE
on channel MDA 010 (Ideal Reference Mode)

less difference between the performance of the FRLS and LMS adaptation algorithms. Both channels exhibited errors bursts due to fading.

The largest numbers of consecutive blocks containing errors were 6 for the LMS algorithm on MDA011 and 19 for the LMS algorithm in MDA010.

The last 16,000 symbols on channel MDA010 were transmitted error free in all cases. Tables 3.3 and 3.4 summarize the error statistics.

It is clear that for these two channels in the ideal reference mode, the FRLS adaptation algorithm did not show a spectacular performance advantage over the simpler LMS algorithm. This lack of a clear-cut advantage for the fast-adapting FRLS algorithm may be attributed to the low signal-to-noise ratios experienced in these tapes.

Figures 3.11 and 3.12 show the block error distribution, for channels MDA011 and MDA010 respectively in the decision directed mode. In order to avoid crash situations, the same decision directed scheme described in the previous sections has been adopted. It was assumed that an ideal reference "training sequence" of 30 data symbols was transmitted once every 330 symbol intervals. Thus 9.1% of the transmitted data was an ideal reference training sequence. In the case of the FRLS algorithm, the training sequence coincided with each restart period.

It is seen that the absence of an ideal reference except during the periodic training periods, greatly increased the total number of symbol errors. The fraction of blocks containing one or more errors did not increase much for the receiver using the decision-directed

TABLE 3.3

SUMMARY OF ERROR RATES FOR CHANNEL MDA011

	Ideal Reference Mode		Decision Directed	
	LMS	FRLS	LMS	FRLS
Block error rate	.103	.06	.193	.06
Bit error rate	0.005	.003	.077	.010

TABLE 3.4

SUMMARY OF ERROR RATES FOR CHANNEL MDA010

	Ideal Reference Mode		Decision Directed	
	LMS	FRLS	LMS	FRLS
Block error rate	.140	.137	.153	.133
Bit error rate	.014	.012	.056	.045

Fig. 3.11 Error Statistics for FRLS and LMS DFE
on Channel MDA 011 (Decision-directed mode)

Fig. 3.12 Error Statistics For FRLS and LMS DFE
on Channel MDA 010 (Decision-directed mode)

FRLS algorithm. However, the block error rate tripled for the LMS receiver on channel MDA011. This was the only case observed in which the FRLS algorithm displayed markedly superior performance. Table 3.3 for channel MDA011 and Table 3.4 for channel MDA010 summarize the average error rates observed in each of the 30,000 - symbol simulations of DFE receivers in both ideal and decision directed mode. Figure 3.13 shows the blocks (numbered, from 1 to 300) containing one or more errors in each simulation. Isolated erroneous blocks are shown as vertical lines, groups of two or more consecutive erroneous blocks are shown as rectangles. The ideal reference results are shown above, and the decision-directed results below the horizontal lines. Again, the very bursty nature of the decision errors is apparent. The time of occurrence of error bursts coincides with dips in the channels' signal-to-noise ratio shown in Figures 3.14 and 3.15 for channels MDA011 and MDA010 respectively. The measurements in Figures 3.14 and 3.15 were made using an adaptive channel identification algorithm; the mean squared error used to determine the SNR was averaged over 500 symbol intervals, so as to present an overall picture over 30,000 symbol intervals, minimizing instantaneous fluctuations. Superimposed on these figures are the error bursts shown in Figure 3.13.

In conclusion, we should again mention that as far as the real data simulations are concerned, the restart FRLS algorithm displayed markedly superior performance over the LMS algorithm only in channel MDA011. Other measurements though [11] showed that MDA010 exhibited less time variation than MDA011 apart from one deep sustained fade.

Fig. 3.13 Blocks Containing One or More Errors.
 Ideal Reference Cases Shown Above, Decision-
 Directed Cases Shown Below.

Fig.3.14 SNR Versus Time For Channel MDA010
(SNR Averaged With 500-Symbol Interval Time Constant)

Fig.3.15 SNR Versus Time For Channel MDA011
(SNR Averaged With 500-Symbol-Interval Time Constant)

3.8 Summary

In this chapter we have focussed on reception techniques for data transmission over 3 KHz bandlimited fading multipath HF channels. The relatively long impulse responses which can occur due to multipath propagation argue for the adoption of a DFE receiver instead of the more powerful but complex, MLSE receiver.

Implementation of the DFE receiver with a fractional spaced forward section was found to be necessary. In essence, fractional spaced equalizers rule out the sensitivity of choice of sampling phase and function more effectively on heavily distorted channels.

We have shown that the tracking behaviour of the RLS DFE does not depend on the channel characteristics to which the LMS DFE is subject, but only on the exponential weighting factor λ . The resulting error rate performance advantage of the FRLS DFE over the LMS DFE was evident in all Rayleigh fading simulated channels. In the case of the recorded real data the difference in performance was less significant.

The problems of instability due to accumulating roundoff errors in FRLS transversal algorithms was met and dealt with by means of an efficient periodic restart procedure. As well as the periodically applied restart procedure for the FRLS adaptation algorithm, it was found necessary to modify the normal decision-directed mode of operation (use of receiver decisions in the DFE feedback signal and for filter adaptation) by periodically inserting a training sequence. Although a small portion of the transmitted data capacity was thereby lost,

periodic retraining allowed the adaptive receiver to recover from the effects of severe fading.

In addition, another new concept has been tested in which a fractional operating whitening predictor is used as a front-end for a fractional spaced LMS adapted DFE. Although the performance of such a scheme in a HF context is better than the simple LMS DFE it is still inadequate in comparison to the FRLS DFE. The reduced overall complexity offered by the above mentioned structure encourages further study in different applications where computational complexity is crucial.

