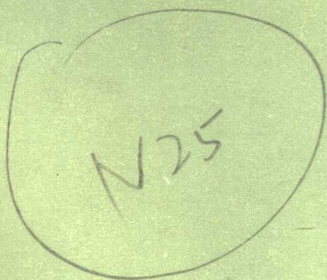


***Least Square Estimation
with Applications to
Digital Signal Processing***

Arthur A. Giordano

Frank M. Hsu



***Least Square Estimation
with Applications to
Digital Signal Processing***

Arthur A. Giordano

Frank M. Hsu

***A WILEY-INTERSCIENCE PUBLICATION
JOHN WILEY & SONS***

New York • Chichester • Brisbane • Toronto • Singapore

Copyright © 1985 by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by Section 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc.

Library of Congress Cataloging in Publication Data:

Giordano, A. A. (Arthur Anthony), 1941-

Least square estimation with applications to digital signal processing.

"A Wiley-Interscience publication."

Includes index.

1. Least squares—Data processing. 2. Estimation theory—Data processing. 3. Signal processing—Digital techniques. I. Hsu, F. M. (Frank Ming), 1946-
II. Title.

QA275.G56 1985 621.38'043 84-19496

ISBN 0-471-87857-X

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Preface

Least squares error techniques were devised independently by Gauss and Legendre in the early 1800s as a method for estimating parameters from noisy measurements. These techniques initiated the development of what is now a voluminous body of mathematical and scientific literature describing investigations on an extensive variety of least square error applications. Each new application seems to spawn its own theoretical formulation. However, in many instances least square error principles with geometrically based foundations provide a unifying thread among seemingly unrelated problems. A dual objective of this text, then, is to establish the mathematical framework of least square error principles, and to subsequently demonstrate the utility and widespread use of these principles in a variety of digital signal processing applications.

Important questions regarding the purpose and objectives of this book require further clarification, specifically:

1. Why is this book important?
2. What can be found in this book that is unavailable elsewhere?
3. What circumstances justify a treatise on this subject at this time?
4. Who is the intended audience?

Although these questions are related, they will be answered individually, beginning with the third question.

Digital computers and real-time digital processors have forced dramatic changes in such diverse scientific disciplines as communications, control, radar, seismology, bioelectronics, etc. Least square error algorithms, which were originally developed to process data, often involve an extensive amount of iterative computation. Thus, special-purpose digital signal processors, programmable digital signal processors and/or digital computers are ideally suited for implementing least square error algorithms. In addition, the research and

development of least square error algorithms, which previously had been constrained by a scientist's perseverance in performing hand computations, have been enhanced and extended by the computational power and speed of modern digital signal processors and digital computers. As a result, new applications and new least square error algorithms supporting a variety of applications are now possible that, until the last decade, could not have been conceived.

In response to the second question, this book documents important least square error algorithms in a unified way, with consistent notation. Both deterministic and probabilistic formulations are presented in a geometrical framework. The subject matter treated here is typically scattered in the available literature, or is developed within the context of a narrow specific application. By presenting these algorithms within a single reference source, unique and/or common features associated with the various algorithms can be identified. In Part I, a mathematical formulation of the least square error algorithms is provided. In Part II, certain digital signal processing applications that have achieved widespread use are selected as examples. One goal of this book, then, is to offer the reader an opportunity to comprehend the theory in a manner which permits a transfer of the technology to the specific application at hand.

The importance of the book (by way of an answer to the first question) can be summarized as follows: (1) The generality and widespread applicability of least square error algorithms in digital signal processing is not well known; (2) a consistent unified treatment of least square error algorithms, which permits the reader to implement these algorithms in hardware and/or software, has heretofore not been available. Many numerical examples which can be followed by hand computation are used in the book to help the reader understand detailed computational procedures required for least square error algorithms.

The fourth and final question posed above concerns the definition of the intended audience. This book is specifically written for practicing engineers and scientists involved in digital signal processing and for advanced students interested in digital signal processing. Desirable prerequisites include courses in matrix algebra, probability and stochastic processes, and digital signal processing.

