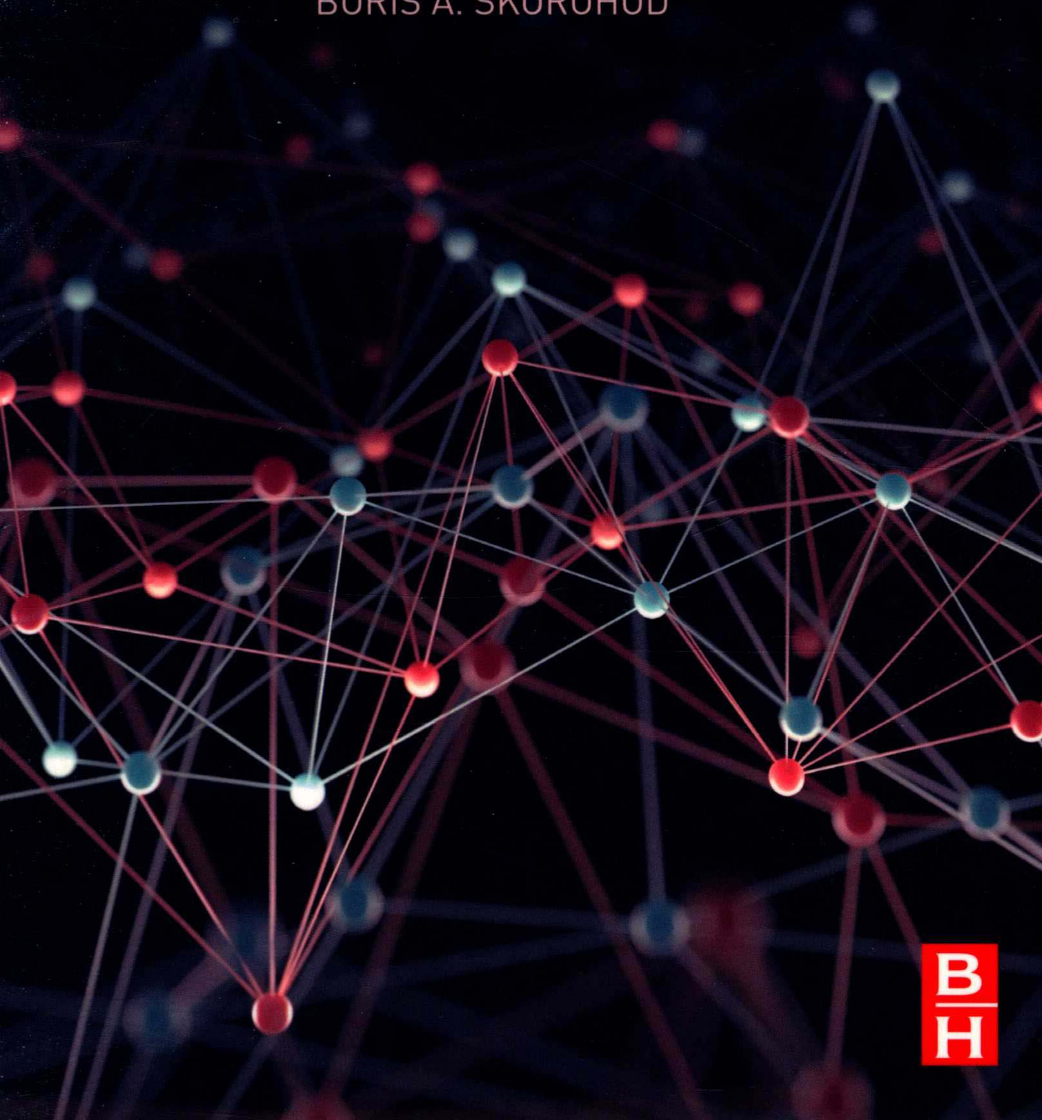
DIFFUSE ALGORITHMS FOR NEURAL AND NEURO-FUZZY NETWORKS

With Applications in Control Engineering and Signal Processing

BORIS A. SKOROHOD



DIFFUSE ALGORITHMS FOR NEURAL AND NEURO-FUZZY NETWORKS

With Applications in Control Engineering and Signal Processing

BORIS A. SKOROHOD

ELSEVIER

Butterworth-Heinemann
An imprint of Elsevier
elsevier.com

Butterworth-Heinemann is an imprint of Elsevier
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN: 978-0-12-812609-7

For Information on all Butterworth-Heinemann publications visit our website at https://www.elsevier.com/books-and-journals

Publisher: Joe Hayton
Acquisition Editor: Sonnini R. Yura
Editorial Project Manager: Ashlie Jackman
Production Project Manager: Kiruthika Govindaraju
Cover Designer: Limbert Matthew

Typeset by MPS Limited, Chennai, India

DIFFUSE ALGORITHMS FOR NEURAL AND NEURO-FUZZY NETWORKS

PREFACE

The problem of neural and neuro-fuzzy networks training is considered in this book. The author's attention is concentrated on the approaches which are based on the use of a separable structure of plants models—nonlinear with respect to some unknown parameters and linear relating to the others. It may be, for example, a multilayered perceptron with a linear activation function at its output, a radial base neural network, a neuro-fuzzy Sugeno network, or a recurrent neural network, which are widely used in a variety of applications relating to the identification and control of nonlinear plants, time series forecasting, classification, and recognition.