Concluding, the experimental results on both simulated and real data indicate that fade rates approximately as high as 0.5 Hz form an upper limit regarding an acceptable error rate performance of a bit by bit adapting serial DFE receiver. For fading bandwidths higher than that, even the powerful fast adapting RLS algorithms reach their limitations resulting in poor performance. We should emphasize once more that although the results of our investigation might look grim in certain cases, they have been drawn under worst case scenarios, i.e., fairly long impulse responses and relatively high fade rates:

APPENDIX A-3

In the FRLS transversal DFE the following quantities are stored and updated at each iteration:

- (1) $\underline{\xi}(n), \underline{\rho}(n)$: p-dimensional complex vectors
- (2) $\gamma(n)$: real scalar
- (3) $\underline{F}(n), \underline{B}(n)$: Nxp complex matrices
- (4) $\underline{E}(n)$: pxp complex matrix
- (5) $\underline{f}(n), \underline{f}'(n), \underline{b}(n), \underline{\mu}(n)$: p-dimensional complex vectors
- (6) $\underline{G}(n)$: N+p-dimensional complex vector
- (7) $\underline{M}(n), \underline{K}(n), \underline{X}(n), \underline{C}(n)$: N-dimensional vectors

To initialize the algorithm we set:

for $n = 0$

$$\underline{C}(0) = \underline{K}(0) = \underline{X}(0) = \underline{0}_{NR} \quad (A-3-1)$$

$$\underline{F}(0) = \underline{B}(0) = \underline{0}_{Np} \quad (A-3-2)$$

$$\underline{E}(0) = \delta \underline{I}_{pp} \quad (A-3-3)$$

$$\gamma(0) = 1 \quad (A-3-4)$$

where $\underline{0}_N$ is the N-dimensional zero vector, $\underline{0}_{Np}$ is the Nxp zero matrix,

$\underline{I}_{pp}$ is the pxp unit matrix and δ is a small positive constant.

for $n \geq 1$

at each time do

begin

$$\underline{f}(n) = \underline{x}(n) - F(n-1)^* \underline{x}(n-1) \quad (A-3-5)$$

$$F(n) = F(n-1) + K(n-1) \underline{f}(n)^* \quad (A-3-6)$$

$$\underline{f}'(n) = \gamma(n-1) \underline{f}(n) \quad (A-3-7)$$

$$E(n) = \lambda E(n-1) + \underline{f}(n) \underline{f}'(n)^*$$

$$\underline{G}(n) = S \left[\frac{E(n)^{-1} \underline{f}'(n)}{K(n-1) - F(n) E(n)^{-1} \underline{f}'(n)} \right] \quad (\text{dimension } N+p \text{ by } p) \quad (A-3-8)$$

$$\begin{bmatrix} \underline{M}(n) \\ \underline{\mu}(n) \end{bmatrix} = Q^T \underline{G}(n) \quad (\text{dimension } N+p \text{ by } p) \quad (A-3-9)$$

$$\underline{b}(n) = \underline{\rho}(n) - B(n-1) \underline{x}(n) \quad (A-3-10)$$

$$B(n) = (B(n-1) + \underline{M}(n) \underline{b}(n)^*) (I_{pp} - \underline{\mu}(n) \underline{b}(n)^*)^{-1} \quad (A-3-11)$$

$$K(n) = \underline{M}(n) + B(n) \underline{\mu}(n)$$

$$\gamma(n) = (\gamma(n-1) - \underline{f}'(n)^* E(n)^{-1} \underline{f}'(n)) (1 - \underline{b}(n)^* \underline{\mu}(n))^{-1} \quad (A-3-12)$$

end

Updating of the equalizer tap coefficient vector $\underline{C}(n)$ then takes place according to (2.1-2) and (2.1-7), using the vector $\underline{K}(n)$.

The $(N+p) \times (N+p)$, Q , S matrices rearrange the components of $\underline{x}(n)$. Their definition can be found in [25]. Thus multiplication by these permutations matrices simply amounts to a permutation of vector components.

For a $T/2$ -spaced DFE $p = 3$. We also can assume $N = 2N_1 + N_2$, i.e., $2N_1$ forward coefficients and N_2 feedback coefficients. In this case, the newest and oldest inputs in the DFE are:

$$\underline{\xi}(n) = \begin{pmatrix} x(nT) \\ x(nT - T/2) \\ \hat{a}_{n-1} \end{pmatrix} \quad \underline{\rho}(n) = \begin{pmatrix} x(nT - N_1 T + T) \\ x(nT - N_1 T + T/2) \\ \hat{a}_{n-N_2} \end{pmatrix} \quad (\text{A-3-13})$$

REFERENCES - 3

- [1] B. Goldberg: "300 KHz - 30 MHz MF/HF", IEEE Trans. on Comm. Technology, Vol.COM-14, pp.767-784, December 1960.
- [2] N.M. Maslin: "High Data Rate Transmission Over HF Links", Radio and Electronic Engineer, Vol.52, pp.75-87, February 1982.
- [3] M. Schwartz, W.R. Bennett and S. Stein: "Communication Systems and Techniques", McGraw-Hill Book Company, 1966.
- [4] J.G. Proakis: "Digital Communications", McGraw-Hill Book Company, 1983.
- [5] A.A. Giordano and F.M. Hsu: "Least Square Estimation With Applications to Digital Signal Processing", Wiley Interscience, 1985.
- [6] "Data Communications Via Fading Channels", edited by K. Brayer, IEEE Press Selected Reprint Series, 1975.
- [7] C.C. Waterson, J.R. Juroshek and W.D. Benbema: "Experimental Confirmation of An HF Channel Model", IEEE Trans. Commun. Technol., Vol.COM-18, pp.792-803, December 1970.
- [8] D.D. Falconer, H.M. Hafez, A.U.H. Sheikh, E. Eleftheriou, M. Tobis: "A Comparative Study of Maximum Likelihood and Adaptive Equalization for High Rate High Frequency Data Communications", Department of Systems and Computer Engineering, Carleton University, Tech. Report, April 1983.
- [9] D.D. Falconer: "Jointly Adaptive Equalization and Carrier Recovery in Two-Dimensional Digital Communication Systems", BSTJ, Vol.55, pp.317-334, March 1976.
- [10] J.G. Proakis: "Advances in Equalization for Intersymbol Interference", in Advances in Communication Systems, Vol.4, pp.123-198, Academic Press, 1975.
- [11] D.D. Falconer, A.U.H. Sheikh, E. Eleftheriou, M. Tobis: "Comparison of DFE and MLSE Receiver Performance on HF Channels", Proceedings Globecom '83, San Diego, December 1983.
also in IEEE Trans. Commun., Vol.COM-33, pp.484-486, May 1985.
- [12] P. Monsen: "Feedback Equalization for Fading Dispersive Channels", IEEE Trans. Inform. Theory, Vol.IT-17, pp.56-64, January 1971.

