**FEED-FORWARD
NEURAL NETWORKS**
*Vector Decomposition Analysis,
Modelling and Analog Implementation*

**THE KLUWER INTERNATIONAL SERIES
IN ENGINEERING AND COMPUTER SCIENCE**

ANALOG CIRCUITS AND SIGNAL PROCESSING

Consulting Editor
Mohammed Ismail
Ohio State University

Related Titles:

FREQUENCY COMPENSATION TECHNIQUES LOW-POWER OPERATIONAL AMPLIFIERS, Ruud Easchauzier, Johan Huijsing

ISBN: 0-7923-9565-4

ANALOG SIGNAL GENERATION FOR BIST OF MIXED-SIGNAL INTEGRATED CIRCUITS, Gordon W. Roberts, Albert K. Lu

ISBN: 0-7923-9564-6

INTEGRATED FIBER-OPTIC RECEIVERS, Aaron Buchwald, Kenneth W. Martin

ISBN: 0-7923-9549-2

MODELING WITH AN ANALOG HARDWARE DESCRIPTION LANGUAGE,

H. Alan Mantooth, Mike Fiigenbaum

ISBN: 0-7923-9516-6

LOW-VOLTAGE CMOS OPERATIONAL AMPLIFIERS: Theory, Design and Implementation,

Satoshi Sakurai, Mohammed Ismail

ISBN: 0-7923-9390-7

ANALYSIS AND SYNTHESIS OF MOS TRANSLINEAR CIRCUITS, Remco J. Wiegink

ISBN: 0-7923-9390-2

COMPUTER-AIDED DESIGN OF ANALOG CIRCUITS AND SYSTEMS, L. Richard Carley,

Ronald S. Gyuresik

ISBN: 0-7923-9351-1

HIGH-PERFORMANCE CMOS CONTINUOUS-TIME FILTERS, José Silva-Martínez, Michiel

Steyaert, Willy Sansen

ISBN: 0-7923-9339-2

SYMBOLIC ANALYSIS OF ANALOG CIRCUITS: Techniques and Applications, Lawrence

P. Huelsman, Georges G. E. Gielen

ISBN: 0-7923-9324-4

DESIGN OF LOW-VOLTAGE BIPOLAR OPERATIONAL AMPLIFIERS, M. Jeroen Fonderie,

Johan H. Huijsing

ISBN: 0-7923-9317-1

STATISTICAL MODELING FOR COMPUTER-AIDED DESIGN OF MOS VLSI CIRCUITS,

Christopher Michael, Mohammed Ismail

ISBN: 0-7923-9299-X

SELECTIVE LINEAR-PHASE SWITCHED-CAPACITOR AND DIGITAL FILTERS, Hussein

Baher

ISBN: 0-7923-9298-1

ANALOG CMOS FILTERS FOR VERY HIGH FREQUENCIES, Bram Nauta

ISBN: 0-7923-9272-8

ANALOG VLSI NEURAL NETWORKS, Yoshiyasu Takefuji

ISBN: 0-7923-9273-6

ANALOG VLSI IMPLEMENTATION OF NEURAL NETWORKS, Carver A. Mead,

Mohammed Ismail

ISBN: 0-7923-9049-7

AN INTRODUCTION TO ANALOG VLSI DESIGN AUTOMATION, Mohammed Ismail, José

Franca

ISBN: 0-7923-9071-7

INTRODUCTION TO THE DESIGN OF TRANSDUCTOR-CAPACITOR FILTERS,

Jaime Kardontchik

ISBN: 0-7923-9195-0

VLSI DESIGN OF NEURAL NETWORKS, Ulrich Ramacher, Ulrich Ruckert

ISBN: 0-7923-9127-6

LOW-NOISE WIDE-BAND AMPLIFIERS IN BIPOLAR AND CMOS TECHNOLOGIES, Z.

Y. Chang, Willy Sansen

ISBN: 0-7923-9096-2

ANALOG INTEGRATED CIRCUITS FOR COMMUNICATIONS: Principles, Simulation and

Design, Donald O. Pederson, Kartikeya Mayaram

ISBN: 0-7923-9089-X

**FEED-FORWARD
NEURAL NETWORKS**
*Vector Decomposition Analysis,
Modelling and Analog Implementation*

by

Anne-Johan Annema
*MESA Research Institute
University of Twente*

SPRINGER-SCIENCE+BUSINESS MEDIA, LLC

ISBN 978-1-4613-5990-6 ISBN 978-1-4615-2337-6 (eBook)
DOI 10.1007/978-1-4615-2337-6

Library of Congress Cataloging-in-Publication Data

A C.I.P. Catalogue record for this book is available
from the Library of Congress.

Copyright © 1995 by Springer Science+Business Media New York
Originally published by Kluwer Academic Publishers, New York in 1995
Softcover reprint of the hardcover 1st edition 1995

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, mechanical, photo-copying, recording, or otherwise, without the prior written permission of the publisher, Springer-Science+Business Media, LLC

Printed on acid-free paper.