Static neural and neuro-fuzzy networks training can be regarded as a problem of minimizing the quality criterion in respect to unknown parameters included in the description of them for a given training set. It is well-known that it is a complex, multiextreme, often ill-conditioned nonlinear optimization problem. In order to solve its various algorithms, that are superior to the error backpropagation algorithm and its numerous modifications in convergence rate, approximation accuracy and generalization ability have been developed. There are also algorithms that directly take into account separable character of networks structure. Thus, in Ref. [1] the VP (variable projection) algorithm for static separable plants models is proposed. According to this algorithm, the initial optimization problem is transformed into a new problem, but only with relation to nonlinear input parameters. Under certain conditions the stationary points sets of two problems coincide, but at the same time dimensionality decreases, and, as a consequence, there is no need for selecting initial values to linearly incoming parameters. Moreover, the new optimization problem is better conditioned [2–4], and if the same method is used for initial and transformed optimization problems, the VP algorithm always converges after a smaller number of iterations. At the same time, the VP algorithm can be implemented only in a batch mode and, in addition, the procedure of determining partial derivative of modified criteria in respect to the parameters becomes considerably more complicated. The hybrid procedure for a Sugeno fuzzy network training is proposed in Refs. [5,6], that is based on the successive use of the recursive least-square method (RLSM) for determining linearly entering parameters and gradient method for nonlinear ones. The extreme learning machine (ELM) approach is developed in Refs. [7,8]. On the basis of this approach only linearly incoming parameters are trained and nonlinear ones are drawn at random without taking into account a training set. However, it is well-known that this approach can provide quite low accuracy at a relatively small size of the training set. It also should be noted that while the ELM and the hybrid algorithms initialization use the RLS algorithm, it is necessary to select the initial values for the matrix which satisfies the Riccati equation. Moreover, as a priori information about the estimated parameters is absent, its elements are generally put proportional to a large parameter, which may lead to divergence in the case of even linear regression.

The purpose of this book is to present new approaches to training of neural and neuro-fuzzy networks which have a separable structure. It is assumed that in addition to the training set a priori information only about the nonlinearly incoming parameters is given. This information may be obtained from the distribution of a generating sample, a training set, or some linguistic information. For static separable models the problem of minimizing a quadratic criterion that includes only that information is considered. Such a problem statement and the Gauss–Newton method (GNM) with linearization around the latest estimate lead to new online and offline training algorithms that are robust in relation to unknown a priori information about linearly incoming parameters. To be more precise, they are interpreted as random variables with zero expectation and a covariance matrix proportional to an arbitrarily large parameter μ (soft constrained initialization). Asymptotic representations as $\mu \to \infty$ for the GNM, which we call diffuse training algorithms (DTAs), are found. We explore the DTA properties. Particularly the DTAs' convergence in case of the limited and unlimited sample size is studied. The problem specialty is connected with the observation model separable character, and the fact that the nonlinearly inputting parameters belong to some compact set, and linearly inputting parameters should be considered as arbitrary numbers.

It is shown that the proposed DTAs have the following important characteristics:

1. Unlike their prototype, the GNM with a large but finite μ, the DTAs are robust with respect to round-off error accumulation.
2. As in Refs. [1–4] initial values choice for linearly imputing parameters is not required, but at the same time there is no need to evaluate the projection matrix partial derivative.

3. Online and offline regimes can be used.
4. The DTAs are followed with the ELM approach and the hybrid algorithm of the Sugeno neuro-fuzzy network training [6,7], and presented modeling results show that developed algorithms can surpass them in accuracy and convergence rate.

With a successful choice of a priori information for the nonlinear parameters, rapid convergence to one of the acceptable minimum criteria points can be expected. In this regard, the DTAs' behavior analysis at fixed values of the nonlinear parameters, when a separable model is degenerating into a linear regression, is very important. We attribute this to the possible influence of the properties of linear estimation problem on the DTAs. The behavior of the RLSM with soft and diffuse initialization in a finite time interval, including a transition stage, is considered. In particular, the asymptotic expansion for the solution of the Riccati equation, the gain rate in inverse powers of μ, and conditions for the absence of overshoot in the transition phase are obtained. Recursive estimation algorithms (diffuse) as $\mu \to \infty$ not depending on a large parameter μ which leads to the divergence of the RLSM are proposed.

The above-described approach is generalized in the training problem of separable dynamic plant models—a state vector and numerical parameters are simultaneously evaluated using the relations for the extended diffuse Kalman filter (DKF) obtained in this book. It is assumed that in addition to the training set a priori information only on nonlinearly inputting parameters and an initial state vector, which can be obtained from the distribution of a generating sample, is used. Linearly inputting parameters are interpreted as random variables with a zero expectation and a covariance matrix proportional to arbitrarily large parameter μ. Asymptotic relations as $\mu \to \infty$, which describe the extended KF (EKF), are called the diffuse extended KF.