The structure of the book is as follows: Chapters 2–5 (Part I) present the least square error algorithms. The remaining chapters (Part II) cover the digital signal processing applications. In Chapter 2 a Fourier-series expansion using orthogonal functions is presented. The Fourier-series coefficients are derived both by differentiating the mean square error and by applying the orthogonality principle. Subsequently, geometric concepts are presented to introduce least square optimization in Hilbert space. Both the orthogonality principle and the Gram-Schmidt orthogonalization procedure are then used to derive the normal equations. In Chapter 3 the Durbin, Levinson and Burg algorithms are derived. In the derivations a digital communications model is assumed using signals with either known correlation functions or correlation functions which can be estimated from the observed data. The Durbin algorithm provides a recursive

solution to the Yule-Walker equations. The Levinson algorithm provides a recursive solution to the normal equations that is referred to as a *Wiener filter*. The Burg algorithm provides a solution using forward and backward error prediction with an implementation which assumes a lattice structure form. In Chapter 4 a general least square lattice algorithm that is recursive in both time and order is described for the nonstationary signal case. In Chapter 5 the Kalman recursive least square estimation algorithm is derived by generalizing the minimum-variance-weighted-least-square method. A computationally stable form of the Kalman algorithm, known as the square root Kalman algorithm, concludes the chapter.

Part II describes applications involving equalization, spectral analysis, digital whitening, adaptive arrays, interchannel interference mitigation, and digital speech processing. The first application, equalization, receives the greatest coverage, since many of the algorithms presented in Part I are utilized. The discussion of equalization algorithms also provides an opportunity to identify a relationship between the Kalman algorithm and generalized least squares lattice structure formulations. Spectral analysis, treated in Chapter 7, considers the application of autoregressive methods, such as the Burg algorithm, for attaining high-resolution spectral estimates. In Chapter 8, which deals with digital whitening, use of the Durbin implementation of a Wiener filter and the Burg implementation of a maximum entropy filter result in substantial performance improvement by suppressing narrowband interference in spread-spectrum communications. The chapters on adaptive arrays and interference mitigation (Chapters 9 and 10) describe techniques for combining multiple signals using minimum mean square error methods to process digital signals efficiently and reduce distortion from interference. Chapter 11, dealing with speech processing, presents applications of linear prediction theory and efficient time-domain waveform coding schemes. These applications represent examples of the use of least square algorithms in digital signal processing and are not intended to be exhaustive.

We would like to express our appreciation to General Telephone and Electronics (GTE) for giving us the opportunity to work on the variety of problems presented in the applications part of the text. We are also indebted to GTE's support during manuscript preparation. We would like to thank numerous colleagues who contributed to the original research at GTE, including P. Anderson, Dr. A. Levesque, J. Lindholm, Dr. H. Nichols, Dr. H. dePedro, M. Sandler, Dr. T. Schonhoff and Dr. John Proakis of Northeastern University. We owe special thanks to Dr. David Freeman, who reviewed the entire manuscript and made many worthwhile suggestions. We are also indebted to L. Carroll for her careful editing and typing of the original manuscript.

ARTHUR A. GIORDANO
FRANK M. HSU

Notation

Frequently used symbols required in the text are provided in this section.*† Every attempt has been made to use a consistent notation. Often, subscripts and index parameters are the only means of distinguishing symbols. Unfortunately, very little consistency exists in the literature as a result of the wide variety of researchers and applications. Thus, care should be exercised in comparing algorithms presented here with references available elsewhere.

Among the characters and operators used in the text are

$\mathcal{E}(x)$	ensemble average of random variable x
$V(x)$	variance of random variable x or power of x ($\sigma_x^2 = V(x)$ in Chapters 1–8, 10, 11)
\hat{x}	estimate of variable x
$\{x_k\}$	set of variables x_k for k ranging over a specified interval
x^*	complex conjugate of variable x
X^t	transpose of vector, [†] X
$\delta(t)$	impulse function defined by $\int_{-\infty}^{\infty} \delta(t) dt = 1$ and $\delta(t) = 0$ for $t \neq 0$
$\delta(l)$	unit sample function $= \begin{cases} 1, & l = 0 \\ 0, & l \neq 0 \end{cases}$
P_{\min}	minimum value of the quantity P

