MATHEMATICAL ASPECTS OF DEEP LEARNING

In recent years the development of new classification and regression algorithms based on deep learning has led to a revolution in the fields of artificial intelligence, machine learning, and data analysis. The development of a theoretical foundation to guarantee the success of these algorithms constitutes one of the most active and exciting research topics in applied mathematics.

This book presents the current mathematical understanding of deep learning methods from the point of view of the leading experts in the field. It serves both as a starting point for researchers and graduate students in computer science, mathematics, and statistics trying to get into the field and as an invaluable reference for future research.

PHILIPP GROHS is Professor of Applied Mathematics at the University of Vienna and Group Leader of Mathematical Data Science at the Austrian Academy of Sciences.

GITTA KUTYNIOK is Bavarian AI Chair for Mathematical Foundations of Artificial Intelligence at Ludwig-Maximilians-Universität München and Adjunct Professor for Machine Learning at the University of Tromsø.

MATHEMATICAL ASPECTS OF DEEP LEARNING

Edited by

PHILIPP GROHS
Universität Wien, Austria

GITTA KUTYNIOK
Ludwig-Maximilians-Universität München

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314-321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781316516782
DOI: 10.1017/9781009025096

First published 2023

Printed in the United Kingdom by TJ Books Limited, Padstow Cornwall

A catalogue record for this publication is available from the British Library.

ISBN 978-1-316-51678-2 Hardback

Cambridge University Press & Assessment
978-1-316-51678-2 — Mathematical Aspects of Deep Learning
Edited by Philipp Grohs , Gitta Kutyniok
Frontmatter
More Information

Contents

Cambridge University Press & Assessment
978-1-316-51678-2 — Mathematical Aspects of Deep Learning
Edited by Philipp Grohs , Gitta Kutyniok
Frontmatter
More Information

Contributors

Aviad Aberdam *Electrical Engineering Department, Technion – Israel Institute of Technology, Haifa 32000, Israel*

Leila Arras *Department of Artificial Intelligence, Fraunhofer Heinrich Hertz Institute, Einsteinufer 37, 10587 Berlin, Germany*

Julius Berner *Faculty of Mathematics, University of Vienna, Universitätsring 1, 1010 Wien, Austria*

Yoshua Bengio *University of Montreal, 2900 Edouard Montpetit Blvd, Montreal, QC H3T 1J4, Canada*

Joan Bruna *Courant Institute and Center for Data Science, New York University, 251 Mercer Street, New York, NY 10012, USA*

Nadav Cohen *Department of Computer Science and Applied Mathematics Weizmann Institute of Science Rehovot 7610001, Israel*

Alexandros G. Dimakis *Department of Electrical and Computer Engineering, University of Texas, Austin TX 78712, USA*

Weinan E *Department of Mathematics, Princeton University, Princeton, NJ 08544–1000, USA.*

Philipp Grohs *Faculty of Mathematics and Research Platform Data Science, University of Vienna, Oskar Morgenstern Platz 1, 1090 Wien, Austria*

Ingo Gühring *Institut für Mathematik, Technische Universität Berlin, Straße des 17. Juni 136, 10623 Berlin*

Benjamin D. Haeffele *Mathematical Institute for Data Science, Department of Biomedical Engineering, Johns Hopkins University, 3400 N. Charles Street, Baltimore MD 21218, USA*

Jiequn Han *Center for Computational Mathematics, Flatiron Institute, 162 5th Ave. New York, NY 10010, USA*

Leslie Pack Kaelbling *Massachusetts Institute of Technology, Computer Science & Artificial Intelligence Laboratory, 32 Vassar St., Cambridge, MA 02139, USA*

Kenji Kawaguchi *Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA*

Gitta Kutyniok *Mathematisches Institut der Universität München, Theresienstraße, 80333 München, Germany*

Yoav Levine *School of Computer Science and Engineering, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel*

Qianxiao Li *Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076.*

Thomas Merkh *Department of Mathematics and Department of Statistics, UCLA, Los Angeles, CA 90095, USA*

Grégoire Montavon *Institute of Software Engineering and Theoretical Computer Science, TU Berlin, D-10587 Berlin, Germany*

Guido Montúfar *Department of Mathematics and Department of Statistics, UCLA, CA 90095, USA and Max Planck Institute for Mathematics in the Sciences, 04103 Leipzig, Germany*

Klaus-Robert Müller *Institute of Software Engineering and Theoretical Computer Science, TU Berlin, D-10587 Berlin, Germany*

Ahmed Osman *Department of Artificial Intelligence, Fraunhofer Heinrich Hertz Institute, Einsteinufer 37, 10587 Berlin, Germany*

Philipp Petersen *Faculty of Mathematics and Research Platform Data Science, University of Vienna, Oskar Morgenstern Platz 1, 1090 Wien, Austria*