Contents

Foreword	ix
Acknowledgments	xi
1 Introduction	1
1.1 Neural networks	1
1.2 Feed-Forward Networks	6
Architecture of feed-forward neural networks	6
Applications for feed-forward neural networks	9
Capabilities of feed-forward neural networks: some theorems	10
1.3 Back-Propagation	16
1.4 Realizations of feed-forward networks	17
1.5 Outline of the book	20
1.6 References	22
2 The Vector Decomposition Method	27
2.1 Introduction	27
2.2 The basics of the VDM	29
2.3 Some notations and definitions	30
2.4 The VDM in more detail	33
Decomposition basics	33
The actual vector decomposition	34
Quantification of vector components	35
An illustration	36
The neuron response	36
2.5 A summary of the VDM	37
2.6 References	37

3	Dynamics of Single Layer Nets	39
3.1	Introduction	39
3.2	Weight vector adaptation with the VDM	42
	Weight adaptation of one neuron with the VDM	42
	Average adaptation of β^h and β^{bias}	43
	Adaptation of $\underline{W}^e$	43
3.3	The effect of the learning rate on learning	46
3.4	The effect of scaling η and $\underline{U}$	47
3.5	The effect of bias-input signal on learning: simple case	48
3.6	The effect of bias-input signal on learning: general case	51
3.7	Conclusions	55
3.8	References	56
4	Unipolar Input Signals in Single-Layer Feed-Forward Neural Networks	57
4.1	Introduction	57
4.2	Translations towards unipolar input signals	58
	Centre-of-gravity	59
	Minimum training time for fixed learning rate η	59
	Minimum training time, including scaling of η	60
	Discussion	61
4.3	References	61
5	Cross-talk in Single-Layer Feed-Forward Neural Networks	63
5.1	Introduction	63
5.2	Coupling between input signals	64
	Analysis of the effect of coupling	64
5.3	Degradation of learning due to coupling	68
5.4	Types of coupling	69
	Capacitive coupling	69
	Resistive coupling	69
	Additive coupling	69
5.5	Calculation & simulation results	70
5.6	Discussion	73
5.7	References	74

6	Precision Requirements for Analog Weight Adaptation Circuitry for Single-Layer Nets	75
6.1	Introduction	75
6.2	The cause and the model of analog imprecision	76
6.3	Estimation of MSE-increment due to imprecision	77
	Basic analysis	77
	The effect on the <i>MSE</i>	78
	An illustration	79
6.4	The effect on correctly classified examples	80
6.5	Rule of thumb	82
	The condition for negligibly small effect of parasitic weight adaptation	83
6.6	Worst-case estimation of precision requirements	85
6.7	Estimation of minimum weight-storage <i>C</i> size	86
6.8	Conclusions	87
6.9	References	87
	Appendix 6.1: Derivation of equation (6.3)	88
	Appendix 6.2: Approximation of error distribution	89
7	Discretization of Weight Adaptations in Single-Layer Nets	91
7.1	Introduction	91
7.2	Basics of discretized weight adaptations	92
7.3	Performance versus quantization: asymptotical	93
	A simple case	93
	A less simple case	95
	A general case	97
7.4	Worst-case estimation of quantization steps	101
	A simple case	101
	A less simple case	103
	A general case	104
7.5	Estimation of absolute minimum weight-storage <i>C</i> size	105
7.6	Conclusions	106
7.7	References	106

8	Learning Behavior and Temporary Minima of Two-Layer Neural Networks	107
8.1	Introduction	107
8.2	A summary	110
	The network and the notation	110
	Back-propagation rule	111
	Vector decomposition	112
	Preview of the analyses	113
8.3	Analysis of temporary minima: introduction	115
	Initial training: a linearized network	116
	Continued training: including network non-linearities	120
8.4	Rotation-based breaking	121
	Discussion	123
8.5	Rotation-based based breaking: an illustrative example	127
8.6	Translation-based breaking	135
8.7	Translation-based breaking: an illustrative example	138
8.8	Extension towards larger networks	141
8.9	Conclusions	144
8.10	References	144
9	Biases and Unipolar Input signals for Two-Layer Neural Networks	147
9.1	Introduction	147
9.2	Effect of the first layer's bias-input signal on learning	148
	Learning behavior: a recapitulation	149
	First layer's bias input versus adaptation in the $\underline{B}^h$ direction	151
	Relation between first layer's bias input and temporary minima	152
	Overall conclusions	154
	An illustration	155
9.3	Effect of the second layer's bias signal on learning	156
	Second layer's bias input versus adaptation in the $\underline{B}^h$ direction	156
	Relation between second layer's bias input and temporary minima	157
	Conclusions	159
	An illustration	160
9.4	Large neural network: a problem and a solution	161
9.5	Unipolar input signals	165
9.6	References	166

