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

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

Needham, Massachusetts
January 1985

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

$\mathscr{E}(x)$	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
$\boldsymbol{X}^t$	transpose of vector,† $\boldsymbol{X}$
$\delta(t)$	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 $\boldsymbol{X}$ has elements $x_1, \cdots x_N$, then $\boldsymbol{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}$.

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
$\boldsymbol{A}_M(N)$	vector of forward prediction error coefficients, that is, $\boldsymbol{A}_M^t(N) = (a_M(1, N), \ldots, a_M(M, N))$
$\bar{\boldsymbol{A}}_M(N)$	extended vector of forward prediction error coefficients, that is, $\bar{\boldsymbol{A}}_M^t(N) = (1, a_M(1, N), \ldots, a_M(M, N))$
$\alpha_M(N)$	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
$\boldsymbol{B}_M(N)$	vector of backward prediction error coefficients, that is, $\boldsymbol{B}_M^t(N) = (b_M(0, N), \ldots, b_M(M-1, N)$
$\bar{\boldsymbol{B}}_M(N)$	extended vector of backward prediction error coefficients, that is, $\bar{\boldsymbol{B}}_M^t(N) = (1, b_M(0, N), \ldots, b_M(M-1, N))$
$\boldsymbol{B}$	vector of Kalman prediction coefficients, that is, $\boldsymbol{B}^t = (b_1, \ldots, b_M)$
$\boldsymbol{B}_k$	vector of Kalman prediction coefficients, that is, $\boldsymbol{B}_k^t = (b_{1,k}, \ldots, b_{M,k})$
$\hat{\boldsymbol{B}}_{k,k-1}$	predicted value of current Kalman coefficient estimate $\hat{\boldsymbol{B}}_k$
β_k	DFE feedback coefficients in Chapter 6
C_k	Fourier-series coefficients
c_k	linear equalizer coefficients in Chapter 6
$\tilde{\boldsymbol{C}}_k$	vector of feedforward and feedback equalizer coefficients, that is, $\tilde{\boldsymbol{C}}_k^t = (\alpha_0, \ldots, \alpha_{M_1}, \beta_1, \ldots, \beta_{M_2})$ (at time instant k)
$\boldsymbol{C}_k$	vector of linear equalizer coefficients (at time instant k), that is, $\boldsymbol{C}_k^t = (c_0, \ldots, c_M)$

$\boldsymbol{C}_M(N)$	vector of feedforward and feedback equalizer coefficients, that is, $\boldsymbol{C}_M^t(N) = (C_M(0, N), \ldots, C_M(M, N))$
$\boldsymbol{C}_{M_T}(N)$	vector of feedforward and feedback lattice equalizer coefficients, that is, $\boldsymbol{C}_{M_T}^t(N) = (C_{M_T}(0, N), \ldots, C_{M_T}(M, N))$
$\boldsymbol{\delta}_k$	zero mean random vector for Kalman algorithm with M elements
$\boldsymbol{D}_M(k, N)$	lattice structure vector with M elements
$\boldsymbol{D}_k$	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
$e_M^f(k, N)$, $e_M^f(N)$	forward error signal
$e_M^b(k, N)$, $e_M^b(N)$	backward error signal
$e_M(k, N)$, $e_M(N)$	error between information symbol and its estimate
$\boldsymbol{E}_M^f(N)$	minimum forward MSE power vector with $M + 1$ elements, that is, $\boldsymbol{E}_M^f(N) = (f_M(N), 0, \ldots, 0)$
$\boldsymbol{E}_M^b(N)$	minimum backward MSE power vector with $M + 1$ elements, that is, $\boldsymbol{E}_M^{bt}(N) = (0, \ldots, 0, r_M(N))$
$\underline{e}_M^f(k, N)$	vector of forward error signals for DFE lattice structure, that is, $\underline{e}_M^{ft}(k, N) = (e_{1,M}^f(k, N)\, e_{2,M}^f(k, N))$
$\underline{e}_M^b(k, N)$	vector of backward error signals for DFE lattice structure, that is, $\underline{e}_M^{bt}(k, N) = (e_{1,M}^b(k, N)\, e_{2,M}^b(k, N))$
$\boldsymbol{E}$	Kalman error vector, that is, $\boldsymbol{E}^t = (\varepsilon_1, \ldots, \varepsilon_N)$
$\boldsymbol{E}_k$	Kalman error vector, that is, $\boldsymbol{E}_k^t = (\varepsilon_{1k}, \ldots, \varepsilon_{Nk})$
E_b/N_0	ratio of signal energy per bit to noise power spectral density or energy contrast ratio

