

## 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 & Assessment  
978-1-316-51678-2 — Mathematical Aspects of Deep Learning  
Edited by Philipp Grohs , Gitta Kutyniok  
Frontmatter  
[More Information](#)

---

## 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](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9781316516782](http://www.cambridge.org/9781316516782)

DOI: 10.1017/9781009025096

© Cambridge University Press 2023

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

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 has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

<b>Contributors</b>	<i>page</i> xiii
<b>Preface</b>	xv
<b>1 The Modern Mathematics of Deep Learning</b>	
<i>Julius Berner, Philipp Grohs, Gitta Kutyniok and Philipp Petersen</i>	1
1.1 Introduction	1
1.1.1 Notation	4
1.1.2 Foundations of Learning Theory	5
1.1.3 Do We Need a New Theory?	23
1.2 Generalization of Large Neural Networks	31
1.2.1 Kernel Regime	31
1.2.2 Norm-Based Bounds and Margin Theory	33
1.2.3 Optimization and Implicit Regularization	35
1.2.4 Limits of Classical Theory and Double Descent	38
1.3 The Role of Depth in the Expressivity of Neural Networks	41
1.3.1 Approximation of Radial Functions	41
1.3.2 Deep ReLU Networks	44
1.3.3 Alternative Notions of Expressivity	47
1.4 Deep Neural Networks Overcome the Curse of Dimensionality	49
1.4.1 Manifold Assumption	49
1.4.2 Random Sampling	51
1.4.3 PDE Assumption	53
1.5 Optimization of Deep Neural Networks	57
1.5.1 Loss Landscape Analysis	57
1.5.2 Lazy Training and Provable Convergence of Stochastic Gradient Descent	61
1.6 Tangible Effects of Special Architectures	65
1.6.1 Convolutional Neural Networks	66

1.6.2	Residual Neural Networks	68
1.6.3	Framelets and U-Nets	70
1.6.4	Batch Normalization	73
1.6.5	Sparse Neural Networks and Pruning	75
1.6.6	Recurrent Neural Networks	76
1.7	Describing the Features that a Deep Neural Network Learns	78
1.7.1	Invariances and the Scattering Transform	78
1.7.2	Hierarchical Sparse Representations	79
1.8	Effectiveness in Natural Sciences	81
1.8.1	Deep Neural Networks Meet Inverse Problems	82
1.8.2	PDE-Based Models	84
2	<b>Generalization in Deep Learning</b>	
	<i>K. Kawaguchi, Y. Bengio, and L. Kaelbling</i>	112
2.1	Introduction	112
2.2	Background	113
2.3	Rethinking Generalization	116
2.3.1	Consistency of Theory	118
2.3.2	Differences in Assumptions and Problem Settings	119
2.3.3	Practical Role of Generalization Theory	121
2.4	Generalization Bounds via Validation	121
2.5	Direct Analyses of Neural Networks	122
2.5.1	Model Description via Deep Paths	123
2.5.2	Theoretical Insights via Tight Theory for Every Pair ( $\mathbb{P}, S$ )	125
2.5.3	Probabilistic Bounds over Random Datasets	127
2.5.4	Probabilistic Bound for 0–1 Loss with Multi-Labels	130
2.6	Discussions and Open Problems	131
Appendix A	Additional Discussions	133
A1	Simple Regularization Algorithm	133
A2	Relationship to Other Fields	135
A3	SGD Chooses Direction in Terms of $\bar{w}$	135
A4	Simple Implementation of Two-Phase Training Procedure	136
A5	On Proposition 2.3	136
A6	On Extensions	137
Appendix B	Experimental Details	137
Appendix C	Proofs	138
C1	Proof of Theorem 2.1	139
C2	Proof of Corollary 2.2	139