10	Cost Functions for Two-Layer Neural Networks	167
10.1	Introduction	167
10.2	Discussion of “Minkowski-r back-propagation”	168
	Making an “initial guess”	168
	Analysis of the training time required to reach minima	169
	Analysis of ‘sticking’ time in temporary minima	170
	An illustration	172
10.3	Switching cost functions	172
10.4	Classification performances using non-MSE cost-function	175
10.5	Conclusions	175
10.6	References	176
11	Some issues for $f'(x)$	177
11.1	Introduction	177
11.2	Demands on the activation function for single-layer nets	178
11.3	Demands on the activation functions for two-layer nets	180
12	Feed-forward hardware	187
12.1	Introduction	187
12.2	Normalization of signals in the network	188
12.3	Feed-forward hardware: the synapses	193
	Requirements	193
	The synapse circuit	196
12.4	Feed-forward hardware: the activation function	199
12.5	Conclusions	203
12.6	References	203
	Appendix 12.1: Neural multipliers: overview	204
	Appendix 12.2: Neural activation functions: overview	210

13	Analog weight adaptation hardware	215
13.1	Introduction	215
13.2	Multiplier: the basic idea	215
13.3	Towards a solution	218
13.4	The weight-update multiplier	221
13.5	Simulation results	222
13.6	Reduction of charge injection	223
13.7	Conclusions	228
13.8	References	228
14	Conclusions	229
14.1	Introduction	229
14.2	Summary	230
14.3	Original contributions	231
14.4	Recommendations for further research	231
	Index	235
	Nomenclature	237

Foreword

Artificial neural networks experienced the attention of many researchers in neuroscience and computer science during the last decade. This is usually considered as the revival of neural network research after two decades in which the interest in perceptron-based architectures was lost. The little effort in neural networks in the period 1970 to 1985 can partly be explained by the lack of sufficient data processing power in those days, which prevented researchers to demonstrate the computational capabilities of neural nets. Another reason may be found in the booming progress in VLSI design and realisation, which did attract a lot of attention of the research community.

Meanwhile renewed interest for biologically inspired neural computation focused the attention to the powerful capabilities of this approach for parallel processing. After 1985, the success and progress in VLSI realisation provided the prerequisites for a return to research in neural network realisation. Nowadays the advanced VLSI technology allows the realisation of very cost-effective processors and huge memories which can be used to simulate or emulate parallel neural processing.

Many successful neural network applications have been reported. Several neural network architectures have been investigated by comparing and evaluating simulations on adaptivity and performance. A lot of expert knowledge is gained by experience and analyses of learning systems. This is attended with an exploding number of papers, conference contributions and books on neural networks.

In such circumstances one should have a very good reason when publishing yet another book about neural networks. Fortunately, Anne Johan Annema has such reason. Thanks to his different point of view that originated from the wish to obtain specifications for analog hardware modules, he developed the Vector Decomposition Analysis for feed-forward neural networks with back-propagation learning. In this books he explains the analysis method and illustrates its power.

The Vector Decomposition Method appears applicable for the analysis of feed-forward neural nets, whether they are hardware analog or digital or even software implemented. The key to success is the particular choice for the basis for the Vector Decomposition Analysis, which makes the analytical expressions easy to read and easy to handle. Looking back to the very challenging period during which Anne Johan Annema was with the MESA Research Institute at the University of Twente, I realise that his particular choice for the basis of the Vector Decomposition Analysis offered the key to a great deal of demystification of the neural network learning behaviour. This work, which is a comprehensive cultivation of the PhD thesis has evolved into an interesting and tractable book.

Prof.dr. Hans Wallinga

Acknowledgments

This book is a slightly modified version of my Ph.D. thesis. It describes the results of a research project that was carried out at the MESA Research Institute at the University of Twente, Enschede, The Netherlands. The work has been supported by the Foundation for Fundamental Research on Matter (FOM) and the Netherlands Technology Foundation (STW). A number of people contributed to a nice working atmosphere at MESA or/and to the research project. At the risk of forgetting some persons, I'd like to thank:

- Prof. Hans Wallinga, Klaas Hoen, and Remco Wiegerink for many fruitful discussions and for their comments on the manuscript. I very much liked working together with them.
- My (ex-)colleagues Remco Wiegerink, Eric Klumperink, Peter Masa, Ton Ikkink, Clemens Mensink, Henk de Vries, Roel Wassenaar, Cor Bakker, Jan Hovius, Karel Lippe, Han Speek, Ning, and Jan Niehof for a lot of discussion about all kinds of everything.
- Albert Bos of the Applied Chemistry department of the University of Twente, Stan Gielen of the University of Nijmegen and Peter Johannesma for discussions in the field of neural networks.
- Last but not least, I would like to thank my father, who (among others) made it possible to study at a university, and who always is willing to accompany me when I plan to visit car races or when I have to check if everything is all right under the hood of my car.

Anne-Johan Annema