

Kalman Filtering and Neural Networks, Edited by Simon Haykin
Copyright © 2001 John Wiley & Sons, Inc.
ISBNs: 0-471-36998-5 (Hardback); 0-471-22154-6 (Electronic)

KALMAN FILTERING AND NEURAL NETWORKS

KALMAN FILTERING AND NEURAL NETWORKS

Edited by

Simon Haykin

Communications Research Laboratory,
McMaster University, Hamilton, Ontario, Canada



A WILEY-INTERSCIENCE PUBLICATION

JOHN WILEY & SONS, INC.

New York / Chichester / Weinheim / Brisbane / Singapore / Toronto

Designations used by companies to distinguish their products are often claimed as trademarks. In all instances where John Wiley & Sons, Inc., is aware of a claim, the product names appear in initial capital or ALL CAPITAL LETTERS. Readers, however, should contact the appropriate companies for more complete information regarding trademarks and registration.

Copyright 2001 by John Wiley & Sons, Inc.. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic or mechanical, including uploading, downloading, printing, decompiling, recording or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-Mail: PERMREQ@WILEY.COM.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional person should be sought.

ISBN 0-471-22154-6

This title is also available in print as ISBN 0-471-36998-5.

For more information about Wiley products, visit our web site at www.Wiley.com.

CONTENTS

Preface **xi**

Contributors **xiii**

1 Kalman Filters **1**

Simon Haykin

- 1.1 Introduction / 1
- 1.2 Optimum Estimates / 3
- 1.3 Kalman Filter / 5
- 1.4 Divergence Phenomenon: Square-Root Filtering / 10
- 1.5 Rauch–Tung–Striebel Smoother / 11
- 1.6 Extended Kalman Filter / 16
- 1.7 Summary / 20
- References / 20

**2 Parameter-Based Kalman Filter Training:
Theory and Implementation** **23**

Gintaras V. Puskorius and Lee A. Feldkamp

- 2.1 Introduction / 23
- 2.2 Network Architectures / 26
- 2.3 The EKF Procedure / 28
 - 2.3.1 Global EKF Training / 29
 - 2.3.2 Learning Rate and Scaled Cost Function / 31
 - 2.3.3 Parameter Settings / 32
- 2.4 Decoupled EKF (DEKF) / 33
- 2.5 Multistream Training / 35

2.5.1	Some Insight into the Multistream Technique /	40
2.5.2	Advantages and Extensions of Multistream Training /	42
2.6	Computational Considerations /	43
2.6.1	Derivative Calculations /	43
2.6.2	Computationally Efficient Formulations for Multiple-Output Problems /	45
2.6.3	Avoiding Matrix Inversions /	46
2.6.4	Square-Root Filtering /	48
2.7	Other Extensions and Enhancements /	51
2.7.1	EKF Training with Constrained Weights /	51
2.7.2	EKF Training with an Entropic Cost Function /	54
2.7.3	EKF Training with Scalar Errors /	55
2.8	Automotive Applications of EKF Training /	57
2.8.1	Air/Fuel Ratio Control /	58
2.8.2	Idle Speed Control /	59
2.8.3	Sensor-Catalyst Modeling /	60
2.8.4	Engine Misfire Detection /	61
2.8.5	Vehicle Emissions Estimation /	62
2.9	Discussion /	63
2.9.1	Virtues of EKF Training /	63
2.9.2	Limitations of EKF Training /	64
2.9.3	Guidelines for Implementation and Use /	64
	References /	65

3 Learning Shape and Motion from Image Sequences 69

Gaurav S. Patel, Sue Becker, and Ron Racine

3.1	Introduction /	69
3.2	Neurobiological and Perceptual Foundations of our Model /	70
3.3	Network Description /	71
3.4	Experiment 1 /	73
3.5	Experiment 2 /	74
3.6	Experiment 3 /	76
3.7	Discussion /	77
	References /	81

4 Chaotic Dynamics **83***Gaurav S. Patel and Simon Haykin*

- 4.1 Introduction / 83
- 4.2 Chaotic (Dynamic) Invariants / 84
- 4.3 Dynamic Reconstruction / 85
- 4.4 Modeling Numerically Generated Chaotic Time Series / 87
 - 4.4.1 Logistic Map / 87
 - 4.4.2 Ikeda Map / 91
 - 4.4.3 Lorenz Attractor / 99
- 4.5 Nonlinear Dynamic Modeling of Real-World Time Series / 106
 - 4.5.1 Laser Intensity Pulsations / 106
 - 4.5.2 Sea Clutter Data / 113
- 4.6 Discussion / 119
- References / 121