$E(C)$	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
$\boldsymbol{G}_k$	$M \times N$ Kalman gain matrix
$\boldsymbol{G}_M(N)$	$M + 1$ element Kalman gain vector with order index
$\boldsymbol{\Gamma}_k$	gradient vector for steepest descent algorithms with $M_1 + M_2 + 1$ elements
I_k	kth information symbol
$\tilde{I}_k$	kth information symbol decision
$\hat{I}_M(k, N)$, $\hat{I}_k$	estimate of kth information symbol
$k_M(N)$	scalar in lattice structure
$\boldsymbol{K}_k$	Kalman gain vector, that is, $\boldsymbol{K}_k^t = (K_{1,k}, \ldots, K_{N,k})$
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
$\boldsymbol{P}_k$	$M \times M$ error covariance matrix of Kalman prediction coefficients
$\boldsymbol{P}_{k,k-1}$	$M \times M$ Kalman prediction error covariance matrix
$P(f)$	power spectral density of received signal
$q(N)$	zeroth time average autocorrelation element, that is, $q(N) = \rho_N(0,0)$
$\boldsymbol{Q}_M(N)$	vector of time average autocorrelation coefficients, that is, $\boldsymbol{Q}_M^t(N) = (\rho_N(1,0), \ldots, \rho_N(M,0))$
$\boldsymbol{Q}_k$	$M \times M$ covariance matrix of Kalman random vector $\boldsymbol{\delta}_k$
$\rho_N(m, j)$	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
$\boldsymbol{R}_M$	$M \times M$ covariance matrix of received signals (ensemble average)
$\boldsymbol{R}_M(N)$	$M \times M$ time average autocorrelation matrix of received signals
$\boldsymbol{R}_e$	$N \times N$ Kalman covariance matrix of error signals (ensemble average)
$\boldsymbol{R}_{ke}$	$N \times N$ 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
$\boldsymbol{S}$	Kalman vector of transmitted signal samples, that is, $\boldsymbol{S}^t = (s_1, \ldots, s_N)$
$\boldsymbol{S}_k$	Kalman vector of transmitted signal samples, that is, $\boldsymbol{S}_k^t = (s_{1,k}, \ldots, s_{N,k})$

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 $u_{ij}(k)$
$v_M(N)$	Mth time average autocorrelation coefficient, that is, $v_M(N) = \rho_N(M, M)$
$V_M(N)$	vector of time average autocorrelation coefficients, that is, $V_M^t(N) = (\rho_N(0, M), \ldots, \rho_N(M-1, M))$
ω	weighting coefficient in Chapters 4 and 6
$W_M(N)$	time average crosscorrelation vector between information symbols and received sample, that is, $W_M^t(N) = (w_0, \ldots, w_M)$
$x(t)$	received signal
x_k	received or observed signal samples
$X_M(k)$, $X(k)$	received signal vector, that is, $X_M^t(i) = (x_i, \ldots, x_{i-M})$ (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, $X_M^t = (x_1, \ldots, x_M)$
X, X_k	$N \times M$ Kalman matrix of received signal samples
$X_k(i)$	Kalman vector of received signal samples, that is, $X_k^t(i) = (x_{i,k}, \ldots, x_{i-(M-1),k})$
$\overline{X}_k$	square root Kalman vector of received signal samples, that is, $\overline{X}_k^t = (x_{1,k}, \ldots, x_{N,k})$
X_k'	vector of received signal samples, that is, $X_k'^t = (x_k, \ldots, x_{k+M})$

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

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

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