- [13] P. Monsen: "Theoretical and Measured Performance of a DFE Modem on a Fading Multipath Channel", IEEE, Trans. Commun., Vol.COM-25, pp.1114-1153, October 1977.
- [14] G. Ungerboeck: "Fractional Tap-Spacing Equalizer and Consequences for Clock Recovery in Data Modems", IEEE Trans. Commun., Vol.COM-24, pp.856-864, August 1976.
- [15] S.U.H. Qureshi and G.D. Forney: "Performance and Properties of a T/2 Equalizer", Proceedings NTC '77, December 1977.
- [16] R.D. Gitlin and S.B. Weinstein: "Fractionally Spaced Equalization: An Improved Digital Transversal Equalizer", BSTJ, Vol.60, pp.275-296, February 1981.
- [17] H.E. dePedro, F.M. Hsu, A.A. Giordano, J.G. Proakis: "Signal Design for High-Speed Serial Transmission on Fading Dispersive Channels", Proceedings NTC '78, Alabama, December 1978.
- [18] F.M. Hsu, A.A. Giordano, H.E. de Pedro, J.G. Proakis: "Adaptive Equalization Techniques for High-Speed Transmission on Fading Dispersive HF Channels", Proceedings, NTC '80, Houston, December 1980.
- [19] F.M. Hsu: "Square-Root Kalman Filtering for High-Speed Data Received Over Fading Dispersive HF Channels", IEEE Trans. Inform. Theory, Vol.IT-28, pp.753-763, September 1982.
- [20] F. Ling and J.G. Proakis, "A Generalized Least Squares Lattice Algorithm and Its Application to Decision-Feedback Equalization", Proceedings ICASSP '82, Paris, France, May 1982.
- [21] ----- "Lattice Decision-Feedback Equalizers and Their Application to Fading Dispersive Channels", Proceedings ICC '83, Boston, June 1983.
- [22] F. Ling and J.G. Proakis, "Adaptive Lattice Decision-Feedback Equalizers - Their Performance and Application to Time-Variant Multipath Channels", IEEE Trans. on Commun., Vol.COM-33, pp.348-356, April 1985.
- [23] R. Tront: "Performance of Kalman Decision-Feedback Equalization in HF Data Modems", M.Sc. Dissertation, University of British Columbia, December 1983.
- [24] E. Eleftheriou, and D.D. Falconer: "Restart Methods for Stabilizing FRLS Adaptive Equalizers in Digital HF Transmission", Proceedings Globecom '84, Atlanta, December 1984.

- [25] D.D. Falconer and L. Ljung: "Application of Fast Kalman Estimation to Adaptive Equalization", IEEE Trans. on Commun., pp.1439-1446, October 1978.
- [26] J.M. Cioffi and T. Kailath: "Fast Recursive-Least-Squares Transversal Filters for Adaptive Filtering", IEEE Trans. on Acoustics, Speech and Signal Processing, pp.304-337, April 1984.
- [27] F. Ling and J.G. Proakis: "A Generalized Multichannel Least Squares Lattice Algorithm Based on Sequential Processing Stages", IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-32, pp.381-389, April 1984.
- [28] E.H. Satorious and S.T. Alexander: "Channel Equalization Using Adaptive Lattice Algorithms", IEEE Trans. on Commun., Vol. COM-27, pp.899-905, June 1979.
- [29] J. Salz: "Optimum Mean-Square Decision Feedback Equalization", BSTJ, Vol.52, pp.1341-1372, October 1973.
- [30] J. Salz: "On Mean-Square Decision Feedback Equalization and Timing Phase", IEEE Trans. on Commun., Vol. COM-25, pp.1471-1476, December 1977.
- [31] D.D. Falconer: "Application of Passband Decision Feedback Equalization in Two-Dimensional Data Communications Systems", IEEE Trans. on Commun., Vol. COM-24, pp.1159-1162, October 1976.
- [32] R.W. Lucky, J. Salz, and E.J. Weldon: "Principles of Data Communications", McGraw-Hill Book Company, 1968.
- [33] J.M. Cioffi: "Fast Transversal Filters for Communications Applications", Ph.D. Dissertation, Stanford University, 1984.
- [34] J.M. Cioffi and T. Kailath: "An Efficient, Exact-Least Squares Fractionally Spaced Equalizer Using Intersymbol Interpolations, IEEE Journal of Selected Areas in Comm., Vol. SAC-2, pp.743-756, September 1984.

CHAPTER 4

AN ADAPTIVE MMSE PREDICTOR BASED

DECISION FEEDBACK EQUALIZER

The increasing demand for efficient, fast and more reliable data transmission over a variety of different mediums was and is the inspiration for more sophisticated communication systems to be devised.

The effects of intersymbol interference in data communications has led researchers to consider different receiver structures for its compensation. Chronologically the first and simplest solution to this problem was the linear zero-forcing equalizer [1]. Linear equalization, while almost always effective for telephone channels, can incur a significant noise enhancement penalty when used to eliminate intersymbol interference in more severely distorted channels. Forney's Structure, the so-called Maximum Likelihood Sequence Estimation (MLSE, implemented by the Viterbi Algorithm) receiver [4], while the ultimate solution in coping with intersymbol interference, is unacceptably complicated for implementation in most practical channels. A well known approach by which we obtain some of the performance advantages of the Viterbi Algorithm, but with much lower complexity, is the Decision Feedback Equalizer (DFE).

The idea of subtracting past and future intersymbol interference by utilizing besides a forward linear equalizer, a decision feedback section, dates back to the original work of Austin [2]. Under the complete

elimination of intersymbol interference and the maximization of the signal to noise ratio constraints, Price in [4] has derived the optimum zero-forcing decision feedback equalizer. It was pointed out at the same reference that the front-end of an infinite-length zero-forcing DFE is the same as the whitened matched filter employed in an ideal MLSE receiver [6]. The derivation of the optimum DFE subject to minimization of the mean square error criterion is due to Monsen [5]. Later, Salz in [10] completed Monsen's work by deriving an explicit formula for the minimum MSE. At the same time Falconer and Foschini [11] investigated the application of a DFE receiver in quadrature-amplitude modulation (QAM) transmission. They derived the optimum receiver and joint optimum system for QAM under minimization of the MSE cost function.

In any case, both optimum zero-forcing and MMSE decision feedback equalizers have been shown [4], [5], [7], [10] to consist of a matched filter followed by an anticausal infinite-length forward transversal filter and a causal semi-infinite length feedback transversal filter. Furthermore, it was shown in [4], [5] that as the noise spectral density tends towards zero the two structures become identical.

Messerschmitt in [7] has treated the general equalization problem from a different perspective. He associated the problem of zero-forcing linear equalization and zero-forcing decision feedback equalization to the one of interpolation and prediction of an equivalent stochastic process correspondingly. Using geometrical arguments, he also established the conditions of the existence of both equalization schemes.

As we have already mentioned, computational and storage requirements are one of the chief considerations in any application. The complexity is related not only to the nature of a particular receiver itself, but also to the adaptation method used in any practical system implementation.

The traditional approach in deriving recursively the tap-coefficients for a DFE receiver is joint adaptation of the forward and feedback sections. Defining the error sequence as the difference between the input and output of the decision making device the joint adaptation process usually minimizes iteratively in time the mean square error or the cumulative sum of squared errors. The different input data streams which supply the forward and feedback filters as well as their joint adaptation classify a DFE receiver as a 2-channel (3-channel for fractional implementation) adaptive structure.