*Vectors and matrices appear in boldface print

†All vectors are defined as column vectors; for example, if X has elements x_1, \dots, x_N , then

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}.$$

A consistent definition is maintained throughout the text for the following symbols, with the exceptions noted

a_k	prediction error coefficients in Chapters 3–11, scalars in Chapter 2
$a_{M,k}$	prediction error coefficients with order index
$a_M(m, N)$	forward prediction error coefficients
$A_M(N)$	vector of forward prediction error coefficients, that is, $A_M'(N) = (a_M(1, N), \dots, a_M(M, N))$
$\bar{A}_M(N)$	extended vector of forward prediction error coefficients, that is, $\bar{A}_M'(N) = (1, a_M(1, N), \dots, a_M(M, N))$
$\alpha_M(N)$	scalar in lattice structure
α_k	DFE forward coefficients in Chapter 6
b_k	linear weights or prediction coefficients
$b_{M,k}$	linear weights or prediction coefficients with order index
$b_M(m, N)$	backward prediction error coefficients
$B_M(N)$	vector of backward prediction error coefficients, that is, $B_M'(N) = (b_M(0, N), \dots, b_M(M-1, N))$
$\bar{B}_M(N)$	extended vector of backward prediction error coefficients, that is, $\bar{B}_M'(N) = (1, b_M(0, N), \dots, b_M(M-1, N))$
B	vector of Kalman prediction coefficients, that is, $B' = (b_1, \dots, b_M)$
B_k	vector of Kalman prediction coefficients, that is, $B_k' = (b_{1,k}, \dots, b_{M,k})$
$\hat{B}_{k,k-1}$	predicted value of current Kalman coefficient estimate \hat{B}_k
β_k	DFE feedback coefficients in Chapter 6
C_k	Fourier-series coefficients
c_k	linear equalizer coefficients in Chapter 6
\tilde{C}_k	vector of feedforward and feedback equalizer coefficients, that is, $\tilde{C}_k' = (\alpha_0, \dots, \alpha_{M_1}, \beta_1, \dots, \beta_{M_2})$ (at time instant k)
C_k	vector of linear equalizer coefficients (at time instant k), that is, $C_k' = (c_0, \dots, c_M)$

$C_M(N)$	vector of feedforward and feedback equalizer coefficients, that is, $C_M'(N) = (C_M(0, N), \dots, C_M(M, N))$
$C_{M_T}(N)$	vector of feedforward and feedback lattice equalizer coefficients, that is, $C_{M_T}'(N) = (C_{M_T}(0, N), \dots, C_{M_T}(M, N))$
δ_k	zero mean random vector for Kalman algorithm with M elements
$D_M(k, N)$	lattice structure vector with M elements
D_k	square root Kalman $N \times N$ diagonal matrix with elements $d_i(k)$
e_k	error signal between received signal and its estimate
ϵ_k	error signal between desired signal or information symbol and its estimate
$\left. \begin{matrix} e_M^f(k, N) \\ e_M^f(N) \end{matrix} \right\}$	forward error signal
$\left. \begin{matrix} e_M^b(k, N) \\ e_M^b(N) \end{matrix} \right\}$	backward error signal
$\left. \begin{matrix} e_M(k, N) \\ e_M(N) \end{matrix} \right\}$	error between information symbol and its estimate
$E_M^f(N)$	minimum forward MSE power vector with $M + 1$ elements, that is, $E_M^f(N) = (f_M(N), 0, \dots, 0)$
$E_M^b(N)$	minimum backward MSE power vector with $M + 1$ elements, that is, $E_M^{b'}(N) = (0, \dots, 0, r_M(N))$
$e_M^f(k, N)$	vector of forward error signals for DFE lattice structure, that is, $e_M^{f'}(k, N) = (e_{1,M}^f(k, N) e_{2,M}^f(k, N))$
$e_M^b(k, N)$	vector of backward error signals for DFE lattice structure, that is, $e_M^{b'}(k, N) = (e_{1,M}^b(k, N) e_{2,M}^b(k, N))$
E	Kalman error vector, that is, $E^t = (\epsilon_1, \dots, \epsilon_N)$
E_k	Kalman error vector, that is, $E_k^t = (\epsilon_{1k}, \dots, \epsilon_{Nk})$
E_b/N_0	ratio of signal energy per bit to noise power spectral density or energy contrast ratio