Mones Raslan *Institut für Mathematik, Technische Universität Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany*

Wojciech Samek *Department of Artificial Intelligence, Fraunhofer Heinrich Hertz Institute, Einsteinufer 37, 10587 Berlin, Germany*

Or Sharir *School of Computer Science and Engineering, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel*

Amnon Shashua *School of Computer Science and Engineering, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel*

Jeremias Sulam *Biomedical Engineering Department & Mathematical Institute for Data Science, Johns Hopkins University, Homewood Campus, Baltimore MD 21218, USA*

René Vidal *Mathematical Institute for Data Science and Department of Biomedical Engineering, Johns Hopkins University, Clark 302B, 3400 N. Charles Street, Baltimore MD 21218, USA*

Zhihui Zhu *Department of Electrical & Computer Engineering, University of Denver, 2155 E. Wesley Avenue, Denver CO 80208, USA*

Preface

We currently are witnessing the spectacular success of "deep learning" in both science (for example, in astronomy, biology, and medicine) and the public sector, where autonomous vehicles and robots are already present in daily life. However, the development of a rigorous mathematical foundation for deep learning is at an early stage, and most of the related research is still empirically driven. At the same time, methods based on deep neural networks have already shown their impressive potential in mathematical research areas such as imaging sciences, inverse problems, and the numerical analysis of partial differential equations, sometimes far outperforming classical mathematical approaches for particular classes of problem. This book provides the first comprehensive introduction to the subject, highlighting recent theoretical advances as well as outlining the numerous remaining research challenges.

The model of a deep neural network is inspired by the structure of the human brain, with artificial neurons concatenated and arranged in layers, leading to an (artificial feed-forward) neural network. Because of the structure of artificial neurons, the realization of such a neural network, i.e., the function it provides, consists of compositions of affine linear maps and (non-linear) activation functions $\varrho\colon \mathbb{R} \to \mathbb{R}$. More precisely, the realization of a neural network with L layers, and N_0, N_L, and N_ℓ, $\ell = 1, \ldots, L-1$, the number of neurons in the input, output, and ℓth hidden layer, as well as weight matrices and bias vectors, $W^{(\ell)} \in \mathbb{R}^{N_\ell \times N_{\ell-1}}$ and $b^{(\ell)} \in \mathbb{R}^{N_\ell}$, respectively, is given by

$$\Phi(x, \theta) = W^{(L)} \rho(W^{(L-1)} \cdots \rho(W^{(1)} x + b^{(1)}) + \cdots + b^{(L-1)}) + b^{(L)}, \qquad x \in \mathbb{R}^{N_0},$$

with free parameters $\theta = \big((W^{(\ell)}, b^{(\ell)})\big)_{\ell=1}^{L}$. Given training data

$$(z^{(i)})_{i=1}^{m} := ((x^{(i)}, y^{(i)}))_{i=1}^{m},$$

which arise from a function $g\colon \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}$, the parameters are then learned by minimizing the empirical risk

$$\frac{1}{m}\sum_{i=1}^{m}\mathcal{L}(\Phi(\cdot,\theta),z^{(i)}),$$

with $\mathcal{L}$ a suitable loss function. This is commonly achieved by stochastic gradient descent, which is a variant of gradient descent accommodating the obstacle that the number of parameters and training samples is typically in the millions. The performance is then measured by the fit of the trained neural network to a test data set.

This leads to three main research directions in the theory of deep learning, namely: (1) expressivity, i.e., studying the error accrued in approximating g by the hypothesis class of deep neural networks; (2) optimization, which studies the algorithmic error using minimization of the empirical risk; and (3) generalization, which aims to understand the out-of-sample error. Expressivity is at present from a theoretical viewpoint the most advanced research direction; a current key question is the impact on the overall performance of various architectural components of neural networks, such as their depth. Optimization has recently seen intriguing new results. However, the main mystery of why stochastic gradient descent converges to good local minima despite the non-convexity of the problem is as yet unraveled. Finally, generalization is the direction that is the least explored so far, and a deep theoretical understanding of, for instance, why highly overparametrized models often do not overfit, is still out of reach. These core theoretical directions are complemented by others such as explainability, fairness, robustness, or safety – sometimes summarized as the reliability of deep neural networks. Interestingly, basically the entire field of mathematics, ranging from algebraic geometry through to approximation theory and then to stochastics is required to tackle these challenges, which often even demand the development of novel mathematics. And, in fact, at a rapidly increasing rate, mathematicians from all areas are joining the field and contributing with their unique expertise.