We have already discussed in Chapters 2 and 3 the powerful convergence and tracking capabilities of RLS adaptive algorithms. We have also seen the dramatic increase in computational burden arising in multichannel applications. Such an increase is also found in the case of multichannel RLS DFE.

In this chapter we shall introduce a different structure of the classical feedback equalization receiver. We shall show that a MMSE DFE can be realized by using an off-line MMSE linear predictor. The same prediction coefficient can be used to synthesize both forward and feedback transversal filters. At this point we should mention the work by Belfiore and Park [14]. They have shown that a DFE consists of an

optimum linear equalizer followed by an estimator of a random distortion sequence [14]. This approach leads to a DFE based on a conventional forward filter and a predictor feedback filter. The incorporation of the predictor at the feedback section is the major feature distinguishing our scheme from this in [14].

From the implementation point of view our approach lends itself to a separation of the adaptation process. Adapting separately the forward and feedback sections by using different error sequences not only decreases the overall computational complexity, in a RLS context, but also allows a greater flexibility as far as optimization criterion is concerned. Thus, bearing in mind that the feedback section belongs essentially to the class of system-identification structures we notice that a combination of a RLS predictor and a LMS adapting feedback transversal filter will decrease the computational complexity without necessarily affecting initial convergence or tracking properties of the whole structure.

In addition, the off-line prediction philosophy leads to an adaptive implementation of the whitening matched filter. It is shown that the cascade of a matched filter and a transversal filter with coefficients the time-reversed ones of a MMSE predictor form an approximate, in the MMSE sense, adaptive whitening matched filter. The approximation becomes exact as the noise spectrum density approaches zero.

Finally, in the last part we have discussed issues regarding the fractional implementation of the proposed scheme.

The rest of the chapter is organized as follows: Section 4.1 establishes the equivalence of an infinite length predictor-based-DFE with the traditional MMSE DFE. Section 4.2 discusses the adaptive implementation of a finite length predictor based DFE. Section 4.3 presents simulation results. Section 4.4 deals with the difficulties concerning the fractional implementation of the new structure. Finally, Section 4.5 is the summary of the chapter.

4.1 MMSE Prediction and Channel Autocorrelation Sequence Factorization

4.1.1 Definitions

A block diagram of the transmission model under investigation is depicted in Figure 4.1. We have considered a data signal emerging from a linear channel, with complex impulse response $h(t)$, in the presence of noise. The received signal plus noise is sampled every T seconds. The data symbols $\{a_i\}_{i=-\infty}^{+\infty}$ are independent complex random variables transmitted at a rate $1/T$ and take values from a finite alphabet. The additive noise $v(t)$ is a zero-mean white complex random process. The described model is essentially a general pulse-amplitude-modulation (PAM) communications model followed by a symbol rate sampler. In view of Figure 4.1, the resulting received sampled signal will be

$$x(nT) = \sum_k a_k h(nT-kT) + v(nT) \quad (4.1-1)$$

where we have assumed that $h(t)$ is strictly bandlimited to $|f| \leq \frac{1}{2T}$, so that discrete-time and continuous-time systems are equivalent.

Fig. 4.1 PAM communication model followed by a symbol rate sampler

In addition,

$$E\{v(iT) v(jT)^*\} \triangleq \sigma_v^2 \delta_{i-j,0} \quad (4.1-2)$$

and

$$E\{a_i a_j^*\} \triangleq \sigma_a^2 \delta_{i-j,0} \quad (4.1-3)$$

The symbol $\delta_{i-j,0}$ represents the Kronecker delta which is zero for $i \neq j$ and 1 for $i=j$. Furthermore, we have assumed that the channel response is square summable

$$\sum_k |h(kT)|^2 < \infty \quad (4.1-4)$$

Defining a stationary autocorrelation sequence as

$$r_{i-j} \triangleq E\{x(nT-jT) x(nT-iT)^*\} \quad (4.1-5)$$

then using (4.1-1) it can easily be seen that

$$r_{i-j} = \sigma_a^2 \sum_k h(kT) h(kT-(i-j)T)^* + \sigma_v^2 \delta_{i-j,0} = r_{j-i}^* \quad (4.1-6)$$

It can be shown that any sequence whose z-transform is non-negative on the unit circle and also satisfies the discrete form of the Paley-Wiener condition, can be factored into a causal part and an anticausal part [18], [23]. Thus, the sequence $\{r_i\}_{i=-\infty}^{+\infty}$ being an autocorrelation sequence admits this kind of factorization. In particular, $\{r_i\}_{i=-\infty}^{+\infty}$ can be represented as the discrete convolution of a sequence $\{f_i\}_{i=0}^{+\infty}$ (causal part) and a sequence $\{f_{-i}\}_{i=-\infty}^0$ (anticausal part) i.e.,

$$r_i = \sum_{l=0}^{+\infty} f_l f_{l-i}^* \quad (4.1-7)$$

where

$$f_l = 0, \quad l \in (-\infty, -1) \quad (4.1-8)$$

Defining the bilateral z-transform of the sequence r_1 as

$$R(z) = \sum_{i=-\infty}^{+\infty} r_1 z^{-i} \quad (4.1-9)$$

equation (4.1-7) yields

$$R(z) = F(z) F^* \left(\frac{1}{z^*} \right) \quad (4.1-10)$$

where $F(z)$ represents the z-transform of the sequence $\{f_1\}$. Obviously, if $\{r_1\}$ is real then $F^* \left(\frac{1}{z^*} \right) = F(z^{-1})$. The expression (4.1-10) speaks for itself as far as the origin of the term factorization is concerned.

Furthermore, it can be shown [18], [23] that the factorization problem has a unique solution if we further impose the condition that $\{f_1\}_{i=0}^{+\infty}$ is minimum phase. The minimum phase condition is guaranteed if

$$\sum_{i=0}^{\infty} f_1 z^{-i} \neq 0 \quad \text{for } |z| > 1 \quad (4.1-11)$$

and simultaneously $F(z)$ analytic for $|z| > 1$.