The theoretical results are illustrated with numerical examples of identification, control, signal processing, and pattern recognition problem-solving. It is shown that the DTAs may surpass the ELM and the hybrid algorithms in approximation accuracy and necessary iterations number. In addition, the use of the developed algorithms in a variety of engineering applications, which the author has been interested in at different times, is also described. These are dynamic mobile robot model identification, neural networks-based modeling of mechanical hysteresis deformations, and monitoring of the electric current harmonic components.

The book includes six chapters. The first chapter presents an overview of the known models of objects and results relating to the subject of the book.

The RLSM behavior on a finite interval is considered in Chapter 2, Diffuse Algorithms for Estimating Parameters of Linear Regression. It is assumed that the initial value of the matrix Riccati equation is proportional to a large positive parameter μ. Asymptotic expansions of the Riccati equation solution and the RLSM gain rate in inverse powers of μ are obtained. The limit recursive algorithms (diffuse) as $\mu \to \infty$ not depending on a large parameter μ which leads to the RLSM divergence are proposed and explored. The theoretical results are illustrated by examples of solving problems of identification, control, and signal processing.

In Chapter 3, Statistical Analysis of Fluctuations of Least Squares Algorithm on Finite Time Interval, properties of the bias, the matrix of second-order moments, and the normalized average squared error of the RLSM on a finite time interval are studied. It is assumed that the initial condition of the Riccati equation is proportional to the positive parameter μ and the time interval includes an initialization stage. Based on the Chapter 2, Diffuse Algorithms for Estimating Parameters of Linear Regression results, asymptotic expressions for these quantities in inverse powers of μ for the soft initialization and limit expression for the diffuse initialization are obtained. It is shown that the normalized average squared error of estimation can take arbitrarily large but bounded values as $\mu \to \infty$. The conditions are expressed in terms of signal/noise ratio under which overshoot does not exceed the initial value (conditions for the absence of overshoot).

Chapter 4, Diffuse Neural and Neuro-Fuzzy networks Training Algorithms deals with the problem of multilayer neural and neuro-fuzzy networks training with simultaneous estimation of the hidden and output layer parameters. The hidden layer parameters probable values and their possible deviations are assumed to be known. A priori information about the output layer weights is absent and in one initialization of the GNM they are assumed to be random variables with zero expectations and a covariance matrix proportional to the large parameter and in the other option either unknown constants or random variables with unknown statistical characteristics. Training algorithms based on the GNM with linearization about the latest estimate are proposed and studied. The theoretical results are illustrated with the examples of pattern recognition, and identification of nonlinear static and dynamic plants.

The estimation problem of the state and the parameters of the discrete dynamic plants in the absence of a priori statistical information about initial conditions or its incompletion is considered in Chapter 5, Diffuse Kalman Filter. Diffuse analogues of the Kalman filter and the extended Kalman filter are obtained. As a practical application, the problems of the filter constructing with a sliding window, observers restoring state in a finite time, recurrent neural networks training, and state estimation of nonlinear systems with partly unknown dynamics are considered.

Chapter 6, Applications of Diffuse Algorithms provides examples of the use of diffuse algorithms for solving problems with real data arising in various engineering applications. They are the mobile robot dynamic model identification, hysteresis mechanical deformations modeling on the basis of neural networks, and electric current harmonic components monitoring.

The author expresses deep gratitude to Head of Department Y.B. Ratner from Marine Hydrophysical Institute of RAS, Department of Marine Forecasts and Professor S.A. Dubovik from Sevastopol State University, Department of Informatics and Control in Technical Systems for valuable discussions, and to his wife Irina for help in preparation of the manuscript.

Boris Skorohod,
Sevastopol, Russia
January 2017

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

Contents

1.1 SEPARABLE MODELS OF PLANTS AND TRAINING PROBLEMS ASSOCIATED WITH THEM

1.1.1 Separable Least Squares Method

Let us consider an observation model of the form

$$y_t = \Phi(z_t, \beta)\alpha, \quad t = 1, 2, \ldots, N, \tag{1.1}$$

where $z_t = (z_{1t}, z_{2t}, \ldots, z_{nt})^T \in R^n$ is a vector of inputs, $y_t = (y_{1t}, y_{2t}, \ldots, y_{mt})^T \in R^m$ is a vector of outputs, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_r)^T \in R^r$, $\beta = (\beta_1, \beta_2, \ldots, \beta_l)^T \in R^l$ are vectors of unknown parameters, $\Phi(z_t, \beta)$ is an $m \times r$ matrix of given nonlinear functions, R^l is the space of vectors of length l, $(\cdot)^T$ is the matrix transpose operation, and N is a sample size.

The vector output y_t depends linearly on α and nonlinearly on β. This model is called the separable regression (SR) [1]. If the vector β is known, then Eq. (1.1) is transformed into a linear regression.

Let there be given the set of input–output pairs $\{z_t, y_t\}$, $t = 1, 2, \ldots, N$ and the quality criteria

Diffuse Algorithms for Neural and Neuro-Fuzzy Networks.
DOI: http://dx.doi.org/10.1016/B978-0-12-812609-7.00001-9