5 Dual Extended Kalman Filter Methods **123***Eric A. Wan and Alex T. Nelson*

- 5.1 Introduction / 123
- 5.2 Dual EKF–Prediction Error / 126
 - 5.2.1 EKF–State Estimation / 127
 - 5.2.2 EKF–Weight Estimation / 128
 - 5.2.3 Dual Estimation / 130
- 5.3 A Probabilistic Perspective / 135
 - 5.3.1 Joint Estimation Methods / 137
 - 5.3.2 Marginal Estimation Methods / 140
 - 5.3.3 Dual EKF Algorithms / 144
 - 5.3.4 Joint EKF / 149
- 5.4 Dual EKF Variance Estimation / 149
- 5.5 Applications / 153
 - 5.5.1 Noisy Time-Series Estimation and Prediction / 153
 - 5.5.2 Economic Forecasting–Index of Industrial Production / 155
 - 5.5.3 Speech Enhancement / 157
- 5.6 Conclusions / 163
- Acknowledgments / 164

Appendix A: Recurrent Derivative of the Kalman Gain /	164
Appendix B: Dual EKF with Colored Measurement Noise /	166
References /	170

6 Learning Nonlinear Dynamical System Using the Expectation-Maximization Algorithm 175

Sam T. Roweis and Zoubin Ghahramani

6.1	Learning Stochastic Nonlinear Dynamics /	175
6.1.1	State Inference and Model Learning /	177
6.1.2	The Kalman Filter /	180
6.1.3	The EM Algorithm /	182
6.2	Combining EKS and EM /	186
6.2.1	Extended Kalman Smoothing (E-step) /	186
6.2.2	Learning Model Parameters (M-step) /	188
6.2.3	Fitting Radial Basis Functions to Gaussian Clouds /	189
6.2.4	Initialization of Models and Choosing Locations for RBF Kernels /	192
6.3	Results /	194
6.3.1	One- and Two-Dimensional Nonlinear State-Space Models /	194
6.3.2	Weather Data /	197
6.4	Extensions /	200
6.4.1	Learning the Means and Widths of the RBFs /	200
6.4.2	On-Line Learning /	201
6.4.3	Nonstationarity /	202
6.4.4	Using Bayesian Methods for Model Selection and Complexity Control /	203
6.5	Discussion /	206
6.5.1	Identifiability and Expressive Power /	206
6.5.2	Embedded Flows /	207
6.5.3	Stability /	210
6.5.4	Takens' Theorem and Hidden States /	211
6.5.5	Should Parameters and Hidden States be Treated Differently? /	213
6.6	Conclusions /	214
	Acknowledgments /	215

Appendix: Expectations Required to Fit the RBFs /	215
References /	216

7 The Unscented Kalman Filter 221

Eric A. Wan and Rudolph van der Merwe

7.1	Introduction /	221
7.2	Optimal Recursive Estimation and the EKF /	224
7.3	The Unscented Kalman Filter /	234
7.3.1	State-Estimation Examples /	237
7.3.2	The Unscented Kalman Smoother /	240
7.4	UKF Parameter Estimation /	243
7.4.1	Parameter-Estimation Examples /	2
7.5	UKF Dual Estimation /	249
7.5.1	Dual Estimation Experiments /	249
7.6	The Unscented Particle Filter /	254
7.6.1	The Particle Filter Algorithm /	259
7.6.2	UPF Experiments /	263
7.7	Conclusions /	269
Appendix A: Accuracy of the Unscented Transformation /		269
Appendix B: Efficient Square-Root UKF Implementations /		273
References /		277

Index 283

PREFACE

This self-contained book, consisting of seven chapters, is devoted to Kalman filter theory applied to the training and use of neural networks, and some applications of learning algorithms derived in this way.

It is organized as follows:

- Chapter 1 presents an introductory treatment of Kalman filters, with emphasis on basic Kalman filter theory, the Rauch–Tung–Striebel smoother, and the extended Kalman filter.
- Chapter 2 presents the theoretical basis of a powerful learning algorithm for the training of feedforward and recurrent multilayered perceptrons, based on the decoupled extended Kalman filter (DEKF); the theory presented here also includes a novel technique called multistreaming.
- Chapters 3 and 4 present applications of the DEKF learning algorithm to the study of image sequences and the dynamic reconstruction of chaotic processes, respectively.
- Chapter 5 studies the dual estimation problem, which refers to the problem of simultaneously estimating the state of a nonlinear dynamical system and the model that gives rise to the underlying dynamics of the system.
- Chapter 6 studies how to learn stochastic nonlinear dynamics. This difficult learning task is solved in an elegant manner by combining two algorithms:
 1. The expectation-maximization (EM) algorithm, which provides an iterative procedure for maximum-likelihood estimation with missing hidden variables.
 2. The extended Kalman smoothing (EKS) algorithm for a refined estimation of the state.