Apart from the development of a mathematical foundation of deep learning, deep learning has also a tremendous impact on mathematical approaches to other areas such as solving inverse problems or partial differential equations. In fact, it is fair to say that the area of inverse problems, in particular imaging science, has already undergone a paradigm shift towards deep-learning-based approaches. The area of the numerical analysis of partial differential equations has been slower to embrace these novel methodologies, since it was initially not evident what their advantage would be for this field. However, by now there exist various results of both a numerical and a theoretical nature showing that deep neural networks are capable of beating the curse of dimensionality while providing highly flexible and fast solvers. This observation has led to the fact that this area is also currently being swept by deep-learning-type approaches, requiring the development of a theoretical foundation as well.

This book is the first monograph in the literature to provide a comprehensive survey of the mathematical aspects of deep learning. Its potential readers could be researchers in the areas of applied mathematics, computer science, and statistics, or a related research area, or they could be graduate students seeking to learn about the mathematics of deep learning. The particular design of this volume ensures that it can serve as both a state-of-the-art reference for researchers as well as a textbook for students.

The book contains 11 diverse chapters written by recognized leading experts from all over the world covering a large variety of topics. It does not assume any prior knowledge in the field. The chapters are self-contained, covering the most recent research results in the respective topic, and can all be treated independently of the others. A brief summary of each chapter is given next.

Chapter 1 provides a comprehensive introduction to the mathematics of deep learning, and serves as a background for the rest of the book. The chapter covers the key research directions within both the mathematical foundations of deep learning and deep learning approaches to solving mathematical problems. It also discusses why there is a great need for a new theory of deep learning, and provides an overview of the main future challenges.

Chapter 2 provides a comprehensive introduction to generalization properties of deep learning, emphasizing the specific phenomena that are special to deep learning models. Towards analyzing the generalization behavior of deep neural networks, the authors then present generalization bounds based on validation datasets and an analysis of generalization errors based on training datasets.

Chapter 3 surveys a recent body of work related to the expressivity of model classes of neural networks. The chapter covers results providing approximation rates for diverse function spaces as well as those shedding light on the question of why the depth of a neural network is important. The overview not only focuses on feed-forward neural networks, but also includes convolutional, residual, and recurrent ones.

Chapter 4 presents recent advances concerning the algorithmic solution of optimization problems that arise in the context of deep learning, in the sense of analyzing the optimization landscape of neural network training. A specific focus is on linear networks trained with a squared loss and without regularization as well as on deep networks with a parallel structure, positively homogeneous network mapping and regularization, and that have been trained with a convex loss.

Chapter 5 summarizes recent approaches towards rendering deep-learning-based classification decisions interpretable. It first discusses the algorithmic and theoretical aspects of an approach called Layer-wise Relevance Propagation (LRP). This is a propagation-based method, allowing us to derive explanations of the decisions of a variety of ML models. The authors also demonstrate how this method can be applied to a complex model trained for the task of visual question answering.

Chapter 6 introduces stochastic feed-forward neural networks, one prominent example of which is deep belief networks. The authors first review existing expressivity results for this class of networks. They then analyze the question of a universal approximation for shallow networks and present a unified analysis for several classes of such deep networks.

Chapter 7 explores connections between deep learning and sparsity-enforcing algorithms. More precisely, this chapter reviews and builds on previous work on a novel interpretation of deep neural networks from a sparsity viewpoint, namely as pursuit algorithms aiming for sparse representations, provided that the signals belong to a multilayer synthesis sparse model. The authors then present extensions of this conceptual approach and demonstrate the advantage of the resulting algorithms in a specific supervised learning setting, leading to an improvement of performance while retaining the number of parameters.

Chapter 8 provides a comprehensive introduction of the scattering transform. The author presents both mathematical results, showing that geometric stability indeed plays a key role in deep learning representations, and applications to, for instance, computer vision. Also, more general group-invariant feature descriptors in terms of Lie groups and non-Euclidean domains are described.

Chapter 9 focuses on the application of deep neural networks to solving inverse problems. The author provides an introduction to the use of generative deep learning models as priors in the regularization of inverse problems. Also, the specific setting of a compressed sensing problem is studied and both mathematical and numerical results in compressed sensing for deep generative models are presented.

Chapter 10 introduces a reformulation of the training process for residual neural networks as well as a corresponding theory. More precisely, the dynamical systems viewpoint regards the back-propagation algorithm as a simple consequence of variational equations in ordinary differential equations, whereas the control theory viewpoint regards deep learning as one instance of mean-field control where all agents share the same control. The authors finally introduce a new class of algorithms for deep learning as one application of these conceptual viewpoints.

Chapter 11 illuminates the connections between tensor networks and convolutional neural networks. These are established by relating one of the current goals of the field of many-body physics, namely the efficient representation of highly entangled many-particle quantum systems, to the area of deep learning. As one application of this framework, the authors derive a new entanglement-based deep learning design scheme which allows theoretical insight in a wide variety of customarily used network architectures.