4.1.2 MMSE Prediction and Decision Feedback Equalization

Having stated some of the fundamental results on the factorization of an autocorrelation sequence we shall proceed to relate the one-sided minimum phase sequence $\{f_1\}_{i=0}^{+\infty}$ with the coefficients of an optimum discrete-time MMSE predictor.[†] For this reason we shall consider the following problems. Choose coefficients $\{c_i\}_{i=0}^N$ to minimize the mean square error (MSE)

$$E\{|e_p(nT)|^2\} \quad (4.1-12)$$

[†] This differs from reference [4] in that baud-rate channel samples (not from a matched filter) are considered.

where

$$e_p(nT) = \sum_{i=0}^N c_i^* x(nT-iT) \quad (4.1-13)$$

subject to

$$c_0 = 1 \quad (4.1-14)$$

We can see that this problem is exactly one of mean-square linear prediction [20], i.e., predict the present sampled channel output $x(n)$ from the past $\{x(nT-iT)\}_{i=1}^N$ samples. Minimization of (4.1-13) as a means of inverting a minimum phase channel using an one-sided linear equalizer has been considered in [9]. This method obviously does not require knowledge of the data sequence $\{a_n\}$ and in that sense can be considered as a "blind" equalization scheme. Our approach will be different. We shall try by minimizing (4.1-12) to establish a relation between $\{c_i\}_{i=0}^N$ where $N \rightarrow \infty$, and the forward transversal section of a MMSE DFE.

Using (4.1-5), (4.1-6) and (4.1-13) equation (4.1-12) yields:

$$E\{|e_p(nT)|^2\} = r_0 + 2 \operatorname{Re}\left\{\sum_{i=1}^N c_i^* r_1\right\} + \sum_{i=1}^N \sum_{j=1}^N c_i^* c_j^* r_{j-i} \quad (4.1-15)$$

It can easily be shown that the MSE attains its minimum value if and only if the coefficients $\{c_i\}_{i=0}^N$ satisfy the following set of linear equations

$$\sum_{i=0}^N c_i r_{j-i} = 0, \quad c_0 = 1, \quad \forall j \in [1, N] \quad (4.1-16)$$

Some results which are useful for the purpose of our analysis are obtained if the number of prediction coefficients becomes very large, (i.e., $N \rightarrow \infty$).

In this case (4.1-16) yields

$$\sum_{i=0}^{\infty} c_i r_{j-i} = 0, \quad \forall j \in [1, +\infty) \quad c_0 = 1 \quad (4.1-17)$$

Notice that (4.1-17) represents a semi-infinite set of equations with a semi-infinite number of unknowns. Therefore, it cannot be solved with the standard transform techniques. For its solution consider instead of (4.1-17) the following augmented set of equations

$$\sum_{i=-\infty}^{+\infty} c_i r_{j-i} = g_j \quad \forall j \in (-\infty, +\infty) \quad (4.1-18)$$

where the auxiliary sequence $\{g_i\}_{i=-\infty}^{+\infty}$ must satisfy

$$g_i = 0 \quad \forall i \in [1, +\infty) \quad (4.1-19)$$

Also

$$c_i = 0 \quad \forall i \in [-\infty, -1] \quad (4.1-20)$$

If $C(z)$ and $G(z)$ are the bilateral z -transforms of g_i and c_i correspondingly, then

$$C(z) F(z) F^*\left(\frac{1}{z^*}\right) = G(z) \quad (4.1-21)$$

where the factorization result of (4.1-10) has been used. Note that $F(z)$ has all its poles and zeros inside the unit circle and $F^*\left(\frac{1}{z^*}\right)$ has them outside. Since $g_i = 0, i \in [1, +\infty]$ the right hand side of (4.1-21) is analytic only outside of the unit circle. On the other hand, by definition, $C(z)$ can have no singularities in the unit circle. Therefore (4.1-21) is true if and only if the product $C(z) F(z)$ is a constant, i.e.,

$$C(z) F(z) = b$$

(4.1-22)

Applying the causality properties of both c_1 and f_1 sequences, a time domain representation of (4.1-22) will be

$$\sum_{l=0}^{\infty} c_l f_{n-l} = b \delta_{n,0}$$

(4.1-23)

Equation (4.1-23) implies that $c_0 f_0 = b$. Thus using the fact that $c_0 = 1$ we finally get

$$b = f_0$$

(4.1-24)

Therefore (4.1-22) becomes

$$C(z) = \frac{f_0}{F(z)}$$

(4.1-25)

The importance of the above expression is decisive. It indicates that a ~~SE~~ predictor inverts the z-transform of the minimum phase part of the factored autocorrelation sequence $\{r_1\}_{i=-\infty}^{+\infty}$.

In the sequel we shall use the formula (4.1-25) in order to determine the tap-coefficients of the forward and feedback filters of a symbol-spaced MMSE decision feedback equalizer. For this reason we shall consider the receiver front-end consisting of a matched filter and a semi-infinite purely anticausal transversal filter with coefficients $\{c_{-1}\}_{i=-\infty}^{*0}$, i.e., the time-reverse ones of an infinite length predictor. The fact that

$$\{c_{-1}\}_{i=-\infty}^{*0} \xrightarrow{z} C^*\left(\frac{1}{z^*}\right) = f_0^* \left(F^*\left(\frac{1}{z^*}\right)\right)^{-1}$$

(4.1-26)

reveals that this front-end and the one of Mosen's MMSE DFE (see [5],

eq.(6)) are governed within a multiplication constant, by the same equations. Figure 4.2 shows the representation of an infinite length MMSE DFE based on a linear MMSE predictor. This structure and the MMSE DFE described in [5], [10] are equivalent. Notice that if we neglect the noise component of the received signal then the sequence $\{r_1\}_{1=-\infty}^{+\infty}$ becomes solely the channel's autocorrelation sequence. In this case the overall z-transform of the feedforward section will be $\frac{F(z)}{f_0}$ which proves the causal nature of the cascade of the matched filter and the forward transversal filter of a DFE. The feedback filter yields a z-transform $\frac{F(z)}{f_0} - 1$. Subtracting those branches just prior to the decision element will produce a delta function. This behavior is characteristic for a zero-forcing DFE and has been discussed extensively in [7]. Naturally, in practical applications the noise cannot be neglected. Besides, the MSE criterion is more general and robust than the zero-forcing criterion, because the noise presence is taken more effectively into account. Under those conditions the MMSE off-line predictor at the output of the channel produces the spectrum $\frac{f_0}{F(z)}$ by factoring not just the channel autocorrelation spectrum but the combined autocorrelation spectrum of the channel plus noise. Such a factorization is the essence of the equivalence between the proposed scheme and the MMSE DFE in [5], [10].

4.2 Adaptation of the Prediction Based DFE Scheme

In the previous section, we established the equivalence of the MSE DFE with the DFE structure which is based on a MMSE predictor (pred-DFE).

Fig. 4.2 z-transform representation of MMSE DFE based on a MMSE predictor

We have also discussed the condition under which the latter scheme degenerates to the zero-forcing DFE [4], [7]. Our approach was based on infinite length structures and in that sense it is rather unrealistic for practical consideration.