- Chapter 7 studies yet another novel idea—the unscented Kalman filter—the performance of which is superior to that of the extended Kalman filter.

Except for Chapter 1, all the other chapters present illustrative applications of the learning algorithms described here, some of which involve the use of simulated as well as real-life data.

Much of the material presented here has not appeared in book form before. This volume should be of serious interest to researchers in neural networks and nonlinear dynamical systems.

SIMON HAYKIN

*Communications Research Laboratory,
McMaster University, Hamilton, Ontario, Canada*

Contributors

Sue Becker, Department of Psychology, McMaster University, 1280 Main Street West, Hamilton, ON, Canada L8S 4K1

Lee A. Feldkamp, Ford Research Laboratory, Ford Motor Company, 2101 Village Road, Dearborn, MI 48121-2053, U.S.A.

Simon Haykin, Communications Research Laboratory, McMaster University, 1280 Main Street West, Hamilton, ON, Canada L8S 4K1

Zoubin Ghahramani, Gatsby Computational Neuroscience Unit, University College London, Alexandra House, 17 Queen Square, London WC1N 3AR, U.K.

Alex T. Nelson, Department of Electrical and Computer Engineering, Oregon Graduate Institute of Science and Technology, 19600 N.W. von Neumann Drive, Beaverton, OR 97006-1999, U.S.A.

Gaurav S. Patel, 1553 Manton Blvd., Canton, MI 48187, U.S.A.

Gintaras V. Puskorius, Ford Research Laboratory, Ford Motor Company, 2101 Village Road, Dearborn, MI 48121-2053, U.S.A.

Ron Racine, Department of Psychology, McMaster University, 1280 Main Street West, Hamilton, ON, Canada L8S 4K1

Sam T. Roweis, Gatsby Computational Neuroscience Unit, University College London, Alexandra House, 17 Queen Square, London WC1N 3AR, U.K.

Rudolph van der Merwe, Department of Electrical and Computer Engineering, Oregon Graduate Institute of Science and Technology, 19600 N.W. von Neumann Drive, Beaverton, OR 97006-1999, U.S.A.

Eric A. Wan, Department of Electrical and Computer Engineering, Oregon Graduate Institute of Science and Technology, 19600 N.W. von Neumann Drive, Beaverton, OR 97006-1999, U.S.A.

KALMAN FILTERING AND NEURAL NETWORKS

Adaptive and Learning Systems for Signal Processing, Communications, and Control

Editor: Simon Haykin

Beckerman / ADAPTIVE COOPERATIVE SYSTEMS

Chen and Gu / CONTROL-ORIENTED SYSTEM IDENTIFICATION: An \mathcal{H}_∞ Approach

Cherkassky and Mulier / LEARNING FROM DATA: Concepts, Theory, and Methods

Diamantaras and Kung / PRINCIPAL COMPONENT NEURAL NETWORKS: Theory and Applications

Haykin / KALMAN FILTERING AND NEURAL NETWORKS

Haykin / UNSUPERVISED ADAPTIVE FILTERING: Blind Source Separation

Haykin / UNSUPERVISED ADAPTIVE FILTERING: Blind Deconvolution

Haykin and Puthussarypady / CHAOTIC DYNAMICS OF SEA CLUTTER

Hrycej / NEUROCONTROL: Towards an Industrial Control Methodology

Hyvärinen, Karhunen, and Oja / INDEPENDENT COMPONENT ANALYSIS

Kristić, Kanellakopoulos, and Kokotović / NONLINEAR AND ADAPTIVE CONTROL DESIGN

Nikias and Shao / SIGNAL PROCESSING WITH ALPHA-STABLE DISTRIBUTIONS AND APPLICATIONS

Passino and Burgess / STABILITY ANALYSIS OF DISCRETE EVENT SYSTEMS

Sánchez-Peña and Sznaler / ROBUST SYSTEMS THEORY AND APPLICATIONS

Sandberg, Lo, Fancourt, Principe, Katagiri, and Haykin / NONLINEAR DYNAMICAL SYSTEMS: Feedforward Neural Network Perspectives