xx NOTATION

$E(C)$	linear equalizer MSE (ensemble average)
$E(\alpha, \beta)$	DFE MSE (ensemble average)
E	time average of continuous squared error signal
$f_m(N)$	minimum forward MSE power
f_k	channel coefficients in Chapter 6
$\gamma_M(k, N)$ $\gamma_M(N)$	} scalar in lattice structure
g_k	
	crosscorrelation between received and transmitted signals (ensemble average)
	crosscorrelation between received sample and information symbol in Chapter 6
G_k	$M \times N$ Kalman gain matrix
$G_M(N)$	$M + 1$ element Kalman gain vector with order index
Γ_k	gradient vector for steepest descent algorithms with $M_1 + M_2 + 1$ elements
I_k	k th information symbol
\tilde{I}_k	k th information symbol decision
$\hat{I}_M(k, N)$ \hat{I}_k	} estimate of k th information symbol
$k_M(N)$	
	scalar in lattice structure
K_k	Kalman gain vector, that is, $K_k^T = (K_{1,k}, \dots, K_{N,k})$
L_f	loss function
M	order of prediction filter related to number of linear equalizer tap coefficients in Chapter 6, that is, either $M + 1$ or $2M + 1$ total number of taps
M_1	number of feedforward DFE coefficients minus one
M_2	number of feedback DFE coefficients
M_T	total number of DFE lattice equalizer coefficients, that is, $M_T = M_1 + M_2 + 1$
M_a	alphabet size
$\mu_M(N)$	scalar in lattice equalizer structure
N	total number of signal samples
n_k	additive noise samples

$\eta(t)$	additive channel noise
N_0	noise power spectral density
$\Phi(k, k-1)$	$M \times M$ transition matrix for Kalman coefficients dynamic model
$p(t)$	transmit filter pulse with symbol duration
P_M	ensemble average MSE for either received or desired signals
$P_M(N)$	time average MSE
$P_M^b(N)$	time average backward error power
$P_M^f(N)$	time average forward error power
P_k	$M \times M$ error covariance matrix of Kalman prediction coefficients
$P_{k,k-1}$	$M \times M$ Kalman prediction error covariance matrix
$P(f)$	power spectral density of received signal
$q(N)$	zeroth time average autocorrelation element, that is, $q(N) = \rho_N(0, 0)$
$Q_M(N)$	vector of time average autocorrelation coefficients, that is, $Q_M'(N) = (\rho_N(1, 0), \dots, \rho_N(M, 0))$
Q_k	$M \times M$ covariance matrix of Kalman random vector δ_k
$\rho_N(m, j)$	time average autocorrelation of received signal in prewindowed method (in some instances the weighting coefficient $\omega = 1$)
r_k	ensemble average received signal autocorrelation coefficients
R_M	$M \times M$ covariance matrix of received signals (ensemble average)
$R_M(N)$	$M \times M$ time average autocorrelation matrix of received signals
R_e	$N \times N$ Kalman covariance matrix of error signals (ensemble average)
R_{k_e}	$N \times N$ Kalman covariance matrix of error signals (ensemble average)
$r_M(N)$	minimum backward MSE power
s_k	desired or transmitted signal samples
$s(t)$	transmitted signal
S	Kalman vector of transmitted signal samples, that is, $S' = (s_1, \dots, s_N)$
S_k	Kalman vector of transmitted signal samples, that is, $S_k' = (s_{1,k}, \dots, s_{N,k})$