In this section we shall address the issue of adaptive implementation of the pred-DFE by using finite-length filter structures. The forward filter will consist of the cascade of a discrete-time symbol-spaced matched filter and a symbol-spaced transversal filter with coefficients those of a time-reversed predictor. These predictor coefficients can be derived by an off-line forward MMSE linear predictor with let us say N tap-coefficients. The same coefficients can be used in a different arrangement at the feedback filter. This kind of configuration is depicted in Figure 4.3. If $Q(z)$ is the z -transform of $q(n)$, i.e., the output of feedback filter, then for a single delta function as input we shall have

$$Q(z) = - \left(\sum_{i=1}^N c_i z^{-i} \right) (1 + Q(z)) \quad (4.2-1)$$

or

$$Q(z) = \frac{- \sum_{i=1}^N c_i z^{-i}}{1 + \sum_{i=1}^N c_i z^{-i}} = \frac{1}{\sum_{i=0}^N c_i z^{-i}} - 1 \quad (4.2-2)$$

where the fact that $c_0 = 1$ has been used. Hence

$$Q(z) = \frac{1}{C(z)} - 1 = \frac{F(z)}{f_0} - 1 \quad (4.2-3)$$

Equation (4.2-3) is consistent with the transfer function of the feedback filter of an optimum infinite-length DFE we have already discussed in the previous section.

Although the scheme shown in Figure 4.3 seems to be attractive in terms of computation complexity the infinite impulse response (IIR) nature of the feedback filter does not guarantee stable operation in a symbol-by-symbol adaptation of the structure. For this reason we shall resort to a finite impulse response transversal feedback filter. For such an implementation the receiver is completely specified by the off-line forward predictor tap-coefficients $\{c_i\}_{i=1}^{N_1-1}$ and the backward transversal filter coefficients $\{b_i\}_{i=0}^{N_2}$. The discrete-time matched filter can be implemented easily by an off-line channel estimation or identification algorithm. The details of this structure appear in Figure 4.4. Note that the tap-coefficient b_0 plays the role of an automatic gain control system in order to compensate for the gain factor $\frac{\sigma_a^2}{|f_0|^2}$ which appears in the forward section. The most important advantage of the predictor based DFE is the fact that the forward and feedback filters can be adapted separately by using different uncoupled error sequences. Such a split in the adaptation process allows one to use different adaptive algorithms. Thus, in fast start-up or in nonstationary applications a combination of a RLS and a LMS algorithm can be used for the adaptation of the forward predictor and the feedback filter correspondingly. Such an approach is possible since the feedback section is in essence a system identification structure. As we have

Fig. 4.4 An adaptive pred-DFE receiver

already shown (Chapter 2), in system identification applications of short time-varying channels the LMS algorithm can track as well as any one of the more complex RLS algorithms. Hence, using the combination of a single channel RLS and a LMS algorithm we accomplish the same performance without resorting to the much more complex multi-channel RLS algorithms.

Another potential advantage of the proposed DFE scheme is the fact that the predictor does not need past decisions for its adaptation. This property is very important especially in the decision-directed mode where wrong decisions may affect the efficiency of the adaptive receiver.

On the other hand, the necessity of a symbol-rate matched filter could be considered by some as a disadvantage. We note though that a matched filter is nothing more than an off-line channel estimator. Thus its implementation by a simple LMS algorithm will not affect significantly the overall computational burden of the DFE receiver.

Finally we should mention that the cascade of the symbol-rate matched filter and the transversal filter with coefficients the time-reversed ones of the off-line MMSE predictor lends itself to adaptation and consists an approximate form of the whitened matched filter. As the variance of the noise tends towards zero the approximation becomes exact.

4.3 Simulation Results

The previously discussed newbaud-rate DFE scheme was tested on a digital computer. The number of coefficients in the finite impulse response channel is three with an approximate eigenvalue ratio of 419. The discrete-time channel is described by the following transfer function:

$$H(z) = 0.3 + 0.5 z^{-1} + 0.2 z^{-2} \quad (4.3-1)$$

The forward off-line predictor shown in Figure 4.4 has $N_1 = 11$ tap coefficients, and $N_2 = 2$ feedback tap-coefficients. One additional coefficient is required for the AGC unit. The predictor was adapted according to a FRLS algorithm [26], while the feedback section was driven by a LMS adaptation algorithm. An LMS algorithm was also used for the off-line matched filter. Due to the presence of the AGC coefficient at the feedback loop, the adaptation of the feedback section calls for more discussion. If $\underline{B}(n)$ represents the feedback filter tap-coefficients, including the gain factor $b_0(n)$, at time nT , i.e.,

$$\underline{B}(n) = (b_0(n), b_1(n), \dots, b_{N_2}(n))^T \quad (4.3-2)$$

then the tap-coefficients vector is updated as

$$\underline{B}(n+1) = \underline{B}(n) + \mu e(n) \begin{pmatrix} \hat{a}_n \\ \hat{a}_{n-1} \\ \hat{a}_{n-N_2} \end{pmatrix} \quad (4.3-3)$$

where

$$\begin{aligned}
 e(n) &= \hat{a}_n b_0(n) - (z(n) - \sum_{k=1}^{N_2} b_k(n) \hat{a}_{n-k}) \\
 &= \sum_{k=0}^{N_2} b_k(n) \hat{a}_{n-k} - z(n)
 \end{aligned} \tag{4.3-4}$$

and $\hat{a}_n$ is the receiver's most recently decided data symbol. The parameter μ represents the LMS step size, while $z(n)$ is the forward section output at time nT .

Figure 4.5 shows initial convergence of the predictor based DFE and the traditional baud-rate sampled DFE adapted according to a 2-channel FRLS adaptation algorithm. The additive noise variance was set to $\sigma_v^2 = 0.01$. In all simulations the exponential weighting factor of the FRLS adaptation algorithm was fixed to $\lambda = 1.0$.

The averaging of the mean-squared error (MSE), shown in Figure 4.5, at the n th iteration is as follows:

$$MSE(n) = 0.9 MSE(n-1) + 0.1 |e(n)|^2 \tag{4.3-5}$$

For the pred-DFE $e(n)$ is given by (4.3-4) whereas for the traditional DFE the error $e(n)$ is defined as the difference between the decision making element's input and its output. Apparently these two quantities are different. In particular

$$e_{DFE}(n) = \frac{e_{\text{pred-DFE}}(n)}{b_0(n)} \tag{4.3-6}$$

This explains the different steady-state values to which the two adaptive feedback receivers converge. From the plots we observe that the two

Fig. 4.5 MSE, in dB, versus time (initial convergence)

structures converge with about the same rate.