Tao and Kokotović / ADAPTIVE CONTROL OF SYSTEMS WITH ACTUATOR AND SENSOR NONLINEARITIES

Tsoukalas and Uhrig / FUZZY AND NEURAL APPROACHES IN ENGINEERING

Van Hulle / FAITHFUL REPRESENTATIONS AND TOPOGRAPHIC MAPS: From Distortion- to Information-Based Self-Organization

Vapnik / STATISTICAL LEARNING THEORY

Werbos / THE ROOTS OF BACKPROPAGATION: From Ordered Derivatives to Neural Networks and Political Forecasting

INDEX

- A priori covariance matrix, 7
- Air/fuel ratio control, 58
- Approximate error covariance matrix, 24, 29–34, 49, 63
- Artificial process-noise, 48–50
- Attentional filtering, 80
- Automatic relevance determination (ARD), 205
- Automotive applications, 57
- Automotive powertrain control systems, 57
- Avoiding matrix inversions, 46

- Backpropagation, 30, 39, 44, 51, 56
- Backpropagation process, 55
- Backward filtering, 12
- Bayesian methods, 203
- Bayes' rule, 181
- BPTT(h), 45

- Cayley–Hamilton theorem, 212
- Central difference interpolation, 230
- Chaotic (dynamic) invariants, 84
- Chaotic dynamics, 83
- Cholesky factorization, 11
- Closed-loop controller, 60
- Closed-loop evaluation, 88, 93, 100, 108, 115
- Colored, 166

- Comparison of chaotic invariances of Ikeda map, 97
- Comparison of chaotic invariants of logistic map, 90
- Comparison of chaotic invariances of Lorenz series, 102
- Comparison of chaotic invariants of sea clutter, 114
- Computational complexity, 24, 33, 34, 39, 46, 63
- Conditional mean estimator, 4
- Constrained weights, 51
- Correlation dimension, 84
- Cortical feedback, 80
- Cost functions, 64
- Covariance matrix of the process noise, 31, 32
- Cross-entropy, 54

- Decoupled extended Kalman filter (DEKF), 26, 33, 39, 47
- Decoupled extended Kalman filter (NDEKF) algorithm, 69
- DEKF algorithm, 34
- Delay coordinate method, 86
- Derivative calculations, 43, 56
- Derivative matrices, 31, 34, 38
- Derivative matrix, 30, 31, 33
- Derivatives of network outputs, 44
- Divergence phenomenon, 10

- Double inverted pendulum, 234
- Dual EKF, 213
- Dual estimation, 123, 130, 224, 249
- Dual Kalman, 125
- Dynamic pattern classifiers, 62
- Dynamic reconstruction, 85
- Dynamic reconstruction of the laser series, 109
- Dynamic reconstruction of the Lorenz series, 101
- Dynamic reconstruction of the noisy Lorenz series, 105
- Dynamic reconstruction of the noisy Ikeda map, 98
- EKF, 37, 43, 52, 54, 56, 62
 - see* Extended Kalman filter
- EKF procedure, 28
- Elliott sigmoid, 53
- EM algorithm, 142
- Embedding delay, 86
- Embedding dimension, 86
- Embedding, 211
- Engine misfire detection, 61
- Entropic cost function, 54, 55
- Error covariance propagation, 8
- Error covariance matrices, 48
- Error covariance matrix, 26
- Error covariance update, 49
- Error vector, 29–31, 34, 38, 52
- Estimation, 124
- Expectation–maximization (EM)
 - algorithm, 177, 182
- Extended Kalman filter (EKF), 16, 24, 123, 182, 221, 227
- Extended Kalman filtering (EKF)
 - algorithm, 179
- Extended Kalman filter-recurrent multilayered perceptron, 83
- Extended Kalman filter, summary of, 19
- Factor analysis (FA), 193
- Filtering, 3
- Forward filtering, 12
- Fully decoupled EKE, 25, 34
- Gauss–Hermite quadrature rule, 230
- GEKF, 30, 33, 34, 39, 62
 - see* Global EKE
- GEKF, decoupled EKE algorithm, 25
- Generative model, 178
- Givens rotations, 49, 50
- Global EKE (GEKF), 24, 26
- Global scaling matrix, 29, 31, 38
- Global scaling matrix \mathbf{A}_k , 34
- Global EKF training, 29
- Graphical models, 178, 179
- Hidden variables, 177
- Hierarchical architecture, 71
- Identifiability, 206
- Idle speed control, 59
- Ikeda map, 91
- Inference, 176
- Innovations, 7
- Jensen’s inequality, 183
- Joint EKF, 213
- Joint estimation, 137
- Joint extended Kalman filter, 125
- “Joseph” version of the covariance update equation, 8
- Kalman filter, 1, 5, 177
- Kalman filter, information formulation of, 13
- Kalman gain, 6
- Kalman gain matrix, 29, 30, 31, 33, 49
- Kalman gain matrices, 34, 38
- Kaplan–York dimension, 85
- Kernel, 192
- Kolmogorov entropy, 85
- Laser intensity pulsations, 106
- Layer-decoupled EKF, 34
- Learning rate, 31, 32, 48