xxii NOTATION

T	symbol interval or duration
T_s	sampling interval
T_c	chip interval
T_0	period
T_i	time interval or integration interval
U_k	$N \times N$ square root Kalman upper triangular matrix with elements $u_{ij}(k)$
$v_M(N)$	M th time average autocorrelation coefficient, that is, $v_M(N) = \rho_N(M, M)$
$V_M(N)$	vector of time average autocorrelation coefficients, that is, $V_M'(N) = (\rho_N(0, M), \dots, \rho_N(M-1, M))$
ω	weighting coefficient in Chapters 4 and 6
$W_M(N)$	time average crosscorrelation vector between information symbols and received sample, that is, $W_M'(N) = (w_0, \dots, w_M)$
$x(i)$	received signal
x_k	received or observed signal samples
$\left. \begin{matrix} X_M(k) \\ X(k) \end{matrix} \right\}$	received signal vector, that is, $X_M'(i) = (x_i, \dots, x_{i-M})$ (In DFE lattice $X_M(i)$ is the vector of received samples and information symbols.)
$X_{M_T}(k)$	vector of received samples and information symbol decisions for DFE lattice with M_T elements
X_M	received signal vector, that is, $X_M' = (x_1, \dots, x_M)$
$\left. \begin{matrix} X \\ X_k \end{matrix} \right\}$	$N \times M$ Kalman matrix of received signal samples
$X_k(i)$	Kalman vector of received signal samples, that is, $X_k'(i) = (x_{i,k}, \dots, x_{i-(M-1),k})$
\bar{X}_k	square root Kalman vector of received signal samples, that is, $\bar{X}_k' = (x_{1,k}, \dots, x_{N,k})$
X_k'	vector of received signal samples, that is, $X_k'' = (x_k, \dots, x_{k+M})$

$\tilde{\mathbf{X}}_k$ vector of received samples and information symbol decisions in Chapter 6

$$\tilde{\mathbf{X}}_k^t = (x_k, \dots, x_{k+M_1}, \tilde{I}_{k-1}, \dots, \tilde{I}_{k-M_2})$$

NOTES

1. MSE = mean square error
2. DFE = decision feedback equalizer
3. Subscripts on received signal are used to distinguish Kalman and lattice structures. The time index k is used in the Kalman case and the order index M is used in the lattice case.

Contents

NOTATION	xvii
1 INTRODUCTION	1
1.1 Basis of Least Square Theory	2
1.2 Summary of Least Square Algorithms and Associated Applications	3
1.3 Estimation by Least Squares and other Methods	7
1.3.1 MAP Estimation, 9	
1.3.2 ML Estimation, 9	
Part 1 Mathematical Formulations of the Least Square Error Algorithms	
2 INTRODUCTION TO LEAST SQUARES THEORY	15
2.1 Classical Least Squares	15
2.1.1 Derivation of Fourier Series Coefficients by Differentiation, 16	
2.1.2 Derivation of Fourier Series Coefficients by Orthogonality Principle, 18	
2.1.3 Evaluation of the Minimum Average Square Error, 18	
2.1.4 Orthogonal Function Expansions, 19	
2.2 Least Squares Optimization in Hilbert Space	20
2.2.1 Hilbert Space Definition, 20	
2.2.2 Generalized Orthogonal Function Expansion, 24	
	xi

2.3	Orthogonality Principle in Hilbert Space	25
2.3.1	Geometric Interpretation of Orthogonality Principle, 25	
2.3.2	Gram-Schmidt Orthogonalization Procedure, 27	
2.3.3	Formal Theory of Optimization in Hilbert Space, 28	
3	LEAST SQUARE ESTIMATION ALGORITHMS—PART I	40
3.1	Durbin Algorithm	40
3.1.1	Model Description, 40	
3.1.2	Derivation of Yule-Walker Equations by Differentiation, 42	
3.1.3	Derivation of Yule-Walker Equations by Orthogonality Principle, 43	
3.1.4	Durbin Recursive Solution, 44	
3.2	Levinson Algorithm	49
3.2.1	Model Description, 49	
3.2.2	Derivation of Normal Equations by Differentiation, 49	
3.2.3	Derivation of Normal Equations by the Orthogonality Principle, 50	
3.2.4	Levinson Recursive Solution, 51	
3.3	Cholesky Decomposition Algorithm	58
3.4	Autocorrelation Coefficients Computational Methods	63
3.5	Forward and Backward Prediction	66
3.5.1	Model Description, 66	
3.5.2	Lattice Structure with Known Autocorrelation Coefficients, 69	
3.5.3	Burg Algorithm, 70	
4	LEAST SQUARE ESTIMATION ALGORITHMS—PART II	80
4.1	Recursive Least Squares Lattice Algorithms	80
4.1.1	Model Description, 80	
4.1.2	Order Update for Coefficients, 84	
4.1.3	Order Update for Minimum Mean Square Error, 90	
4.1.4	Error Updates, 91	
4.1.5	Time Update, 93	