	<i>Contents</i>	vii
	C3 Proof of Theorem 2.7	140
	C4 Proof of Theorem 2.9	141
	C5 Proof of Theorem 2.10	142
	C6 Proof of Proposition 2.5	143
<b>3</b>	<b>Expressivity of Deep Neural Networks</b>	
	<i>Ingo Gühring, Mones Raslan, and Gitta Kutyniok</i>	149
3.1	Introduction	149
3.1.1	Neural Networks	151
3.1.2	Goal and Outline of this Chapter	154
3.1.3	Notation	154
3.2	Shallow Neural Networks	155
3.2.1	Universality of Shallow Neural Networks	156
3.2.2	Lower Complexity Bounds	159
3.2.3	Upper Complexity Bounds	160
3.3	Universality of Deep Neural Networks	161
3.4	Approximation of Classes of Smooth Functions	163
3.5	Approximation of Piecewise Smooth Functions	167
3.6	Assuming More Structure	172
3.6.1	Hierachical Structure	172
3.6.2	Assumptions on the Data Manifold	174
3.6.3	Expressivity of Deep Neural Networks for Solutions of PDEs	175
3.7	Deep Versus Shallow Neural Networks	177
3.8	Special Neural Network Architectures and Activation Functions	180
3.8.1	Convolutional Neural Networks	180
3.8.2	Residual Neural Networks	184
3.8.3	Recurrent Neural Networks	185
<b>4</b>	<b>Optimization Landscape of Neural Networks</b>	
	<i>René Vidal, Zhihui Zhu, and Benjamin D. Haeffele</i>	200
4.1	Introduction	201
4.2	Basics of Statistical Learning	205
4.3	Optimization Landscape of Linear Networks	206
4.3.1	Single-Hidden-Layer Linear Networks with Squared Loss and Fixed Size Regularization	207
4.3.2	Deep Linear Networks with Squared Loss	212
4.4	Optimization Landscape of Nonlinear Networks	214
4.4.1	Motivating Example	215
4.4.2	Positively Homogeneous Networks	221
4.5	Conclusions	225

<b>5</b>	<b>Explaining the Decisions of Convolutional and Recurrent Neural Networks</b>	
	<i>Wojciech Samek, Leila Arras, Ahmed Osman, Grégoire Montavon, Klaus-Robert Müller</i>	229
5.1	Introduction	229
5.2	Why Explainability?	231
5.2.1	Practical Advantages of Explainability	231
5.2.2	Social and Legal Role of Explainability	232
5.2.3	Theoretical Insights Through Explainability	232
5.3	From Explaining Linear Models to General Model Explainability	233
5.3.1	Explainability of Linear Models	233
5.3.2	Generalizing Explainability to Nonlinear Models	235
5.3.3	Short Survey on Explanation Methods	236
5.4	Layer-Wise Relevance Propagation	238
5.4.1	LRP in Convolutional Neural Networks	239
5.4.2	Theoretical Interpretation of the LRP Redistribution Process	242
5.4.3	Extending LRP to LSTM Networks	248
5.5	Explaining a Visual Question Answering Model	251
5.6	Discussion	258
<b>6</b>	<b>Stochastic Feedforward Neural Networks: Universal Approximation</b>	
	<i>Thomas Merkh and Guido Montúfar</i>	267
6.1	Introduction	268
6.2	Overview of Previous Works and Results	271
6.3	Markov Kernels and Stochastic Networks	273
6.3.1	Binary Probability Distributions and Markov Kernels	273
6.3.2	Stochastic Feedforward Networks	274
6.4	Results for Shallow Networks	276
6.4.1	Fixed Weights in the Output Layer	277
6.4.2	Trainable Weights in the Output Layer	278
6.5	Proofs for Shallow Networks	278
6.5.1	Fixed Weights in the Output Layer	279
6.5.2	Trainable Weights in the Second Layer	283
6.5.3	Discussion of the Proofs for Shallow Networks	285
6.6	Results for Deep Networks	286
6.6.1	Parameter Count	288
6.6.2	Approximation with Finite Weights and Biases	288

	<i>Contents</i>	ix
6.7	Proofs for Deep Networks	289
6.7.1	Notation	289
6.7.2	Probability Mass Sharing	290
6.7.3	Universal Approximation	293
6.7.4	Error Analysis for Finite Weights and Biases	296
6.7.5	Discussion of the Proofs for Deep Networks	298
6.8	Lower Bounds for Shallow and Deep Networks	299
6.8.1	Parameter Counting Lower Bounds	299
6.8.2	Minimum Width	301
6.9	A Numerical Example	302
6.10	Conclusion	306
6.11	Open Problems	307
<b>7</b>	<b>Deep Learning as Sparsity-Enforcing Algorithms</b>	
	<i>A. Aberdam and J. Sulam</i>	314
7.1	Introduction	314
7.2	Related Work	316
7.3	Background	317
7.4	Multilayer Sparse Coding	320
7.4.1	ML–SC Pursuit and the Forward Pass	321
7.4.2	ML–SC: A Projection Approach	323
7.5	The Holistic Way	324
7.6	Multilayer Iterative Shrinkage Algorithms	327
7.6.1	Towards Principled Recurrent Neural Networks	329
7.7	Final Remarks and Outlook	332
<b>8</b>	<b>The Scattering Transform</b>	
	<i>Joan Bruna</i>	338
8.1	Introduction	338
8.2	Geometric Stability	339
8.2.1	Euclidean Geometric Stability	340
8.2.2	Representations with Euclidean Geometric Stability	341
8.2.3	Non-Euclidean Geometric Stability	342
8.2.4	Examples	343
8.3	Scattering on the Translation Group	346
8.3.1	Windowed Scattering Transform	346
8.3.2	Scattering Metric and Energy Conservation	349
8.3.3	Local Translation Invariance and Lipschitz Continuity with Respect to Deformations	351
8.3.4	Algorithms	354
8.3.5	Empirical Analysis of Scattering Properties	357