- Least-squares solution, 42
- Logistic map, 87
- Lorenz attractor, 99
- Lyapunov exponents, 84
- Lyapunov dimension, 85

- MAP estimation, 135
- Mammalian neocortex, 70
- Marginal estimation, 140
- Markov-chain Monte Carlo, 258
- Matrices \mathbf{H}_k^i , 52
- Matrix of derivatives, 29
- Matrix factorization lemma, 48
- Matrix inversion lemma, 14
- Matrix inversions, 55, 63
- Maximum a posteriori (MAP), 135
- Maximum-likelihood cost, 140
- Measurement covariance, 30
- Measurement equation, 3
- Measurement error covariance matrix, 38
- Mixture of factor analyzers (MFA), 193
- MMSE estimator, 225
- Modeling, 124
- Model selection, 203
- Monte Carlo simulation, 255
- Moving object, 80
- Multistream, 39, 40
- Multistream EKF, 26, 38
- Multistream EKF training, 63
- Multistream Kalman recursion, 42
- Multistream training, 34, 36, 45

- Neurobiological foundations, 70
- Node-decoupled extended Kalman filter (NDEKF) algorithm, 69
- Node-decoupled EKF, 25, 34, 46
- Noise, 166
- Noisy time-series estimation, 153, 235
- Noisy Ikeda series, 95
- Noisy Lorenz series, 103
- Nonlinear dynamics, 175
- Nonlinear dynamic modelling of real-world time series, 106
- Non-rigid motion, 80
- Nonstationarity, 202

- Occlusions, 75, 78
- On-line learning, 201,
- Open-loop evaluation, 87, 92, 99, 108, 115
- Optimal recursive estimation, 224
- Optimization with constraints, 51
- Optimum smoothing problem, 11
- Overfitting, 204

- Parameter estimation, 223, 240
- Partial M-step, 200
- Particle filters, 182, 259
- Perceptual foundations, 70
- Prediction error, 126
- Prediction, 3, 124
- Priming length, 37
- Principles of orthogonality, 4
- Process equation, 2
- Process noise, 29
- Proposal distribution, 254

- Radial basis function (RBF) networks, 188
- Rauch–Tung–Streibel (RTS) smoother, 11, 17, 180
- Rauch–Tung–Striebel, 15, 25
- Real-time-recurrent learning (RTRL), 25
- Recency effect, 36
- Recency phenomenon, 25, 26
- Reconstruction failures, 119
- Recurrent derivative, 131, 164
- Recurrent multilayered perceptron (RMLP), 26, 28, 30, 61, 69
- Recurrent network, 28, 44, 59
- Recurrent multiplayer perceptron, 69
- Recurrent neural networks, 60, 62, 63
- Recursive least-squares (RLS) algorithm, 201

- Rescaled extended Kalman recursion, 31
- Sea clutter data, 113
- Sensor-catalyst modeling, 60
- Sequential DEKF, 47
- Sequential importance sampling, 255
- Sequential update, 47
- Shape and motion perception, 80
- Signal-to-ratio (SER), 87
- Simultaneous DEKF, 47
- Singular-value decomposition, 46
- Smoothing, 3
- Speech enhancement, 157
- Square-root filtering, 10, 48–50
- Square-root UKF, 273
- Stability, 210
- Stability and robustness, 118
- State-error vector, 5
- State estimation, 127, 176, 222
- Sum of squared error, 29, 64
- Summary of the Rauch–Tung–Striebel smoother, 17
- Summary of the Kalman filter, 10
- Takens embedding, 86
- Taken’s theorem, 211
- Teacher forcing, 42
- Tracking, 70
- Training cost function, 31
- Trajectory length, 37
- Truncated backpropagation through time (BPTT(h)), 25, 37, 44
- Unscented transformation, 228
- Unscented filter, 182
- Unscented particle filter, 254
- Unscented Kalman filter (UKF), 221, 228, 230
- Unscented Kalman smoother, 237
- Variance estimation, 149
- Variational approximations, 210
- Vehicle emissions estimation, 62
- Weather Data, 197
- Weight estimation, 128
- Weighting matrix, 31, 32
- “What” pathway, 79
- “Where” pathway, 80