4.1.6	Summary of Recursive Least Squares Lattice Algorithm, 93	
4.1.7	Lattice Structure Development, 95	
4.1.8	Specialization of Lattice Structure to Burg Algorithm, 98	
Appendix	Derivation of the Time Update Recursion	102
5	LEAST SQUARE ESTIMATION ALGORITHMS—PART III	109
5.1	Simple Least-Squares Estimation	109
5.2	Minimum Variance Weighted Least Squares (Gauss-Markov)	111
5.3	Minimum Variance Least Squares (Kalman)	114
5.3.1	Kalman Recursive Estimation Algorithm, 117	
5.3.2	Derivation of Kalman Recursive Estimation Algorithm Using the Loss Function, 120	
5.4	Square Root Kalman Algorithm	126
5.4.1	U-D Factorization, 126	
5.4.2	Mathematical Development of U-D Kalman Algorithm, 126	
5.4.3	Summary of U-D Kalman Algorithm, 135	
Part 2	Application of Least Square Analysis to Digital Signal Processing	
6	EQUALIZATION	139
6.1	Discrete Model of a Continuous-Time Communication System	139
6.2	Wireline and Radio Channels	143
6.3	Equalization Techniques	147
6.3.1	Linear Equalization Using Transversal Filters, 148	
6.3.2	Decision-Feedback Equalization, 150	
6.3.3	Adaptive Equalization Techniques, 152	
6.3.3.1	Steepest Descent Method, 153	
6.3.3.2	Lattice Structure Method for Linear Equalization, 154	
6.3.3.3	Gradient Lattice Structure for Linear Equalization, 162	

6.3.3.4	<i>Lattice Structure for Decision Feedback Equalization, 164</i>	
6.3.3.5	<i>Kalman Method, 170</i>	
6.4	Equalizer Performance Examples	174
6.4.1	Linear Equalizer Performance Examples, 175	
6.4.1.1	<i>Telephone Channel Example, 175</i>	
6.4.1.2	<i>Multipath Radio Channel Example, 184</i>	
6.4.2	Decision Feedback Equalizer Performance Examples, 198	
6.4.2.1	<i>Troposcatter Channel Example, 198</i>	
6.4.2.2	<i>HF Channel Example, 207</i>	
6.4.3	Performance of Lattice Structures, 223	
Appendix	Derivation of Recursive Equations Needed in Gradient Lattice Structure for Linear Equalization	229
7	POWER SPECTRAL ESTIMATION	238
7.1	Historical Overview	238
7.2	Traditional Power Spectral Estimation Technique	239
7.2.1	Power Spectrum Estimation Using Periodogram Methods, 244	
7.3	Autoregressive Method	250
7.3.1	AR Technique, 250	
7.3.2	Durbin Method, 253	
7.3.3	Selection of the Order of the Prediction Filter, 254	
8	DIGITAL WHITENING IN SPREAD SPECTRUM (SS) COMMUNICATIONS	266
8.1	Introduction	266
8.2	Overview of Spread Spectrum Communications	266
8.3	Augmentation of PN Spread Spectrum Communications with Digital Whitening	268
8.4	SS Communication System Model with Digital Whitening	269
8.5	Digital Whitening Algorithms	272
8.6	Decision Variable for Digitally Whitened SS Communication System	275
8.6.1	Moments of the PN Correlator Output without Digital Whitening, 275	