Figure 4.6 shows the steady-state evolution of $1/\text{MSE}(n)$ with time over a period of 6,000 iterations. In order to present a more objective comparison the MSE at the output of pred-DFE has been divided by an estimate of the squared value of the AGC coefficient $b_0(n)$. Again the two receiver structures demonstrate similar mean-square error performances.

4.4 Fractionally-Spaced Implementation Considerations

In our previous discussion we have assumed that the overall channel plus transmit filter impulse response $h(t)$, has a bandwidth restricted in the Nyquist band, $|f| \leq \frac{1}{2T}$. Thus, inversion of the minimum-phase-factored part of the autocorrelation spectrum by a symbol spaced predictor placed at the output of the channel would not introduce sensitivity to sampling phase.

Under more realistic conditions though a small percentage of excess bandwidth (i.e., the Fourier transform extends beyond $\frac{1}{2T}$ Hz) is unavoidable for the pulse $h(t)$. In this case the off-line predictor, being placed before the matched filter will factor the aliased autocorrelation spectrum causing performance degradation. This degradation is evident from the dependance on the timing epoch of the mean squared error (MSE) at the output of the pred-DFE.

Despite the fact [27], [28] that the MSE of a DFE is less sensitive to timing phase than that of a linear equalizer, there is always the

Fig: 4.6 $1/\text{MSE}$, in dB, versus time
(steady state behaviour)

possibly that a badly-chosen timing phase will affect the receiver's performance.

The natural question is: How will the performance of such a predictor based DFE be independent of the receiver sampling phase? An off-hand answer would be to use a fractional spaced forward predictor with tap-spacing less than the inverse of twice the highest frequency of the baseband signal. In the sequel we shall illuminate the difficulties associating with the adaptation of such a fractional spaced pred-DFE.

Consider the transmission model shown in Figure 4.1, where now the channel output signal is sampled twice the baud rate. Thus

$$x(n\frac{T}{2}) = \sum_k a_k h(n\frac{T}{2} - kT) + v(n\frac{T}{2}) \quad (4.4-1)$$

Note that the definitions (4.1-2) to (4.1-4) are holding again. In such a case, a $T/2$ -sampled channel autocorrelation function can always be achieved by defining the following stationary autocorrelation sequence, i.e.,

$$r_{i-j} \triangleq \frac{1}{2} \{E\{x(n\frac{T}{2} - j\frac{T}{2})x(n\frac{T}{2} - i\frac{T}{2})\} + E\{x(n\frac{T}{2} - j\frac{T}{2} - \frac{T}{2})x(n\frac{T}{2} - i\frac{T}{2} - \frac{T}{2})^*\}\} \quad (4.4-2)$$

Using (4.4-1) and (4.4-2) we can easily show that

$$r_{i-j} = \frac{1}{2} \sigma_a^2 \sum_k h(k\frac{T}{2}) h(k\frac{T}{2} - (i-j)\frac{T}{2})^* + \sigma_v^2 \delta_{i-j,0} \quad (4.4-3)$$

Hence, the question becomes how we efficiently can factor adaptively the above defined sequence.

The simplistic solution of factoring the autocorrelation spectrum by a $T/2$ forward predictor would lead to the following set of equations:

$$\sum_{i=0}^N c_i r_{i,j} = 0, \quad c_0 = 1, \quad \forall j \in [1, N] \quad (4.4-4)$$

Observe that in this case

$$r_{i,j} \triangleq E\{x(n\frac{T}{2} - j\frac{T}{2}) x(n\frac{T}{2} - i\frac{T}{2})^*\} \quad (4.4-5)$$

is no longer a stationary sequence and thus the $T/2$ predictor coefficients only approximately invert the combined autocorrelation spectrum of the $T/2$ -sampled impulse response plus noise. Furthermore, although oversampling has removed the problem "aliasing", lack of synchronization between the every $T/2$ seconds adapting forward predictor and the once per symbol interval updated feedback section does not allow the receiver to compensate adaptively for any sampling phase.

This problem could be solved by using 2-dimensional predictor structures. In this case the predictor will be updated once per baud while each new entry will be a 2×1 vector quantity. Unfortunately, such a scheme does not offer any computational complexity advantage over the traditional 3-channel DFE. Further work is needed in order to find ways of efficient adaptation of a fractional pred-DFE.

4.5 Summary

We have shown analytically that the forward section of an optimum MMSE DFE can be realized by cascading a matched filter and a transversal

filter with coefficients the time-revised ones of an off-line forward MMSE predictor.

The efficiency of an adaptive T-spaced pred-DFE in data systems which make efficient use of the available bandwidth by using sharp roll-off filters has been demonstrated by computer simulations.

Additionally, the off-line predictor concept gives the means for a good approximation of an adaptive whitened matched filter.

Finally, in the last part we have discussed the difficulties associated with the adaptation of a $T/2$ spaced predictor based DFE.

REFERENCES - 4

- [1] R.W. Lucky, J. Salz, and E.J. Weldon: "Principles of Data Communication", McGraw-Hill, Inc., New York, 1969.
- [2] M.E. Austin: "Decision-Feedback Equalization for Digital Communication Over Dispersive Channels", Tech. Report 461, Res. Lab. Electronics, MIT, August 1967.
- [3] D.A. George, R.R. Bowen, J.R. Storey: "An Adaptive Decision Feedback Equalizer", IEEE Trans. Commun. Tech., COM-19, pp.281-293, June 1971.
- [4] R. Price: "Nonlinearity Feedback-Equalized PAM vs. Capacity for Noisy Filter Channels", Proceedings ICC '72, Philadelphia, 1972.
- [5] P. Monsen: "Feedback Equalization for Fading Dispersive Channels", IEEE Trans. Inform. Theory, IT-17, pp.56-64, January 1971.
- [6] G.D. Forney: "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference", IEEE Trans. Inform. Theory, IT-18, pp.363-378, May 1972.
- [7] D.G. Messerschmitt: "A Geometric Theory of Intersymbol Interference, Part I", BSTJ, Vol.59, pp.1483-1518, November 1973.
- [8] D.G. Messerschmitt: "Design of a Finite Impulse Response for the Viterbi Algorithm and Decision Feedback Equalizer", Proceedings ICC '74.
- [9] J.B. Allen and J.E. Mazo: "A Decision-Free Equalization Scheme for Minimum Phase Channels", IEEE Trans. on Comm., pp.1732-1733, October 1974.
- [10] J. Salz: "Optimum Mean-Square Decision Feedback Equalization", BSTJ, Vol.52, pp.1341-1373, October 1973.
- [11] D.D. Falconer and G.J. Foschini: "Theory of Minimum Mean Square-Error QAM Systems Employing Decision Feedback Equalization", BSTJ, Vol.52, pp.1821-1849, November 1973.
- [12] E.A. Robinson: "Multichannel Time Series Analysis With Digital Computer Programs", Holden-Day, San Francisco, 1967.
- [13] W.U. Lee and F.S. Hill, "A Maximum-Likelihood Sequence Estimator With Decision Feedback Equalization", IEEE Trans. on Comm., Vol.COM-25, pp.971-979, September 1977.