8.3.6	Scattering in Modern Computer Vision	362
8.4	Scattering Representations of Stochastic Processes	363
8.4.1	Expected Scattering	363
8.4.2	Analysis of Stationary Textures with Scattering	367
8.4.3	Multifractal Analysis with Scattering Moments	369
8.5	Non-Euclidean Scattering	371
8.5.1	Joint versus Separable Scattering	372
8.5.2	Scattering on Global Symmetry Groups	372
8.5.3	Graph Scattering	375
8.5.4	Manifold Scattering	383
8.6	Generative Modeling with Scattering	384
8.6.1	Sufficient Statistics	384
8.6.2	Microcanonical Scattering Models	385
8.6.3	Gradient Descent Scattering Reconstruction	387
8.6.4	Regularising Inverse Problems with Scattering	389
8.6.5	Texture Synthesis with Microcanonical Scattering	391
8.7	Final Remarks	393
<b>9</b>	<b>Deep Generative Models and Inverse Problems</b>	
	<i>Alexandros G. Dimakis</i>	400
9.1	Introduction	400
9.2	How to Tame High Dimensions	401
9.2.1	Sparsity	401
9.2.2	Conditional Independence	402
9.2.3	Deep Generative Models	403
9.2.4	GANs and VAEs	404
9.2.5	Invertible Generative Models	405
9.2.6	Untrained Generative Models	405
9.3	Linear Inverse Problems Using Deep Generative Models	406
9.3.1	Reconstruction from Gaussian Measurements	407
9.3.2	Optimization Challenges	409
9.3.3	Extending the Range of the Generator	410
9.3.4	Non-Linear Inverse Problems	410
9.3.5	Inverse Problems with Untrained Generative Priors	412
9.4	Supervised Methods for Inverse Problems	414
<b>10</b>	<b>Dynamical Systems and Optimal Control Approach to Deep Learning</b>	
	<i>Weinan E, Jiequn Han, and Qianxiao Li</i>	422
10.1	Introduction	422
10.1.1	The Problem of Supervised Learning	423

	<i>Contents</i>	xi
10.2	ODE Formulation	424
10.3	Mean-Field Optimal Control and Pontryagin’s Maximum Principle	425
10.3.1	Pontryagin’s Maximum Principle	426
10.4	Method of Successive Approximations	428
10.4.1	Extended Pontryagin Maximum Principle	428
10.4.2	The Basic Method of Successive Approximation	428
10.4.3	Extended Method of Successive Approximation	431
10.4.4	Discrete PMP and Discrete MSA	433
10.5	Future Work	435
<b>11</b>	<b>Bridging Many-Body Quantum Physics and Deep Learning via Tensor Networks</b>	
	<i>Yoav Levine, Or Sharir, Nadav Cohen and Amnon Shashua</i>	439
11.1	Introduction	440
11.2	Preliminaries – Many-Body Quantum Physics	442
11.2.1	The Many-Body Quantum Wave Function	443
11.2.2	Quantum Entanglement Measures	444
11.2.3	Tensor Networks	447
11.3	Quantum Wave Functions and Deep Learning Architectures	450
11.3.1	Convolutional and Recurrent Networks as Wave Functions	450
11.3.2	Tensor Network Representations of Convolutional and Recurrent Networks	453
11.4	Deep Learning Architecture Design via Entanglement Measures	453
11.4.1	Dependencies via Entanglement Measures	454
11.4.2	Quantum-Physics-Inspired Control of Inductive Bias	456
11.5	Power of Deep Learning for Wave Function Representations	460
11.5.1	Entanglement Scaling of Deep Recurrent Networks	461
11.5.2	Entanglement Scaling of Overlapping Convolutional Networks	463
11.6	Discussion	467

## 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  $\varrho: \mathbb{R} \rightarrow \mathbb{R}$ . More precisely, the realization of a neural network with  $L$  layers, and  $N_0$ ,  $N_L$ , and  $N_\ell$ ,  $\ell = 1, \dots, L-1$ , the number of neurons in the input, output, and  $\ell$ th 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

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

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

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

which arise from a function  $g: \mathbb{R}^{N_0} \rightarrow \mathbb{R}^{N_L}$ , 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.