- [14] C.A. Belfiore and J.H. Park: "Decision Feedback Equalization", Proceedings of IEEE, Vol.67, pp.1143-1156, August 1979.
- [15] G. Ungerboeck: "Adaptive ML Receiver for Carrier-Modulated Data-Transmission Systems", IEEE Trans. Commun. Technol., Vol.COM-22, pp.624-636, May 1974.
- [16] S.U.H. Qureshi and E.E. Newhall: "An Adaptive Receiver for Data Transmission Over Time-Dispersive Channels", IEEE Trans. Inform. Theory, Vol.IT-19, pp.448-457, July 1973.
- [17] D.D. Falconer and F.R. Magee: "Adaptive Channel Memory Truncation for Maximum Likelihood Sequence Estimation", BSTJ, Vol.52, pp.154-1562, November 1973.
- [18] A. Papoulis: "Signal Analysis", McGraw-Hill Company, 1972.
- [19] A.V. Oppenheim and R.W. Schaffer: "Digital Signal Processing", Prentice-Hall, 1975.
- [20] J. Makhoul: "Linear Prediction: A Tutorial Review", Proceedings of IEEE, Vol.63, pp.561-580, April 1975.
- [21] J. Makhoul: "A Class of All-Zero Lattice Digital Filters: Properties and Applications", IEEE Trans. on Acoust. Speech and Signal Proc., Vol.ASSP-26, pp.304-314, August 1978.
- [22] U. Grenander and M. Rosenblatt: "Statistical Analysis of Stationary Time Series", John Wiley and Sons, 1957.
- [23] U. Grenander and G. Szegő: "Toeplitz Forms and Their Applications", University of California Press, 1958.
- [24] M.S. Mueller and J. Salz: "A Unified Theory of Data-Aided Equalization", BSTJ, Vol.60, pp.2023-2038, November 1981.
- [25] Ib. N. Andersen: "Sampled-Whitened Matched Filters", IEEE Trans. on Inform. Theory, Vol.IT-19, pp.633-660, September 1973.
- [26] D.D. Falconer and L. Ljung: "Application of Fast Kalman Estimation to Adaptive Equalization", IEEE Trans. on Commun., pp.1439-1446, October 1978.
- [27] J. Salz: "On Mean-Square Decision Feedback Equalization and Timing Phase", IEEE Trans. on Commun., Vol.COM-25, pp.1471-1476, December 1973.
- [28] D.D. Falconer: "Application of Passband Decision Feedback Equalization in Two-Dimensional Data Communication Systems", IEEE Trans. on Commun., Vol.COM-24, pp.1159-1162, October 1976.

- [29] R.D. Gitlin and S.B. Weinstein: "Fractionally-Spaced Equalization: An Improved Digital Equalizer", BSTJ, Vol.60, pp.275-296, February 1981.

CHAPTER 5

CONCLUSIONS

5.1. Concluding Remarks

The principal focus of this thesis was on the steady-state performance characteristics of the wide class of RLS adaptive algorithms. Our work provides a mathematical treatment of the topic giving insight on properties which sometimes had been conjectured from experimental evidence. The "estimation-noise" and especially the "lag-error" formulas are analytical expressions of the behaviour of the RLS adaptation algorithm in time-invariant as well as time-varying media correspondingly. The lag error analysis proves, in general, the superior tracking capabilities of the RLS adaptive filter. It also sets the conditions under which the simple LMS algorithm exhibits tracking properties equivalent to the more complex RLS algorithms. Equations (2.2-15), (2.2-21), (2.3-7) and (2.3-11) equip us with the necessary knowledge regarding the tradeoffs involved in selecting the appropriate algorithm in the different areas of application.

Transversal FRLS adaptive structures although more prone to finite precision error problems, seem to be more attractive, especially in communication applications, because their computational complexity is lower than that of equivalent lattice filter adaptation algorithms. The novel restart procedures we proposed guarantee stable propagation

of the algorithm without significantly affecting its capability to track an unknown time-varying parameter vector.

Another essential contribution was the in depth study of adaptive equalization techniques in narrowband serial data transmission over the extremely difficult ionospheric channel. The superior tracking properties as well as the high fade rate limitations of the fractional spaced FRLS decision-feedback equalizer have been demonstrated experimentally on simulated and real HF data. In conjunction to the HF equalization techniques a novel whitening prefilter approach towards improving the convergence and tracking ability of an LMS DFE has also been presented.

Finally, this thesis has introduced a new formulation of the traditional decision feedback equalizer. The key idea is the use of an off-line forward predictor for estimating the coefficients of the anticausal forward tapped delay line. This allows a separation of the adaptation process. Such an approach leads to a more efficient adaptation of the DFE receiver in terms of computational complexity.

5.2 Suggestions for Future Work

Below are outlined suggestions for future study.

- (1) Although restart methods have been proven to be effective in improving numerical robustness of the transversal FRLS algorithms, fully understanding of the finite precision problem has not yet been accomplished. Generally, normalized versions of the FRLS transversal algorithms considerably improve its numerical stability properties. Rescue operations though are still necessary even if invoked much

less frequently. So the question remains if it is possible to completely stabilize the algorithm or to find other simple suboptimum schemes with probably better tracking response in applications involving dynamically changing environments.

- (2) Another area open to future research is computational reduction associated with the FRLS "multichannel" equalizers. As such fractional spaced and decision feedback equalizers could be considered as prime candidates for future exploration. The off-line predictor approach we proposed is a meaningful start towards this direction. But further work is needed in finding efficient adaptation ways for its fractional implementation.
- (3) Our study on the decision-directed mode of the adaptive DFE in HF fading channels, revealed that periodic resynchronization of the adaptive receiver with an ideal reference is necessary and appears to be a practical measure for fading HF channels. Thus an effective method for crash recovery could be transmission of data in blocks preceded by known training sequences. This block or packet transmission strategy lends itself to an alternative receiver signal processing; the block-by-block. In the block-by-block processing, the tap coefficients of the adaptive DFE can be computed by fast autocorrelation or covariance algorithms after the autocorrelation or covariance matrix has been estimated from samples received in a predetermined time interval. Such an approach could be proven to be more effective than the symbol-by-symbol adaptation scheme especially at high fade rates.

END

14.05.87

FIN